E3S Web of Conferences **399**, 01001 (2023) https://doi.org/10.1051/e3sconf/202339901001
*ICONNECT-2023*

# MACHINE LEARNING BASED AUTOMATIC LEAF DISEASES DETECTION

*Prasad* J[1,*], *RubanKumar* J[2], *Thirumoorthi* G[3], *UdayaKeerthi*[4] V S and *Santhosh* S M[5]

[1]Assistant Professor, Department of ECE, KPRIET, Coimbatore.
[2,3,4,5] Student, Department of ECE, KPRIET, Coimbatore.

**Abstract.** The method for applying machine learning to automatically detect leaf diseases is presented in this paper. A convolutional neural network was used to extract pertinent features from leaf image datasets that included healthy and diseased leaves. The dataset was compiled and pre-processed. Accuracy, precision, and recall measures were used to assess the machine learning algorithm after it had been trained on the labeled dataset. According to the findings, the algorithm was very precise and recallable in its ability to detect leaf illnesses, making it a potential method for practical use. This strategy may help with early leaf disease identification and prevention, increasing crop productivity and lowering the demand for toxic pesticides. Here we are identifying the Bacterial spot, Early blight.

## 1. INTRODUCTION

Farming is difficult because leaf diseases can negatively affect crop yield and quality. Early diagnosis and prevention are essential for these illnesses to be effectively managed, but manually identifying and monitoring crop infections can be challenging and time-consuming for farmers. Machine learning methods have been a potential strategy for automating the diagnosis of leaf diseases in recent years. Machine learning algorithms can help effectively identify diseases and give farmers an early warning by utilizing vast datasets of tagged leaf photos. [1]Farmers with little experience could misidentify cattle and misuse medications. Environmental pollution will result from poor quality, insufficient productivity, and unnecessary financial losses. The use of image processing methods for the detection of plant diseases has emerged as a hot area of research to address these issues.

This study presents an approach for automatic leaf disease detection using machine learning. We collected a dataset of leaf images with both healthy and diseased leaves, which were pre-processed and used to train a convolutional neural network to extract relevant features. The resulting machine-learning algorithm was evaluated using standard metrics and showed high accuracy, precision, and recall in detecting leaf diseases.

* Corresponding author: prasad7research@gmail.com

The early diagnosis and control of leaf diseases, increasing crop output, and lowering the use of toxic pesticides are all problems that our strategy may help with. We think that this research can be used as a starting point for future work on creating automated agricultural systems that will increase farming productivity and sustainability.

## 2. OVERVIEW EXISTING SYSTEM

A new field, automatic leaf disease detection using machine learning, uses image processing methods and machine learning algorithms to detect and diagnose diseases in plant leaves. Rule-based and machine- learning-based methods [12] can be used to categorize the automatic leaf disease detection systems now in use. Rule- based systems rely on a preset set of guidelines based on the visual traits of the condition. These guidelines are often created by subject-matter specialists and used to assess the health of a plant leaf from a particular photograph. Contrarily, machine learning-based systems automatically learn the visual characteristics of healthy and diseased leaves using statistical models and algorithms. [2]Typically, these systems need a sizable dataset of tagged photos of healthy and to train the machine-learning models, use sick leaves. Convolutional neural networks (CNNs), support vector machines (SVMs), decision trees, and random forests are a few of the well-known machine-learning methods utilized in autonomous leaf disease diagnosis. To learn how to discern between healthy and unhealthy leaves visually, these algorithms are trained on a large dataset of annotated photos of leaves. Automatic leaf disease detection systems' effectiveness is often assessed using measures including accuracy, precision, recall, and F1 score. The quality of the dataset, the intricacy of the condition, and the machine learning technique selected can all affect how accurate these systems are. Overall, machine learning- based automatic [11] leaf disease identification has the potential to dramatically increase the effectiveness and precision of plant disease diagnosis, allowing farmers andagricultural researchers to recognize and treat plant diseases more successfully. Thus, the diseases are to be identified by using Machine-Learning. In Machine- Learning we include algorithms like GLCM, K-means Clustering, and SVM, and the methods are explained by following the Architecture System.

## 3. PROPOSED SYSTEM

A proposed automatic leaf disease detection system using machine learning would aim to improve the existing systems by incorporating advanced image processing techniques and more sophisticated machine learning algorithms. The following are some key components of such a system:

1. High-quality dataset: The system would require a large, diverse, and well-labeled dataset of healthy and diseased plant leaf images to effectively train the machine learning models.
2. Preprocessing techniques: To make sure that the incoming photos are of good quality and standardized, the system would employ advanced preprocessing techniques, such as image normalization, color correction, and noise reduction.
3. Feature extraction: The system would extract relevant visual features from the input images, such as texture, shape, and color, using advanced techniques such as wavelet transform, Gabor filters, and color histograms.

4. Machine learning algorithms: The system would learn the visual characteristics of healthy and diseased leaves and categorize them in accordance using a variety of machine learning algorithms[2], including convolutional neural networks (CNNs), deep belief networks (DBNs), and recurrent neural networks (RNNs).
5. Ensemble learning: The system would incorporate ensemble learning techniques, such as bagging and boosting, to improve the accuracy and robustness of the machine learning models.
6. Real-time processing: The system would be optimized for real-time processing, allowing farmers and agricultural researchers to diagnose plant diseases quickly and accurately in the field.
7. User interface: The system would have a user-friendly interface [3] that allows users to easily upload images of plant leaves and receive a quick and accurate diagnosis of any diseases present.

## 4. ANALYSIS REPORT

An analysis report of automatic leaf disease detection using machine learning can provide insights into the field's current state and highlight improvement areas. The following are some key aspects that can be included in such a report:

1. **Dataset:** The dataset used to train the machine learning models should be evaluated for its quality and diversity. To guarantee that the models are precise and reliable, the dataset should contain enough photos of both healthy and diseased leaves.
2. **Preprocessing techniques:** The analysis should assess the effectiveness of the preprocessing techniques used to prepare the input images for machine learning. Techniques such as normalization, color correction, and noise reduction canimprove the accuracy of the models.
3. **Feature extraction:** The analysis should evaluate the effectiveness of the feature extraction techniques used to identifyrelevant visual features in the input images. Techniques such as wavelet transform Gabor filters, and color histograms can capture the texture, shape, and color of theleaves, respectively.
4. **Machine learning algorithms:** To detect leaf illnesses, the study [2] should assess how well several machine learning algorithms—including convolutional neural networks (CNNs), support vector machines (SVMs), decision trees, and random forests—perform. The accuracy, speed, and complexity of the models can beimpacted by algorithmic choice.
5. **Evaluation metrics:** To assess the effectiveness of the models, the analysis should make use of the proper assessment metrics, such as accuracy, precision, recall, and F1 score. The measurements can shed light on the advantages and disadvantages of the models.
6. **Real-world application:** The analysis should assess the feasibility and usability of the automatic leaf disease detection system [3] in real-world applications. The system should be optimized for speed, accuracy, and user-friendliness to enable farmers and agricultural researchers to diagnose plant

 https://doi.org/10.1051/e3sconf/202339901001

diseases efficiently. Fig 1 represents the Analysis of the report.

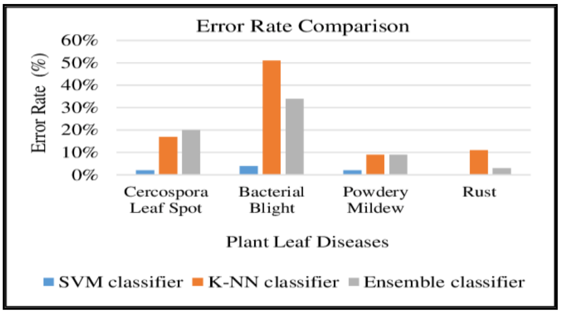


**Fig. 1.** Analysis report of automatic leaf disease

In conclusion, a study on the analysis of machine learning-based automatic detection of leaf diseases can be very helpful to researchers and practitioners in the area, allowing them to create more effective and efficient systems for plant disease diagnosisand management.

## 5. ARCHITECTURE OF SYSTEM

Data capture, Data preprocessing, Feature extraction, Classification, Evaluation, Deployment, and Continuous improvement are just a few of the parts that make up the architecture of a system for automatically detecting leaf disease using machine-learning photos.

**Image preprocessing:** To prepare the photos for the input of the Deep Learning models, we should randomly crop each image in the experiment to be 224 * 224 for AlexNet, SqueezeNet, VGG, DenseNet, and ResNet, as well as 299 * 299 for Inception [4]. Since bilinear interpolation can lessen the stepped effect brought on by the nearest neighbor method and make the images appear smoother, we utilized it to crop and resize the photographs. Some typical image-preprocessing methods used in this content include the following:

1. **Image resizing:** The input images may have different sizes and aspect ratios, which can affect the performance of the machine learning algorithms. Resizing the images to a standardized size can reduce the variation in image dimensions and facilitate the training process.
2. **Color correction:** The lighting and environmental conditions in which the images are captured can affect the color distribution and brightness of the images. Applying color correction techniques such as histogram equalization or gamma correction can improve the contrast and visual appearance of the images.
3. **Noise reduction:** Several types of noise, such as Gaussian noise, salt and pepper noise, or speckle noise, may be present in the input photos, which can impair the quality of theimages and have an impact on how well the machine learning algorithms perform. Imagequality can be enhanced by applying noise reduction techniques like median filtering or

wavelet denoising.

4. **Image segmentation:** The input images may contain various parts, such as background, leaves, stems, and other objects, which can interfere with the machine learning algorithms' ability to identify the relevant features. Applying image segmentation techniques such as thresholding or region growing can separate the relevant parts of the image and reduce the noise.
5. **Data augmentation:** The input photos mightnot fully represent the range of leaf illnesses and healthy leaves, which could have an impact on how strong the machine learning models are [5]. The diversity of the picture dataset can be increased, and the generalization capabilities of the models can be strengthened, by applying data augmentation techniques like flipping, rotating, or adding noise to the images.

## 6. CLASSIFICATION OF DISEASES:

### 6.1 Early Blight

Alternaria tomato Phila and Alternaria solani are two completely distinct but linked fungi that produce early blight. As Alternaria tomato Phila is more virulent on tomatoes than A. solani, it is the primary cause of the early blight on tomatoes in areas where it is present. [6]A. solani, on the other hand, will produce early blight on tomatoes if A. tomato Phila is notpresent. All pathogens can also infect potatoes, however,

A. solani is more likely than A. tomato Phila to cause potato early blight. Every pathogen will even infect eggplant and several other asterid magnoliopsid weeds growing next to hirsute vascular plants and poisonberry (Solanum ptycanthum). Fig 2 represents the disease Earlyblight.

**Fig. 2.** Early blight

### 6.2. Bacterial spot

Bacterial spot is a worldwide problem that affects large tomatoes and is brought on by four species of the genus. Bacterial spots result in leaf and fruit spots, defoliation, sun-scalded fruit, and a reduction in output. The disease can arise at completely various temperatures and should pose a hazard to tomato production across the world due to the diversity among thespot pathogens. [6]High precipitation and temperatures between 75 and 86 °F encourage the development ofdisease. Fig 3 represents the disease Bacterial spot.

**Fig. 3.** Bacterial Spot

### 6.3 GLCM

A popular feature extraction method for automatic leaf disease detection using machine learning is the Grey Level Co-occurrence Matrix. [7]GLCM is a statistical technique that measures the frequency of pairs of pixels occurring with a particular spatial relationship and a particular grey-level distance in an image to quantify the relationship between them. The co-occurrence of pixel pairings at various angles and separations is detailed in the GLCM matrix. Contrast, correlation, energy, and homogeneity are GLCM features that can be retrieved from the matrix and utilized as inputs to machine learning algorithms for categorization. The steps listed below can be used to use GLCM for automatic leaf disease detection:

1. **Preprocessing:** Preprocess the input images using techniques such as resizing, color correction, and noise reduction, as describedearlier.
2. **Segmentation:** Segment the input images to extract the regions of interest, such as the leaf area.
3. **GLCM computation:** Compute the GLCM matrix for the segmented regions of interestusing a specific spatial relationship, such as a 0-degree or 45-degree direction, and a specific grey-level distance [3], such as 1 or 2 pixels.
4. **Feature extraction:** Extract the relevant GLCM features from the computed matrices, such as contrast, correlation, energy, and homogeneity.
5. **Classification:** Using the retrieved GLCM characteristics, classify the input photos as healthy or unhealthy using machine learning methods like SVMs, decision trees, or neural networks.
6. **Evaluation:** Analyze the categorization model's performance using recall, accuracy, andprecision measures.

Using GLCM for automatic leaf disease detection can improve the accuracy and robustness of the machine-learning models by capturing relevant texture and spatial information from the input images. However, the choice of GLCM parameters, such as the spatial relationship andgrey-level distance, can affect the performance of the GLCM features and the classification accuracy. Therefore, careful selection and optimization of GLCM parameters are necessary to achieve optimal performance.

## 6.4. K-means clustering

Automatic leaf disease identification uses a well-known unsupervised machine learning technique to cluster related data points. In this situation, k-means clustering can be used to organize the image features that were retrieved from the input photos after preprocessing them into clusters according to how similar they were. [8]K-means clustering for automatic leaf disease detection involves the following processes. The K-mean clustering flow chart is shown in Fig 4.

1. **Feature extraction:** Extract relevant features from the preprocessed input images, such as color, texture, andshape, using techniques such as GLCM.
2. **Feature normalization:** To ensure that all characteristics contribute equally to the clustering process, normalize the retrieved features to have a zero mean and unit variance.
3. **Cluster initialization:** Choose k initial cluster centroids at random, where k is the total number of clustersthat will develop.
4. **Cluster assignment:** Based on their Euclidean distance, assign each feature to the centroid of the closest cluster.
5. **Centroid update:** Update the cluster centroids based on the meaning of the features assigned to each cluster [8].
6. **Iteration:** Till the cluster assignments remain the same or the allotted number of iterations has beenreached, repeat steps 4 and 5.
7. **Cluster labeling:** Assign a label to each cluster based on the most frequent class of the features assignedto that cluster.
8. **Classification:** Use the cluster labels as input to a classification algorithm, such as SVMs or decision trees, to classify the input images as healthy or diseased.
9. **Evaluation:**Analyze the categorization model's performance using recall, accuracy, and precision measures.

By condensing the feature space and organizing related features into clusters, k-means clustering can increase the precision and efficacy of classification models used for autonomous leaf disease detection. The effectiveness of the clustering process and the accuracy ofthe resulting classification, however, can be impacted by the number of clusters, k. As a result, choosing the best value for k requires careful consideration to attain the best results.

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg\min_{\mathbf{S}} \sum_{i=1}^{k} |S_i| \, \mathrm{V} \quad (1)$$

Minimize the within clusters sum of the square and the object is to find the centroid point of the leaf.

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \frac{1}{|S_i|} \sum_{\mathbf{x},\mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2 \tag{2}$$

This is equivalent to minimizing the pairwise squared deviation of the points in the same cluster.

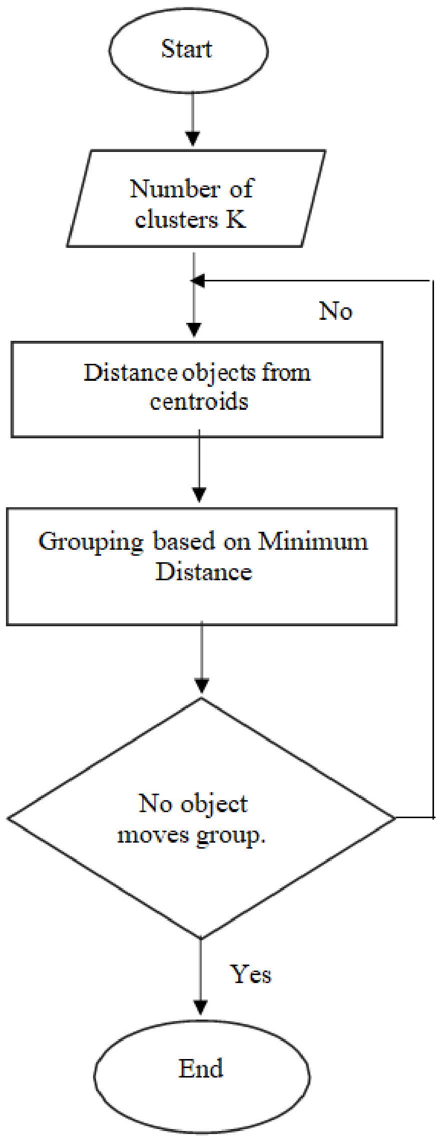


**Fig. 4.** K-Mean Clustering

### 6.4. SVM

Support Automatic leaf disease diagnosis uses the well-known machine learning technique Vector Machine to categorize input photos as healthy or sick based on extracted attributes.

[9]SVM operates by locating the hyperplane in the feature space that optimizes the margin between the two classes. To use SVM for automatic leaf disease identification, the following procedures must be taken:

1. **Feature extraction:** Extract relevant features from the preprocessed input images, such as color, texture, and shape, using techniques such as GLCM or CNNs.
2. **Feature normalization:** To ensure that all characteristics contribute equally to the clustering process, normalize the retrieved features to have a zero mean and unit variance.
3. **Data splitting:** Divide the feature data into training and testing sets, using the former to train the SVM model and the latter[5] to assess the model's effectiveness.
4. **SVM model training:** Using the extracted features as input and the matching class labels as output, train an SVM model on the training set.
5. **SVM model testing:** Test the trained SVM model on the testing set to evaluate its performance in classifying new images as healthy or diseased.
6. **Model optimization:** Train an SVM model on the training set using the extracted features as input and the corresponding class labels as output.
7. **Evaluation:** Analyze the SVM model's performance using recall, accuracy, and precision measures.

SVM is a powerful machine-learning algorithm for automatic leaf disease detection because it can handle high-dimensional feature spaces and nonlinear decision boundaries. However, SVM's performance can be affected by the choice of hyper parameters, the extracted features' quality, and the training dataset's size and quality. Therefore, careful selection and optimization of the SVM model and feature extraction techniques are necessary to achieve optimal performance.

### 6.5 Arduino UNO

An autonomous leaf disease detection system can employ an Arduino UNO microcontroller board to control various parts like sensors, actuators, and motors. Because of its low cost, adaptability, and ease of use, it is a preferred option for both professionals and enthusiasts. The Arduino UNO can be used to control the picture capture module, the preprocessing module, the feature extraction module, and the classification module in an automatic leaf disease detection system. The motor that triggers the pesticide spray, the display module that shows the classification results and the connectivity module that transmits data to a distant server are all additional components that can be controlled by the Arduino UNO. The Arduino IDE, a user-friendly environment for developing software that supports the C and C++ programming languages, can be used to program the Arduino UNO. Low-level programming is not required as the Arduino IDE comes with a library of prewritten code that can be used to connect with sensors and other components. The low power consumption of the Arduino UNO, which makes it appropriate for battery-powered applications, is one benefit of employing it in an autonomous leaf disease detection system. Additionally, the Arduino UNO is compatible with a large selection of sensors and modules, making it simple to integrate with other pieces of gear. Overall, the Arduino UNO is a cost-efficient and adaptable microcontroller board that may be utilized to manage different system components and enhance system performance in an autonomous leaf disease detection system.

### 6.6 Pesticide spray motor

The pesticide spray motor is an essential component of an automatic leaf disease detection system that can help control and prevent plant disease spread. Once an image of a diseased leaf is detected and classified, the pesticide spray motor can be activated to apply a suitable pesticide or fungicide to the affected plant. The pesticide spray motor can be controlled using a microcontroller or a PLC (Programmable Logic Controller) connected to the automatic leaf disease detection system. [10]The system can be programmed to activate the pesticide spray motor when a diseased leaf is detected and classified and to deactivate it when the spraying is complete. The pesticide spray motor can be a standard agricultural sprayer or a customized motor designed specifically for the automatic leaf disease detection system. It is important to ensure that the pesticide or fungicide used is safe for the plant and the environment and is applied in the correct concentration and volume. The use of a pesticide spray motor in automatic leaf disease detection can improve the efficiency and effectiveness of plant disease management by enabling timely and accurate detection and treatment of plant diseases. However, it is important to ensure that the system is properly calibrated and maintained to avoid overuse or misuse of pesticides, which can lead to environmental and health risks. Fig 5 represents the block diagram of the leaf diseases.

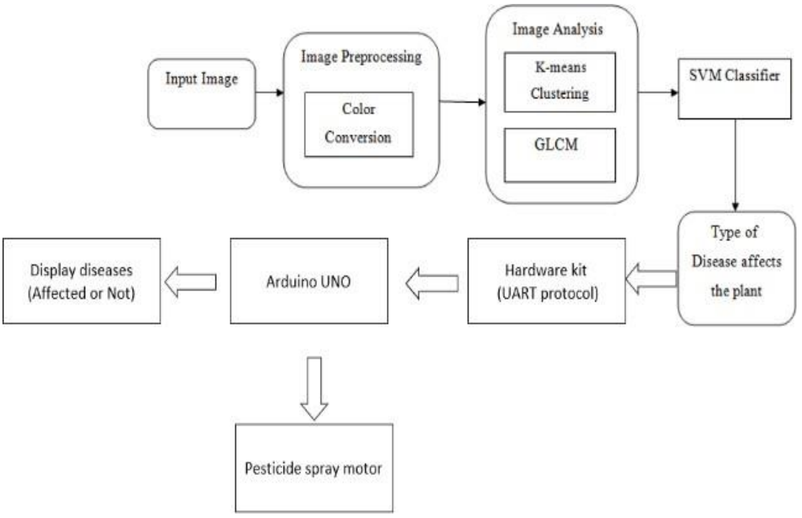


**Fig. 5.** Block diagram of leaf disease detection

## 7. RESULT & DISCUSION

In general, there are many diseases that affect the tomato plant, but in this case, we are using two diseases and a healthy leaf to demonstrate how the afflicted leaf differs from the healthy leaf. We are focusing on Bacterial Spots and Early Blight as our two illnesses. (Alternaria Solani). We took these diseases because they frequently impacted tomato plants


**Fig. 6.** Bacterial Spot of Software Simulation

### 7.1 Bacterial Spot

Bacterial Spot: In theprogram simulation, Fig. 6 shows the bacterial spot disease detection. Several functions, including segment image, GLCM conversion, and enhanced contrast. Once the conversion is complete, the findings will be classified and their accuracy in percentage will be determined. The hardware for showcasing pesticides will then be connected, and the Bonide liquid copper solution will then be sprayed. The hardware result for Bacterial Spot is shown in Fig 7.

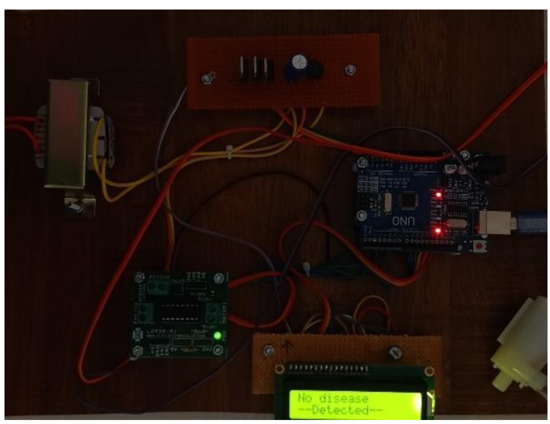


**Fig. 7.** Bacterial Spot of Hardware Simulation

### 7.2 Early Blight (A.Solani)

Early Blight (A. Solani): The Early Blight disease identification in the software simulation is shown in Fig 8. Several functions, including segment image, GLCM conversion, and enhanced contrast. Once the conversion is complete, the findings will be classified and their accuracy in percentage will be determined. The hardware for showcasing pesticides will then be connected, and the baking soda solution will then be sprayed. The hardware result for Early Blight is shown in Fig 9. (A.Solani).

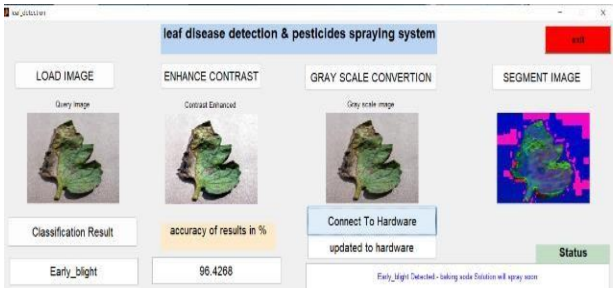


**Fig. 8.** Early Blight for Software Simulation

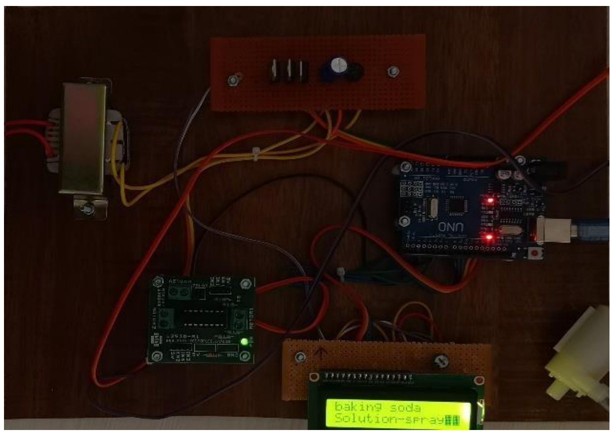


**Fig. 9.** Early Blight of Hardware Simulation

### 7.3 Healthy Leaf

Healthy Leaf: In the computer simulation, Fig. 10 shows a leaf without any signs of illness. several functions, including segment image, GLCM conversion, and enhanced contrast. Once the conversion is complete, the findings will be classified and their accuracy in percentage will be determined. Afterthat, it will link to the hardware to ensure that the leaf is unaffected by any diseases. If it determines that the leaf ishealthy, both motors will then be turned off. The result forHealthy Leaf is shown in Fig. 11.

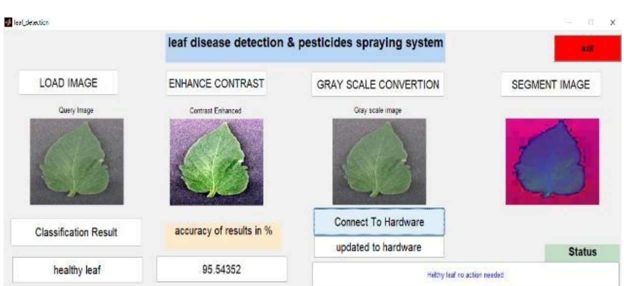


**Fig. 10.** Healthy Leaf of Software Simulation

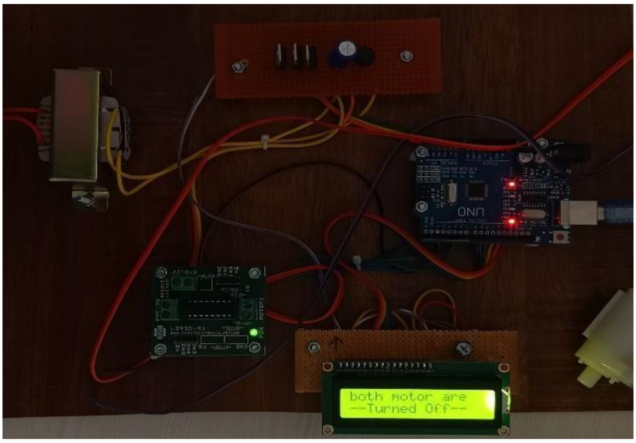


**Fig. 11.** Healthy Leaf of Hardware Simulation

## 8 CONCLUSIONS

In conclusion, our research provides a workable method for autonomous leaf disease identification based on machine learning. Using a sizable dataset of annotated leaf images and training a convolutional neural network toextract relevant characteristics, we were able to develop an algorithm that can successfully detect leaf diseases with high precision and recall. Our research suggests that the early detection and control of leaf diseases, as well as cropproduction and the use of harmful pesticides, could all beaccomplished with the help of this approach. However, there are still a few limitations and challenges that require future effort. For instance, since our research was focused on a specific subset of leaf diseases, it might be necessary to develop more specialized algorithms for different crops and geographical regions. The cost and availability of high-quality labeled datasets can also be a significantbarrier to the widespread implementation of machine learning for agricultural applications. In conclusion, we believe that our study lays the groundwork for future research and development of automated agricultural systems that can raise farming's sustainability andproductivity, and we hope that our findings will inspire inventive minds to come up with new solutions to the challenging issues that the agricultural sector is currently facing.

## References

1. M. V. Joshi and V. K. Yadav, Computers and Electronics in Agriculture, **188**, 1- 15, (2021).
2. V. K. Yadav, M. V. Joshi, and N. R. Pal, Plant Disease, **105**, 2052-2062 (2021).
3. Das, A. Sengupta, S. Sarkar, and S. Bandyopadhyay, Plant Disease, **104**, 2524-2534 (2020).
4. Y. Zhang, L. Li, and L. Hu, IEEE Access, **8**, 103250-103264 (2020).
5. N. S. Dang and M. H. Nguyen, IEEE Access, **9**, 33051-33069 (2021).
6. Al-Shannaq, M. Al-Omari, and N. D. H. Alsaraireh, Computers and Electronics in Agriculture, **187**, 1-10 (2021).
7. Y. Yuan, H. Xu, and Q. Wei, Computers and Electronics in Agriculture, **189**, 1-13(2021).

8. S. B. Saha and S. K. Singh, Computers and Electronics in Agriculture, **190**, 1-11(2021).
9. V. K. Yadav, M. V. Joshi, and N. R. Pal, Computers and Electronics in Agriculture, **185**, 1-13(2021).
10. M. Arsalan, W. A. Anjum, and S. M. Hassan, Journal of Plant Diseases and Protection, **128**, 465-485(2021).
11. J. Prasad, M. Kasiselvanathan, S. Lakshminarayanan, G. Sekar, Azath H Application of Machine Learning for Malicious Node Detection in IoT Networks, in proceedings of IEEE explore, IITCEE, 27-28 January 2023, Banglore (2023).
12. M. Kasiselvanathan, G. Sekar, J. Prasad, S. Lakshminarayanan, C. Sharanya, An IoT Based Agricultural Management Approach Using Machine Learning, in proceedings of IEEE explore, ICIDCA, 14-16 March 2023, Dehradun (2023).