# A REVIEW OF MACHINE LEARNING APPROACHES IN PLANT LEAF DISEASE DETECTION AND CLASSIFICATION

[#1]Dr.M.Rajeswari, Associate Professor, *Department of CSE, Karunya Institute of Technology and Sciences*,Coimbatore, India, rajeswari@karunya.edu.in

[#2]Darla G, [#3]Devagiri Mounika, UG Scholars, *Department of CSE, Karunya Institute of Technology and Sciences,* Coimbatore, India, darlag@karunya.edu.in, devagirimounika@karunya.edu.in

[#4] Dr. Krishnapriya K. S, Assistant Professor, Department of Computer Science, Valdosta State University, Valdosta, GA, USA, kkottakkalsugath@valdosta.edu

Corresponding Author: Dr.M.Rajeswari

*Abstract* — **Plants play an important role in everyone's lives, and they have a significant impact on a healthy environment. This paper describes a method for distinguishing plant leaf infection and a methodology for locating illnesses with caution. The objective is to analyze the sickness of leaves using various machine learning algorithms, such as Support Vector Machine (SVM) and the K-means algorithm. The plant's growth depends on the development of the leaves, which are affected by various illnesses that have been found in recent days. Researchers have started developing various algorithms for leaf disease detection and prevention. This research primarily focuses on the detection of leaf diseases in the following plant species: apple, cherry, grapes, orange, peach, potato, tomato, maize. These diseases include: apple-apple scab, black rot, and cedar apple rust; cherry-powdery mildew; maize-cercospora leaf spot, common rust and northern leaf blight; grape-black rot, esca and leaf blight; peach-black spot; potato-late blight; orange-haunglongbing; tomato-early blight, late blight, leaf mold, and mosaic virus. Its purpose is to categories the name of the leaf-affecting sickness. It is intended to classify the name of the disease that affects the leaf. Also, the percentage of damage due to the leaf disease will be predicted. Various performance metrics, including accuracy, precision, and recall, will be used in the implementation, which will be carried out via MATLAB.**

***Keywords****- Support Vector Machine (SVM), K-Means.*

## I. INTRODUCTION

Cattle farmers cannot control the climate since they cannot maintain control over most agricultural practices. If there are any illnesses that affect the land, they must be changed right away with immediate effect. Checking the leaves will reveal a significant percentage of a plant's illness or deficiency. Cattle farmers are used to monitoring the plants at distinct intervals, but if they do not see the symptoms of a disease, they may apply an incorrect quantity of pesticide or compost. Cattle farmers frequently lack the ability to discriminate between actual and false deficiencies, though. Due to the incorrect manure being used, the plants and the soil are eventually affected. Computerizing the process of diagnosing medical deficiencies is the solution to this problem. Various photo-handling techniques are often used to complete it.

### *A. Leaf Diseases*

With the growth of agricultural technology and the application of artificial intelligence to diagnose plant diseases, it is critical to conduct relevant research to develop sustainable agriculture. Numerous diseases, such as early blight and late blight, significantly impact quality and quantity, and manually interpreting these leaf diseases is time-consuming and difficult. Effective and automated diagnosis of these diseases in their initial stages can help reduce their severity since it takes a high degree of knowledge. Diseases have negative consequences for agricultural and plant areas. Microorganisms, genetic abnormalities, and infectious agents, including bacteria, fungi, and viruses, are the major causes of these illnesses.

### B. Classification

The ability to detect plant illnesses early and take prompt action to reduce crop loss makes the categorization of plant diseases a potent tool for farmer and agricultural researchers. SVM may be used to categorize diverse plant diseases based on characteristics including symptoms, underlying causes, and environmental variables. SVM is frequently used for classification jobs because it can group data into distinct classes based on characteristics gathered from pictures of ill plants. SVM improves classification accuracy by constructing a hyperplane that maximizes the margin between the two classes. In order to increase the effectiveness of plant disease detection systems, SVM can be used with additional methods like feature selection and data augmentation.

## II. LITERATURE REVIEW

SUNIL C. K., JAIDHAR C. D., et al. proposed that the EfficientNetV2 deep learning model is utilised in the suggested method for identifying illnesses in cardamom plants. The method entails gathering a sizable and varied dataset of photos of healthy and sick cardamom plants, preprocessing the dataset, dividing it into training, validation, and test sets, fine-tuning the pre-trained EfficientNetV2 model on the dataset of cardamom plants, assessing the model performance on the test set, and deploying the model in a production setting for autonomous disease detection. Although reliable and effective results are anticipated from the employment of EfficientNetV2, it is crucial to verify that the model is properly tuned and that the dataset is prepared.

Manish Kumar, Ahlad Kumar, et al. discussed how soil sensors, exploratory data analysis (EDA), and machine learning were used to create a system for forecasting plant illnesses. In addition to highlighting the value of early disease identification in agriculture, they go into how soil sensors may be used to gather current information on soil temperature and moisture. They next carry out an EDA to find trends and connections between environmental conditions and the prevalence of plant diseases. Based on the gathered data, a predictive model is created using machine learning techniques. The created system's potential to assist farmers in taking preventative actions and minimising crop losses is highlighted in the article.

HASSAN AMIN, ASHRAF DARWISH, et al. proposes a complete deep-learning model for identifying maize leaf diseases. A pre-processing module, a feature extraction module, and a classification module make up the model's three primary parts. The photos are improved and any noise is removed using the pre-processing module. A pre-trained convolutional neural network (CNN) that has been optimised using the dataset for maize leaf disease serves as the foundation for the feature extraction module. The input picture is categorised into one of the illness categories by the final classification module using a softmax function. The accuracy of the suggested model as tested by the authors on a dataset related to maize leaf disease was 97.43%.

Zhang, X., Qiao, et al. has proposed that the automated diagnosis and conclusion of maize leaf diseases are especially desired in the field of rural data. In this study, enhanced Google Net and Cifar10 models based on deep learning are suggested for the identification of maize leaf infections in order to increase ID accuracy and decrease the number of organizational boundaries. By altering the borders, pooling mixtures, adding dropout activities, correcting direct unit capacities, reducing the number of classifiers, and decreasing the number of classifiers, two better models are obtained and used to create and evaluate nine different types of maize leaf images.

STEFANIA BARBURICEANU, SERBAN MEZA, et al. proposed the use of Convolutional Neural Networks (CNNs) for texture feature extraction and their application in the classification of leaf diseases in precision agriculture. Four steps are included in the framework the authors suggest for classifying leaf diseases: picture capture, image pre-processing, feature extraction, and illness classification. Three distinct CNN architectures are compared in the study for feature extraction: VGG16, ResNet50, and InceptionV3. Using a dataset of 12,000 photos of leaves from four different disease classes—rust, powdery mildew, downy mildew, and healthy leaves—the authors assess the efficacy of these models. They evaluate the performance of the models using a variety of criteria, including accuracy, precision, recall, and F1-score. The results show that all three CNN models are able to achieve high accuracy in leaf disease classification. In particular, InceptionV3 outperforms the other models with an accuracy of 97.17%.

SABBIR AHMED, MD. BAKHTIAR HASAN, et al. proposed a deep neural network that is intended to be lighter and quicker than previous models for classifying tomato leaf diseases. The authors contend that conventional deep neural architectures are frequently too complex and computationally demanding, which causes longer inference times and more memory consumption. The authors suggest a novel design that makes use of a lightweight feature extraction module and a limited number of convolutional layers to overcome this problem. They also use methods like batch normalisation and dropout to enhance the model's functionality and avoid overfitting. Using a collection of photos of tomato leaves with five distinct forms of illnesses, the suggested design is assessed. According to the findings, the novel architecture beats a number of currently used architectures in terms of both accuracy and speed.

Wen-Liang Chen, Yi-Bing Lin, et al. proposed the creation of a rice blast detection system utilising Internet of Things (IoT) and Artificial Intelligence (AI) technology is covered in the article. A fungus called rice blast can severely lower rice harvests and result in financial losses. To reduce the harm this disease causes, early identification and immediate treatment are essential. IoT devices are used in the proposed system to gather real-time data on environmental factors including temperature, humidity, and light levels in rice fields. These devices send the data to a cloud server, where it is analysed by an AI programme to look for probable indications of a rice blast. The AI algorithm classifies the data and determines whether there is a pattern using a deep learning methodology, namely a convolutional neural network (CNN). The system can also provide farmers with alerts and recommendations for treatment based on the severity of the disease.

Dhaware, c. G., & Wanjale, k. H. et al. have proposed Agrarian creation is a feature on which our country's economy is heavily reliant. Picture-preparation processes are used in this manner to detect and recognise plant leaf wretchedness. Recognizing plant leaf sicknesses in connection with any programmed approach is useful since it decreases the massive effort of seeing in extensive lands and, at the initial stage, detects the indicators of sickness. Plant leaf sickness recognition and identification involve phases such as image securing, image pre-preparation, image division, highlight extraction, and characterization.

Siva K. Reddy, Gil Ben-Yashar, et al. proposed the feasibility of early detection of the Tomato Brown Rugose Fruit Virus (ToBRFV) in tomato plants using electrical measures. The objective of the study was to develop a quick and painless approach for the early diagnosis of the virus, which can cause major agricultural losses. The experimentation was done on both healthy and ToBRFV-infected tomato plants. Using a method known as Electrical Impedance Spectroscopy, they assessed the electrical potential difference between various plant sections, such as the stem and leaves (EIS). In order to detect any changes in the electrical characteristics of the plant, EIS measurements were made at various phases of the infection. According to the study, diseased and healthy plants had significantly different electrical characteristics. The electrical impedance decreased and the phase angle increased in the infected plants. Five days after infection, which is significantly earlier than the onset of the virus' symptoms, alterations in the electrical characteristics were already apparent. The study's findings support the possibility of the EIS methodology as a trustworthy, non-destructive tool for ToBRFV early detection in tomato plants.

CHANGJIAN ZHOU, ZHIYAO ZHANG, et al. proposed an approach to identify grape leaf spot, a common disease in grape plants, using limited samples. They stated that several labelled data sets, which can be time and money consuming to acquire, are frequently needed for traditional techniques of illness detection. The authors suggest a fine-grained GAN (Generative Adversarial Network) model to get around this problem and produce realistic leaf pictures from a limited number of input samples. This model is used to create new artificial pictures after being trained on a limited dataset of annotated photos. The original dataset and these artificial pictures are then blended to provide a bigger and more varied training set, which can enhance the performance of the illness classification model. Using a collection of photographs of grape leaves, the authors assessed the suggested method and contrasted it with a number of other cutting-edge approaches. The results showed that their approach outperformed other methods, achieving an accuracy of 94.8%.

## III. PROPOSED METHODOLOGY

The proposed framework is divided into two components: advanced picture evaluation and component extraction from test plant leaves. To carry out the backpropagation, the proposed study aims to examine the condition of the leaf using image processing and Support Vector Machine (SVM). Infections on the leaf are a fundamental problem that causes a dramatic decline in the production of the leaf and related neurons. The most hidden pixel is the lightest pixel in each of the sets of neurons that are arranged in ascending order. The suggested structure included a provision to at least acknowledge the type of relative plant-keeping error. SVM is designed to function, as demonstrated by its ability to learn from examples, provide requirements for results, and recognize designs similarly.

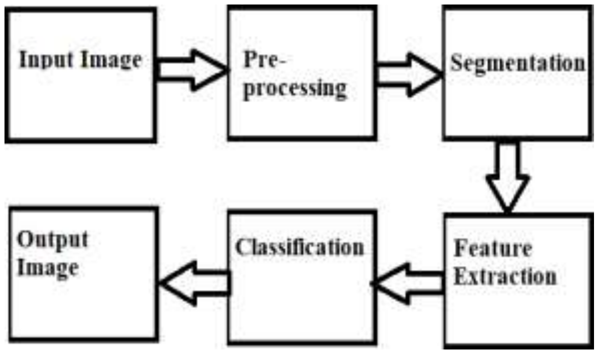



Figure 1. Overall Architecture

### A. *Dataset Collection*

The dataset was collected from the Kaggle website. The new plant disease dataset consists of several healthy and diseased crop leave in the form of images. This dataset was rebuilt via offline augmentation from the original one. About 87,000 RGB photos of healthy and sick crop leaves comprise this collection, divided into 38 classifications. The directory structure is preserved when the training and validation sets are split up into training and validation sets in an 80/20 ratio. Later, 33 test photos are made in a new directory for prediction purposes.

| PLANT | PLANT DISEASE | NO. OF IMAGES |
|---|---|---|
| Apple | Apple scab | 2016 |
| | Black rot | 1987 |
| | Cedar apple rust | 1760 |
| | Healthy | 2008 |
| Blueberry | Healthy | 1816 |
| Cherry | Powdery mildew | 1683 |
| | Healthy | 1826 |
| Maize | Cercospora leaf spot | 1642 |
| | Common rust | 1907 |
| | Healthy | 1859 |
| | Northern leaf blight | 1908 |
| Grape | Black rot | 1888 |
| | Esca | 1920 |
| | Healthy | 1692 |
| | Leaf blight | 1722 |
| Orange | Haunglongbing | 2010 |
| Peach | Bacterial spot | 1838 |
| | Healthy | 1728 |
| Bell | Bacterial spot | 1913 |
| pepper | Healthy | 1988 |
| Potato | Early blight | 1939 |
| | Healthy | 1824 |
| | Late blight | 1939 |
| Raspberry | Healthy | 1781 |
| Soyabean | Healthy | 2022 |
| Squash | Powdery mildew | 1736 |
| Strawberry | Leaf scorch | 1774 |
| | Healthy | 1824 |
| Tomato | Early blight | 1920 |
| | Bacterial spot | 1702 |
| | Healthy | 1926 |
| | Late blight | 1851 |
| | Leaf mold | 1882 |
| | Septoria leaf spot | 1745 |
| | Spider mite | 1741 |
| | Target spot | 1827 |
| | Mosaic virus | 1790 |
| | Yellow leaf curl virus | 1961 |

Table 1. Dataset Specifications

### B. *Preprocessing*

The phrase "image pre-processing" refers to actions taken on images at the most basic level of abstraction with the intention of increasing or suppressing undesired distortions or other image characteristics important for tasks involving further processing and analysis. The photos aren't given any further picture information. Its techniques make use of significant picture redundancy. The brightness value of adjacent pixels that represent the same real item is the same or close to it. If a deformed pixel in the image can be detected, it can be recovered as the average value of itsneighboring pixels.

### C. *Input image*

Images of the plant leaves were subjected to a number of bacterial and parasite-influenced tests. A Sony computerized shading camera is used in the field to capture images ofleaves in the shade. Randomly chosen from the crop field, the leaves used for handling are then photographed under uncontrolled illumination. The image is preprocessed, which comprises scaling it into 256 x 256 aspects, prior to division.

### D. *K-Means Clustering*

To investigate and interpret picture segmentation, K-means clustering is performed. A cluster is a group of items that are both similar and distinctive from other clusters. Clustering is the technique of classifying things into several groups such that the data in each subset can show a common component based on a predefined distance metric. Shapes, textures, and any other data that can be gleaned from a picture itself may be used to classify it. K-means groups the data into a preset number of clusters. First, random calculations are done to find the centroids of chosen clusters.

### *E. Feature Extraction*

This step, known as feature extraction, is important because it makes use of algorithms and techniques to identify and isolate particular desired sections or forms within a picture. The input data will be condensed to a smaller set of attributes for representation if it is too large to analyse or is deemed to be significantly redundant. The core attributes of a feature are its area, perimeter, and eccentricity.

### *F. Classification*

Support vector machines, which analyze data used for characterization and relapse investigation, are directed learning models with associated learning computations. In this method, the photos were arranged according to SVM calculations, and their presentational proportions were established. A Support vector machine, which can provide classifier tasks extremely fast following a preparation phase, is an essential tool for twofold characterization. There are several approaches to handling the adoption of SVMs to order problems with at least three classes. Support vector machines analyse data used for grouping and relapse investigation in artificial intelligence. They have controlled learning models with related learning computations. SVM classifiers are by nature two-class classifiers. It is customary to use one of the approaches when using SVMs to do multiclass order.

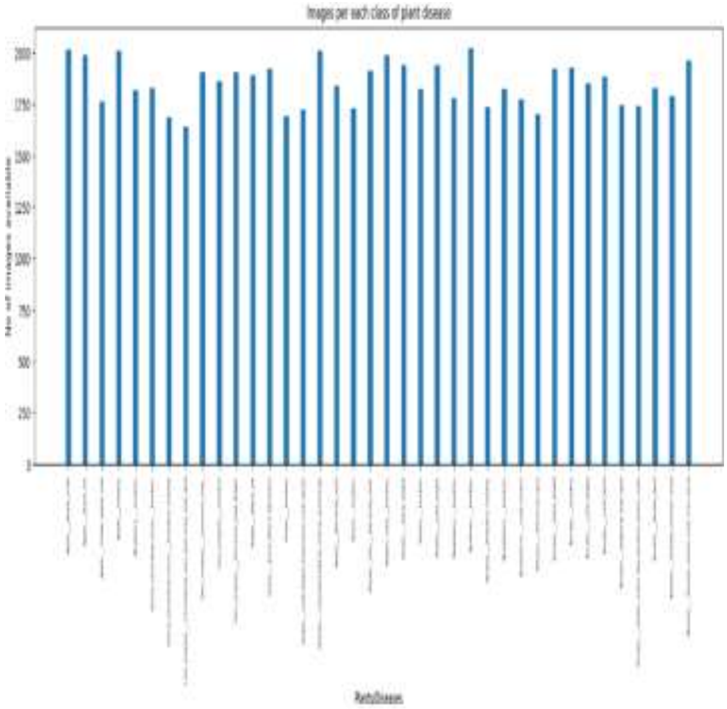
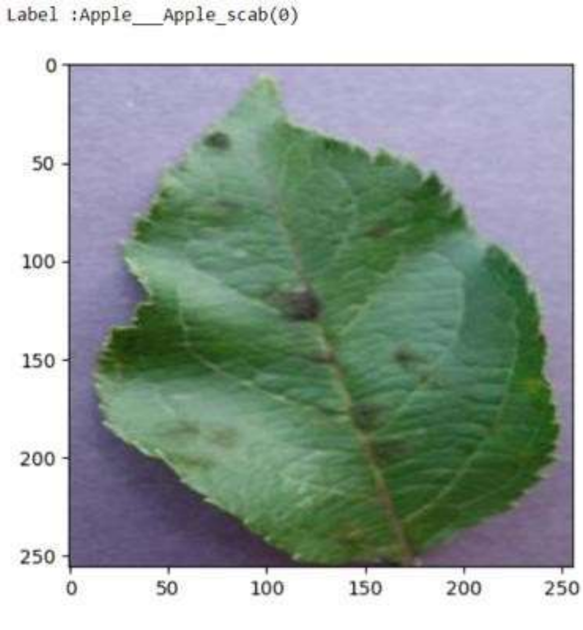


Fig 2. Images per each class of plant disease

## IV. EXPERIMENTAL RESULT

It is necessary to establish a hyperplane in the middle of the informative indexes in order to designate which class the support vector machine belongs to. The element vector provides input to the classifier. Testing and preparing vectors make up the element vectors of the information base pictures. The classifier utilises the preparation set as training data for arranging the testing set. The predicted names and actual attributes are compared to determine the classifier's presentation. Images in the RGB colour space are separated using the K- Means approach. Three groupings—the split-up photographs and the sample input photos—have been chosen as the number.



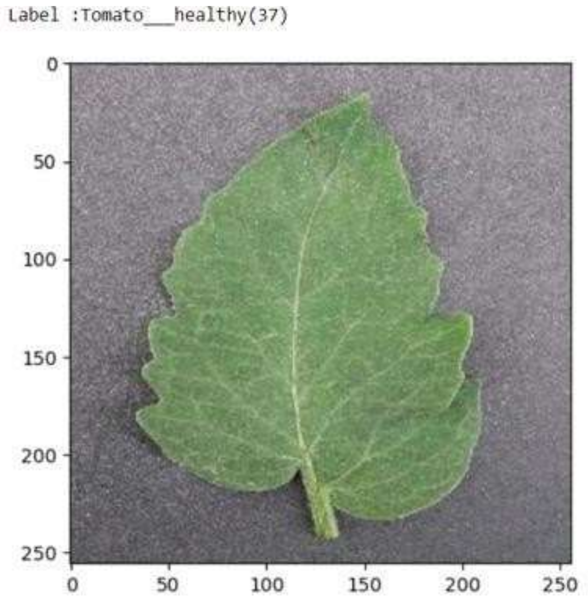



Fig 3,4. Images and outputs generated

## V.CONCLUSION

In conclusion, SVM and K-means algorithms have produced encouraging results in the identification and classification of plant diseases. While SVM has been demonstrated to be a potent classification method that can precisely identify between healthy and unhealthy plants, K-means clustering has been found to be effective at bringing together similar plant photos. Researchers have created efficient disease detection systems that can accurately identify and categorise plant illnesses by integrating these two methods. The quantity and quality of the data used to train the algorithms, however, have a significant impact on the effectiveness of these systems. Future research is needed to organize the numerous plant diseases and improve the precision of characterization. Development in this area could lead to more sophisticated and accurate plant disease detection systems that can help farmers quickly identify and respond to plant diseases, ultimately leading to higher crop yields and more sustainable agriculture. This project uses SVM calculations, another kind of K-implies computation, to classify leaf diseases.

## REFERENCES

[1] Cardamom Plant Disease Detection Approach Using EfficientNetV2
SUNIL C. K., (Student Member, IEEE), JAIDHAR C. D., AND NAGAMMA PATIL

[2] Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning
Manish Kumar, Ahlad Kumar, and Vinay S. Palaparthy

[3] End-to-End Deep Learning Model for Corn Leaf Disease Classification
HASSAN AMIN , ASHRAF DARWISH, (Member, IEEE), ABOUL ELLA HASSANIEN, AND MONA SOLIMAN

[4] Zhang, X., Qiao, Y., Meng, F., Fan, C., & Zhang, M. (2018)"Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks", IEEE Access

[5] Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision Agriculture
STEFANIA BARBURICEANU, SERBAN MEZA, (Member, IEEE), BOGDAN ORZA, RAUL MALUTAN, AND ROMULUS TEREBES, (Member, IEEE)

[6] Less Is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification SABBIR AHMED, MD. BAKHTIAR HASAN, TASNIM AHMED, MD. REDWAN KARIM SONY, AND MD. HASANUL KABIR , (Member, IEEE)

[7] RiceTalk: Rice Blast Detection Using Internet of Things and Artificial Intelligence Technologies
Wen-Liang Chen, Yi-Bing Lin, Fellow, IEEE, Fung-Ling Ng, Chun-You Liu, and Yun-Wei Lin

[8] Dhaware, C. G., & Wanjale, K. H. (2017)"A modern approach for plant leaf disease classification which depends on leaf image processing", International Conference on Computer Communication and Informatics (ICCCI).

[9] Early Sensing of Tomato Brown Rugose Fruit Virus in Tomato Plants via Electrical Measurements
Siva K. Reddy, Gil Ben-Yashar, Yarden Mazal Jahn, Yahav Ben-Shimon, Aviv Dombrovsky, Avital Bechar, and Assaf Ya'akobovitz

[10] Grape Leaf Spot Identification Under Limited Samples by Fine Grained-GAN
CHANGJIAN ZHOU, ZHIYAO ZHANG, SIHAN ZHOU, JINGE XING, QIUFENG WU, AND JIA SONG