Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v29i02.7165

AN EXTENSIVE ANALYSIS OF MACHINE LEARNING METHODS FOR IDENTIFYING PLANT LEAF DISEASES

Deepak Kumar Awasthi[1*], Dr. Arvind Kumar Shukla[2], Ashwani Gupta[3].

[1*,2]School of Computer Science and Application, IFTM University, Moradabad, Uttar Pradesh, India
[3]Assistant Professor, Department of CSIT, MJP Rohilkhand University, Bareilly, Uttar Pradesh -243006, India, Email: posttodeepak@gmail.com.

***Corresponding Author:** Deepak Kumar Awasthi
*School of Computer Science and Application, IFTM University, Moradabad, Uttar Pradesh, India

Abstract
India's economy is dependent on agriculture. Maintaining strong crop yields for food, medicine, and commercial uses is essential as the world's second-largest population. Applications based on IT are frequently utilised for disease identification. By analysing images of different plant sections, data science-based computer vision systems are incredibly effective at detecting diseases in their early phases. It takes a lot of human skill to diagnose the condition by eye inspection, which is a difficult task in and of itself. The disease diagnosis for supervised machine learning approaches for leaf images is critically reviewed in this work. The application of supervised machine learning as a general concept is described. Based on the symptoms of a disease extracted in the form of features, a disease in plants can be identified. Thus, feature extraction methods are crucial in these systems. There is extensive discussion of the Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) approaches, as well as a brief discussion of relevant recent works are presented. It offers a thorough analysis of various visual characteristics for various illnesses in different atmospheric conditions.

1.Introduction
India is an agricultural country. 70 % of the population of India is directly connected to the villages. The main source of income is agriculture[1]. According to the Economic Survey 2020-2021, the contribution of agriculture to GDP has increased to approximately 20 % for the first time in the last 17 years, making it one of the major sector showing prospects for GDP growth in 2020–21[2]. Agriculture has a huge contribution to making the country economically prosperous due to the large productive land area; crop security has a direct impact on the country and its economic condition. Another important aspect is that making them healthy. It becomes a major challenge in view of the increasing population.

Due to increasing awareness in herbal therapy and its lack of adverse side effects compared to allopathic treatment, as well as the need to provide for the medical needs of an expanding human population, medicinal plants have regained a widespread reputation during the COVID-19 pandemic. One of the main reasons for this growth is the increasing popularity of ayurvedic medicines in recent years; especially after the pandemic COVID-19 it attracted the attention of farmers toward the cultivation of therapeutic herbs. The most well-liked and profitable agricultural venture for farmers in India is currently the commercial production of medicinal herbs.

Plant leaves diseases directly affect plant growth and crop production. Often many of our farmers use chemical insecticides without proper identification of plant diseases, due to which they not only inadvertently damage the crop but also destroy the productive capacity of the field. Generally, main causes of these plant diseases are microorganism, bacteria, fungi, and viruses. Among them fungi and bacteria are mainly responsible for plant leaf diseases.

In this paper, the idea of integration of Information and Communications Technology (ICT) with the agriculture sector, motivates the development of an automated system for herb plant leaf disease classification. This survey offers a thorough analysis of different supervised machine-learning classification methods used in plant disease detection. Section 2 of this paper, presents a brief discussion of the traditional approaches, that our farmers have been using for the past several years to identify plant diseases. Section 3 of the paper presents the generalized mechanism for plant leaf classification of a conventional plant leaf disease detection system. The common phases of the mechanism are also summarized in this section. In Section 4 various classification techniques explored in plant disease detection have been given and reviewed. Section 5 provides a summary and discussion of classification techniques and their benefits. Additionally, Section 6 of this paper concludes with future direction. The most common diseases that affect medicated plant leaves are categorized as leaf spot, powdery mildew, leaf curl(Copper deficiency), and leaf rust. The cause and symptoms for each of these diseases are described below.

The disease can affect any section of the blade and appears as circular to irregular patches that are occasionally divided by veins. The upper surface of the infected leaves shows distinct copper-brown markings. When there is high humidity and high temperature in the air, the leaf spot develops from July to November throughout India. The powdery mildew disease appears on the ventral surface of the leaves, white powdery patches first develop, and then distinct dot-like structures start to appear. In the early stage of the disease, it shows chlorotic symptoms on the ventral surface of leaves with curved lamina. Affected leaves become yellow and drop off too soon. Arjun's *(Terminalia arjuna)* crop loss is estimated to be between 25% to 30%. Copper deficiency is the main cause of leaf curl disease. Due to the leaflet's midrib-based folding, leaves have a boat-like shape[26].

Various modern technologies have emerged to protect the crop from diseases, eliminate infections and maximize productivity of crops. Most of the technologies used so far are laboratory-based. Machine learning technology in computer science and IT sector has played a vital role in the timely detection of various diseases of plants. Apart from this, the study of machine learning algorithms is being used in many agricultural works in the modern world.[10-11][60-61].

For this purpose plant leaves were captured using various cameras and also collect from different available on-line repositories. The information on the common disease in herb leaves and the Medicinal use of these plant leaves are depicted in Table 1.

Plant leaf / (Botanical Name)	**Availability Region in India**	**Diseases**	**Medicinal Use**
Arjun *(Terminalia arjuna)*	Uttar Pradesh, Madhya Pradesh, and some parts of south and central India	Leaf spot Powdery mildew Leaf Curl	Cardio-tonic cases [51]
Chinar *(Platanus)*	Uttarakhand and Kashmir	Leaf Scorch Leaf blight leaf spot	Dysentery, toothache[54]
Jamun *(Syzygiumcumini)*	TamilNadu, Maharashtra, Andhra Pradesh, Bihar, and Karnataka.	leafspot fruit rots	Diabetes [53]
Jatropha *(Jatropha Curcas)*	Rajasthan	Powdery mildew Passalora leaf spot	Febrile diseases, Jaundice[55]

Lemon *(Citrus limon)*	Andhra Pradesh, Maharashtra, Gujarat, Odisha and Tamil Nadu	Bacterial blast, Citrus nematode. Dothiorella blight	Coughs[52]
Pongamia Pinnata *(Millettia Pinnata)*	Rajasthan	Leaf spot and blight, Leaf Rust Powdery mildew	Piles, skin diseases, wounds, and ulcers[62]
Pomegranate *(Punica Granatum)*	Maharashtra, Gujarat, Rajasthan, Karnataka, Uttar Pradesh, Punjab and Haryana	Leaf spot	Sore throats, coughs, urinary infections, digestive disorders, skin disorders, arthritis [63]

{https://data.mendeley.com/datasets/hb74ynkjcn/5}

Shri Mata Vaishno Devi University, Jammu, India, provided the image repository of all the above leaves. This image collection procedure was completed in the period months of March to May. The images are taken in an enclosed space. The photographs were taken with an 18-55mm lens, sRGB color representation, 24-bit depth, two resolution units, 1000-ISO, and no flash. Later these pictures were saved in *jpg* file format. Instances of healthy and unhealthy images of above leaves are showen in following figure 1 to figure 10.

Figure 1 Arjun Healthy Leaf.

Figure 2 Arjun Unhealthy Leaf.

Figure 3 Jamun Healthy Leaf.

Figure 4 Jamun Unhealthy Leaf.

Figure 5 Jatropha Healthy Leaf.

Figure 6 Jatropha Unhealthy Leaf.

Figure 7. Lemon Healthy Leaf.

Figure 8. Lemon Unhealthy Leaf.


Figure 9. Pomegranate Healthy Leaf. Figure 10 . Pomegranate Unhealthy Leaf.

2. Traditional Detection Approaches

A set of conventional methods have been developed to detect plant leaf diseases and to avoid crop loss. These methods can be categorized as A) Direct method and B) Indirect method.

A. Direct Method

The direct detection method of plant diseases is based on molecular and serological methods that require a large number of samples to be studied. Bacterial, fungal, and viral problems in plants can be identified using these techniques. Major direct detection methods are –

Fluorescence In-situ Hybridization Method (FISH)

FISH is a molecular detection technique it is a combined approach based on microscopy and plant DNA hy-bridization to detect bacteria[3]. FISH could also be used to detect fungi and viruses and other bacteria that infect plants[4]. FISH also be used to investigate complex microbial communities[5]. Auto-fluorescence materials are a common problem that arise a big question about their reliability[5].

Polymerase Chain Reaction Method (PCR)

Kohler JF, Milstein C, and Mullis were the first to apply PCR technology for the production of monoclonal antibodies and the amplification of nucleic acid sequences. PCR was initially used to diagnose diseases brought on by bacteria and viruses with extreme accuracy[6]. It is now frequently used to identify plant diseases. Reverse-transcription PCR (RT-PCR) is a more sophisticated PCR technique that is developed and introduced recently. Due to its great sensitivity, RT-PCR is utilised to identify plant pathogens[7]. Real-time PCR devices have also been utilised for on-site, quick diagnosis of plant diseases based on bacterial, fungal, and viral nucleic acids[8][9]. Due to the accuracy of DNA amplification, the PCR technique can offer great sensitivity and specificity, but it is constrained by a lack of operational robustness[10].

PCR performance is influenced by inhibitors found in the sample test, polymerase activity, and the effectiveness of DNA extraction.

Enzyme-Linked Immunosorbent Assay Method (ELISA)

A molecular method for identifying plant diseases based on antibodies and colour changes is the enzyme-linked immunosorbent test (ELISA)[11]. In this technique, antibodies are conjugated to an enzyme and used to effectively bind the target antigens from the viruses, bacteria, and fungi. The interaction between the substrate and the mounted enzyme causes colour changes that can be used to view the detection. The performance of ELISA can be improved greatly with the application of specific monoclonal and recombinant antibodies which are commercially available[12][13]. It is only helpful for verifying plant illnesses once visual symptoms manifest; it is a less reliable method for early disease identification before symptoms manifest[7].

Flow Cytometry Method (FCM)

A common method for cell counting and sorting, finding biomarkers, and creating new proteins is the flow cytometry method (FCM). It is an optical direct detection method based on a laser.

Multiple parameters test capacity simultaneously is a benefit of this technology. FCM is still a relatively new approach for plant disease detection even though it has mostly been used to research cell cycle rates and antibiotic sensitivity, enumerate bacteria, distinguish between viable and non-viable bacteria, and describe bacterial DNA and fungal spores[14].

B. Indirect Detection Method

An Indirect detection methodology employs many parameters, such as changes in temperature and environment, as well as morphological changes that happen due to pollution.

Fluorescence Imaging Method

In this method, the fluorescence of the chlorophyll on the leaves is measured as a function of the light that strikes them, and changes in the fluorescence characteristics can be used to investigate pathogen infections[15][16]. This method allowed for the exact detection of leaf rust and powdery mildew infections in plant leaves by analysing the temporal and spatial changes in chlorophyll fluorescence[16]. Although fluorescence measurement offers sensitive identification of anomalies in photosynthesis, the practical applicability of this method in a field environment is limited[17][18][19].

Hyper-spectral Techniques Based Method

Over a broad spectrum, hyper-spectral imaging can be utilised to gather insightful data about plant health. In large-scale agriculture, plant disease identification is done using hyper-spectral imaging. The method offers a quick examination of the imaging data and is quite robust.

By observing the variations in reflectance based on the biophysical and biochemical characteristic changes during infection, hyper-spectral techniques are utilized to detect plant diseases. Phytophthora infestans, Venturia inaequalis, and Magnaporthe grisea infections of rice, tomato and apple trees have all been found and reported using hyper-spectral imaging techniques[20][21][22].

Thermography Method

This method visualizes the variations in leaf surface temperatures. The colour difference of the emitted infrared radiation can be examined using thermographic cameras. The resulting disease can be predicted through thermographic imaging and the amount of water that transpired can be determined, without the external temperature influences[23]. Several research groups have reported the temperature fluctuations brought on by plant pathogen infection[23][24][25].

Thermography has a high sensitivity to changes in the environment during measurements, which limits its practical application for illness monitoring. As a result, it is impossible to determine the type of infection or differentiate between diseases that produce identical thermo-graphic patterns using thermo-graphic detection since it lacks specificity toward diseases.

Merits and Demerits of methods

Method	**Merits**	**Demerits**
PCR	PCR enables highly sensitive pathogen detection in a single experiment for one or more diseases[7].	Irrelevant for routine application.
ELISA	It is useful only for the confirmation of plant diseases after visual symptoms appear	Irrelevant in the early stage of diseases.
FCM	capability for simultaneous measurement of several parameters. Separate germs that are alive from those that are not FCM is used to characterize the genomic sizes of populations of fungi and oomycetes[49].	It is a very expensive method and requires highly skilled operators[49].

Fluorescence Imaging	Used to find infections of powdery mildew and leaf rust on plant leaves.	Field setting is limited for practical application.
Hyperspectral Techniques	For fieldwork, this technique might be more beneficial.[48]	Hyper-spectral cameras are still expensive, making them difficult to be widely applied in agriculture[48]. Hyper-spectral image data takes a lengthy time in acquire, analyzeprocess[48].
Thermography	This method works in all situations when analysing the area or object requires consideration of the temperature difference[50].	Dependence on changes in the environment during measurements. Unable to differentiate between diseases that show similar thermo-graphic patterns.

A review of traditional approaches

All of the aforementioned techniques are laboratory-based techniques. They require a setup for sample preparation. The absence of consistency and power supply makes these tests highly challenging to apply. Apart from that, it costs a significant amount to carry out these tests and techniques since they need expensive reagents, sophisticated labs, and well-trained, professional personnel. So these above methods are beyond the reach of small farmers in far-flung areas.

Even now, farmers attempt to identify plant diseases by using way of their naked eyes, based on their earlier farming experience. However, it is challenging to visually scan the plant disease symptoms due to incomplete soil knowledge, crop variety, unexpected environmental changes, and virus strains. That’s why many scientists and researchers are using computer science-based technology i.e. AI-based concepts, i.e. machine learning, etc.

3. A Generic plant leaf Disease Classification mechanism

The field of plants has substantially benefited from technologies like artificial intelligence and image processing. High-quality pictures are a need for these procedures[21]. By employing a high-resolution camera to capture or gather pictures of plant leaves, a data set is produced. The photographs that were taken might show the influence of environmental factors, lighting, equipment capabilities and other natural conditions. Pre-processing is therefore required before use. During the pre-processing operation, essential tasks like background removal, color space conversion, contrast stretching, and scaling are determined (successfully completed). Feature extraction plays an important role to differentiate healthy leaves and unhealthy leaves or to identify the actual disease as leaves We can identify diseases in the leaves or classify the leaves as healthy or unhealthy based on a variety of characteristics, such as color, texture, and shape. The precision of classification has a significant impact on a system’s effectiveness. In this work, the thorough descriptions of the main machine learning algorithms for categorization proposed in the study are explored and compiled. A generic setup of the system for detecting and classifying plant diseases is depicted in Figure 1.

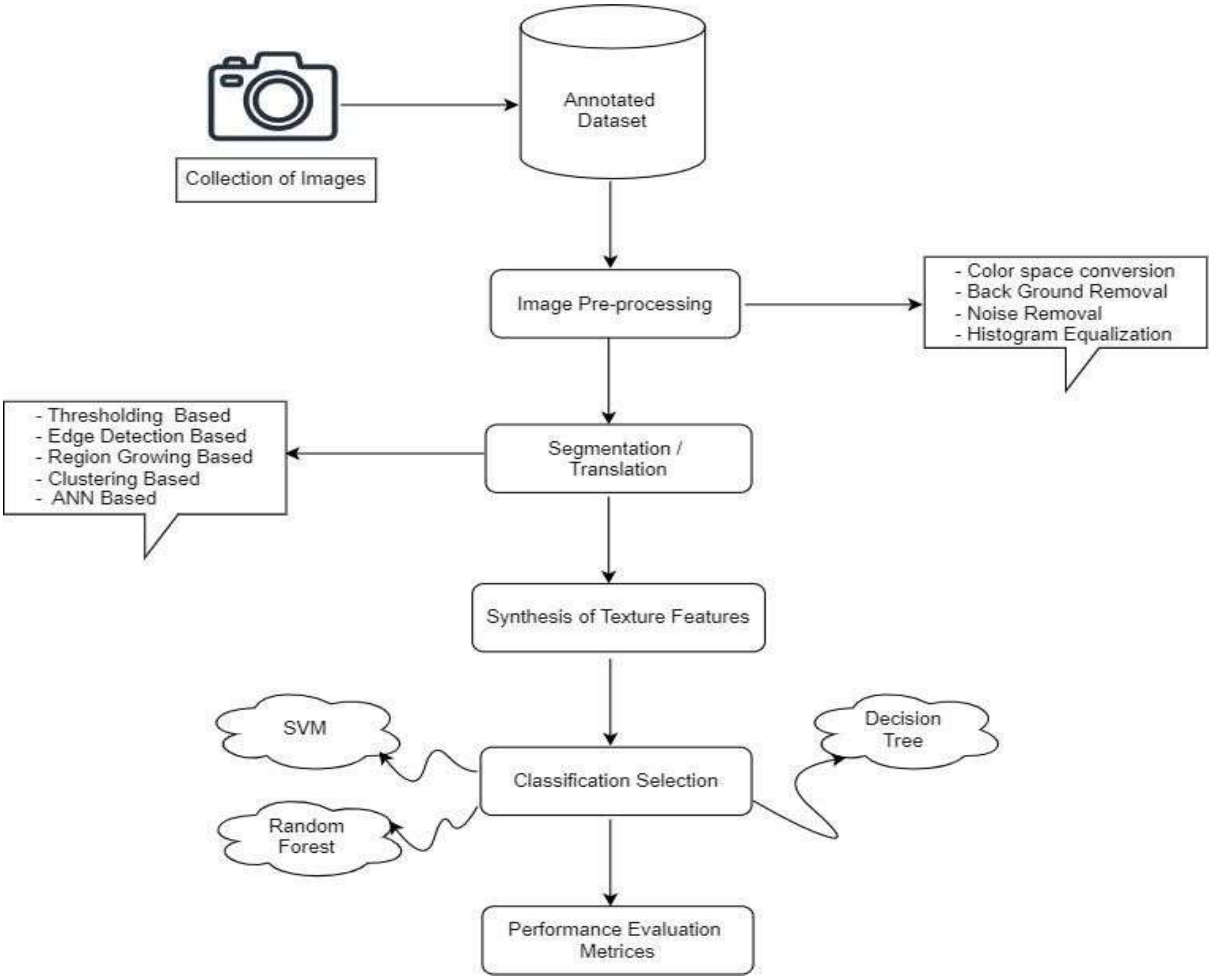

Figure 1 : A generic system for plant leaf diseases Identification.

3.1 Image Collection

The primary stage of the plant disease detection system is image collection. The phase is important since the efficacy of the disease detection system depends on the quality of the images that are collected, regardless of whether they are collected directly from the field or from an archived database. The imaging techniques and tools utilized, such as the digital camera, scanners, or drones have a direct impact on the image quality. Generally, researchers worked with opendatasets like Plant-Village {https://www.kaggle.com/datasets/emmarex/plantdisease}, Shri Mata Vaishno Devi University, Jammu, India {https://data.mendeley.com/datasets/hb74ynkjcn/5}, controlled laboratory or real-time circumstances, or self-generated datasets of images of plants in crop fields. The real-time image-gathering process creates complicated backgrounds, imbalanced lighting, and illumination effects, all of which enhance the complexity of the imaging system that affect the detection system efficiency. During the collection of real-time images, it requires more awareness of device information like- resolution and orientation of camera.

3.2 Annotated Dataset

Annotated Data plays a key component role in supervised machine learning solutions. Labeled datasets are essential for supervised machine learning since machine learning models require understanding input patterns in order to interpret them and produce effective outputs. So In this stage here we label and identify the images as healthy or unhealthy.

3.3 Image Pre-Processing

Gathered images may contain undesired information in form of shadows, noise, undefined distortion, and complex backdrops. These factors have nothing to do with the original work's aim but have a significant impact on the accuracy of the outcome. Preprocessing reduces the impact of this operation. Researchers can implement the image preprocessing process such as colour space conversion, histogram equalization, contrast enhancement, cropping, noise reduction, and smoothing as needed.

The conversion of colour spaces is used to improve outcomes while using a particular colour space and to satisfy important visual attributes. There are many widely used colour spaces accessible, including HSV, RGB, HSI, L*u*v*, L*a*b*, CIELuv, and YCbCr.

- Although the RGB color space is based on the device and acceptable for display systems, its application is limited by strong channel correlation.[27].
- For the purpose of identifying and classifying plant diseases, the HSV color model is frequently employed. Compared to green plants, most background pixels are removed according to their colorvalues[28].
- The L*a*b* color model improves the result of Image classification. Its visual perception is analogous to that of the human eye according to the specific sensitivity of the three types of cone cells in the human eye[29].

A significant challenge is removing noise from real-time acquired images. Therefore, noise removal is one of the regular operations associated with plant disease detection methods. There are various noise removal techniques like mean filters[30], median filters[31], and Gaussian smoothing[32] used by researchers. The contrast of leaf images in plant disease detection systems is improved by histogram equalization[66].

Image sharpening increased focus and placed emphasis on the pattern of the image. In research on plant disease identification, the Laplacian filter is deployed for image sharpening[33].

3.4 Segmentation / Translation

Image segmentation reduces the complexity of the image to make further analysis.The aim of segmentation is to simplify an image and make it more meaningful. In this process researchers can define borders, draw lines, and distinguish the most necessary objects in an image from the other, not necessary objects. There are several segmentation techniques like edge detection-based, region growing-based, clustering based segmentation, and thresholding-based segmentation methods are available.

This edge detection-based method seeks to identify leaf edges inside an image. In these techniques, we identified the pixels that are concerned with the boundary pixels of an object. Sobel, Canny, Laplacian, and fuzzy logic are a few of the techniques used for edge detection.

In the Region-Based Segmentation technique, an image is divided into various parts that are the same type. Regions of the same type were created by combining pixels of the same type[35].

Clustering is a technique for segmentation in which an image is first transformed into a histogram and then clustering is applied to it. K-means, one of the fundamental clustering algorithms, is designed for segmenting textured images[35]. When many diseases' symptoms are present in leaf photos, K-means clustering is shown to be more efficient and appropriate than edge-based and thresholding-based techniques[36].

Thresholding is a method for generating binary images from a grayscale image. The benefit of getting a binary image is that it makes the process of recognition and classification easier and decreases the complexity of the data[34]. Edge-based and threshold-based segmentation are the popular traditional approaches used in the applications of plant disease detection.

3.5 Synthesis of Texture Features

Since ancient times, at first sight, diseases in plants are estimated from their changing nature. Characteristics like size, color and texture of plant leaves are mainly used. Feature extraction allows for the independent description of these attributes. Texture feature extraction is the process of gathering all the features through texture analysis. When a small portion of an image has a lot of color variation, the texture is the effective attribute for that area. The factors that affect how the texture is seen are light, contrast, distance, and direction. Entropy, contrast, skewness, variance, homogeneity, and other factors can all be used to describe the texture of an image. These features are shown as follows[37][38] –

Contrast- It is determined by measuring the intensity of each pixel and its surrounding pixels in an image. It is the difference in the image's color and brightness.

$$\text{Contrast} = \sum_{i,j=0}^{N-1} p(i,j)(i-j)^2 \tag{1}$$

Energy - It is described as a measurement of the degree of repeated pixel pairs. Energy also measures the disorders in textures.

$$\text{Energy (E)} = \sum_{i,j=0}^{N-1} p(ij)^2 \tag{2}$$

Entropy - The disorder or complexity of an image is measured by entropy. If the image's texture is not uniform, the entropy is high.

$$\text{Entropy (ENT)} = -\sum_{i,j}^{N-1} p(i,j)\ln[p(i,j)] \tag{3}$$

Skewness - It depicts the asymmetries in the probability distribution of a random variable with a real value.

$$\text{Skewness } (\mu)_3 = \sum_{i=0}^{L-1} (Z_i - m)^3 p(Z_i) \tag{4}$$

Smoothness - The reason for the smoothness of an image is described by the Inverse Difference Moment(IDM). Identical pixels of an image represent a higher value of IDM.

$$\text{Smoothness (S)} = 1\text{- } (1 / 1+ \sigma^2) \tag{5}$$

Homogeneity - It is also known as Inverse Difference Moment. It Measures visual homogeneity by assuming bigger values for minor changes in grey tone between adjacent elements. It contains a maximum value when all elements in the image are the same.

$$\text{Homogeneity (HOM)} = \sum_{i,j=0}^{N-1} \frac{1}{1+(i-j)2}\, p_{i,j} \tag{6}$$

Mean - It represents the pixel's average intensity values.

$$\text{Mean (m)} = \sum_{i=0}^{L-1} Z_i\ p\,(Z_i) \tag{7}$$

Variance - It displays the distribution values around the mean. When the values of the grey levels vary from their mean, variance increases.

$$\text{Variance (Var)} = \sum i \sum j\ (1-\mu)^2\ Pij \tag{8}$$

Where μ is the mean of p_{ij}

Abbreviations: In the above equations considering a two-dimensional Image, I (i,j) represent the pixel position at row i and column j respectively.

3.6 Classification Selection

3.6.1 Support Vector Machine (SVM)

A support vector machine(SVM) constructs a hyper-plane in a high-dimensional space that effectively separates the data points, which can be used for classification as well as regression problems. Figure 2, depicts a diagram for a multi-plane. It is used in case of finite sample data. It aims at acquiring worthy solutions on the ground of present data rather than the optimal value for infinite samples. Numerous potential hyper-planes might be selected to divide the two classes of data points. A plane with the largest margin between data points from both classes is chosen. Maximizing the margin distance provides greater assurance, with which future data points can be classified more effectively. Data points closer to the hyper-plane, known as support vectors, have an impact on the hyper-plane's position and orientation. By utilizing these support vectors, the classifier's margin is, enhances These are the concepts that support the creation of our SVM.

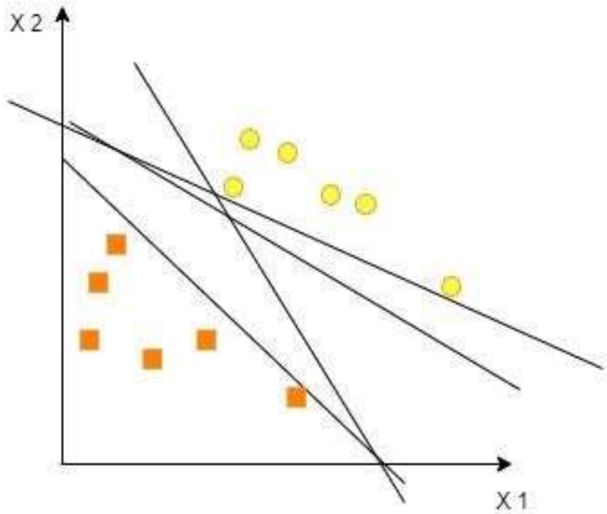

Figure 2. Support Vector Machine(a)

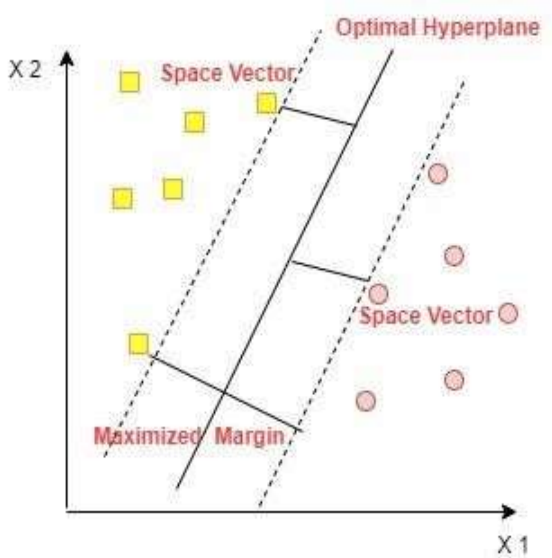

Figure 3. Support Vector Machine(a)

Nandhini M. and V. S. Pream[39], proposed a system to identify turmeric leaf diseases through leaf feature inspection. The authors describe mainly three types of leaf diseases in turmeric plants that are leaf spot, leaf blotch, and rhizome rot. Different machine learning models are used to identify and classify the diseases in turmeric leaves. In this scenario, the author used 800 leaf images of different categories of datasets for pre- processing and segmentization process. In this dataset, the First 200 diseased leaf concerned with leaf spots second 200 diseased leaf concerned with leaf blotch third 200 leaves were affected with rhizome rot disease and the last 200 leaves were healthy leaves. The model was trained with the help of machine learning algorithms like SVM, decision tree, and native Bayes (NB). In this research work, the performance of the model was evaluated using 10fold cross-validation. The study shows that the classification of turmeric leaf diseases using a Support Vector Machine (SVM) gives the highest better accuracy of 93.75% as compared to other algorithms.

Islam Monzurulet al [40], suggests an approach, In this work author utilized image segmentation with multiclass SVM to develop an automated method that classifies diseases on potato plants from a publicly available plant image database that is 'Plant Village' and identified the most prevalent diseases in potatoes, late blight, and early blight. The segmentation approach and utilization of a support vector machine demonstrate disease classification over 300 images with an accuracy of 95% Thus, This approach offers a way to automate extensive plant disease diagnosis.

N. R. Bhimte and V. R. Thool [41], presented a system using an approach for the automatic diagnosis of cotton leaf diseases. SVM classifiers are used to classify images based on the selection of relevant attributes, such as color and texture. Various preprocessing techniques such as filtering, background removal, and enhancement are used by researchers. The diseased segmented section of the cotton leaf is extracted using color-based segmentation. To extract features, a segmented image is utilized. The researchers collected datasets by surveying several fields, the suggested effort would compile the essential database of various cotton illnesses. In this work, There are 130 pictures in the dataset. Out of these, the classifier is trained using 50 images of bacterial blight, 50 images of magnesium shortage, and 30 images of health. Multi-class SVM is used to categorize two diseases, bacterial blight, and magnesium insufficiency. With illness-recognized proof, the constructed classifier performs excellently, providing an accuracy of 98.46%.

Jaisakthi et al[42], used an image processing and machine learning approach. In this approach, researchers have suggested a system for automatically identifying diseases in grape vines. 5675 grape leaves that were obtained from the Plant Village dataset were used by researchers to evaluate the suggested approach. 80% - 20 % of the images have been used for testing and training. The global thresholding method[65] is used for the segmentation process. After the segmentation diseased part of the leaf is more effective for training purposes leaves which leads to improved classification results. The system obtained a better testing accuracy of 93%.

Panigrahi et al[43], focused on supervised machine learning methods for disease diagnosis in maize plants using plant images, including Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF). To choose the model that would forecast plant diseases with the maximum degree of accuracy, the aforementioned categorization strategies are examined and evaluated. The author used a dataset from the plant village website. The data set to implement the suggested methods into practice, the classification model was trained using a labeled dataset. The 3823 images of leaves are labeled in the following category such as common rust, gray leaf spot, northern leaf blight and healthy having 1192 images, 513 images, 956 images and 1162 images respectively. In comparison to the other classification models, the SVM classifier achieved a score of 77.56%.

Sharma et al[44], proposed an approach for disease identification that combines image processing and machine learning using images of potato leaves. Gaussian filtering is used to remove the diseased leaf images, and the K-means clustering approach is then used to identify the desired Region of a leaf. The images of diseased potato leaves are obtained from a publicly available dataset that is “Plant Village”. This dataset includes 1000 images of early and late blight, respectively. On the features that were taken from the segmented images of late blight and early blight potato leaves, an SVM classifier was utilized. The proposed system obtained an accuracy of 93%.

3.6.2Decision Tree

A Decision Tree is another supervised learning technique that can be used for both classification and regression problems. Each leaf node of the decision tree corresponds to a class label, and the interior nodes of the tree are used to represent the attributes toanswer the problem. There are different algorithms of decision tree.

1. C4.5

C4.5 is the decision tree algorithm that is used to classify the data. It can work with both Discrete and Continuous Data. C4.5 can handle the issue of incomplete data very well.

2. CART

The CART algorithm is a type of classification algorithm that is required to build a decision tree on the basis of Gini’s impurity index. CART makes it possible to quicklyclassify the most recent observations.

3. CHAID

The most significant feature is identified by CHAID using a chi-square measurement metric, which is then used recursively until subinformational datasets havea single conclusion.

4. ID3

ID3 (Iterative Dichotomiser 3) uses a top-down approach to build a decision tree. The C4. 5 algorithm is derived from the ID3 algorithm. ID3 is often exclusively utilised for classification issues with only nominal features.

Sabrol H. and K. Satish[45], proposed a plan where five different types of tomato leaves diseases are categorized in the study, including tomato late blight, Septoria spot, bacterial spot, bacterial canker, tomato leaf curl, and images of healthy tomato plant leaves. The classification was carried out by separating the color, shape, and texture characteristics from the images of healthy and diseased tomato leaves. The author employed a dataset of 383 digital images of tomato leaves disease these were collected tasks by a standard digital camera. The image dataset undergoes Otsu's segmentation. They computed a total of nine features from healthy and unhealthy RGB images of the tomato leaves to extract colour attributes. As a classifier, the Decision tree classification achieves a classification accuracy of 97.3%.

Ahmed et al[46], proposed a concept that describes a machine learning-based system for detecting diseases in rice leaves. This research identifies leaf smut, bacterial leaf blight and brown spot illnesses as three of the most prevalent diseases affecting rice plants. The dataset was trained using a variety of different machine-learning techniques after the necessary pre-processing. Data is gathered for this study from the UCI Machine Learning Repository. Waikato Environment for Knowledge Analysis, WEKA, using the Caffe deep learning framework, trained their model on 30880 original images and used 2589 original images for testing. The decision tree approach, after 10-fold cross-validation, produced an accuracy of over 97% when applied to the test dataset.

Rajesh et al[47], proposed a method that gathered the leaf images by downloading them from https://www.kagle.comto categorize leaf diseases. To train the model, images from the dataset are taken. It includes a selection of images captured in diverse positions. More than 1000 images of the leaf can be found in the dataset. Both healthy and diseased images are used to train the model. The suggested system will identify diseased leaves, and it represented an accuracy of greater than 95%.

3.6.3 Random Forest

Leo Breiman and Adele Cutler created the machine learning algorithm known as random forest. The Random forest can be used for both classification and regression problems in machine learning. In order to increase the dataset's predicted accuracy; it contains a number of decision trees on different subsets of the provided dataset and takes the average. A large number of decision trees leads to higher accuracy and avoid the problem of over-fitting.

Ramesh and Shima, et al.[56] proposed a study, in which researchers used publicly available data sets of 160 images to distinguish between healthy and unhealthy leaves using Random Forest. This method provides a simple technique to diagnose plant diseases on a massive scale. To distinguish images that seem to be unhealthy and those that are healthy, the produced datasets are individually trained using Random Forest. In this concept, researchers used the histogram of oriented gradients to determine an image's characteristics (HOG). One benefit of HoG feature extraction is that it works with newly formed cells. Three feature descriptors were used by the researchers in this case. That are Hu moments,

Haralick Texture, and color histogram respectively. The model's classification accuracy was around 70%.

Sandika and Biswas, et al.[57] proposed a work, researchers suggest a technique for categorizing the diseases that typically affect the grape crop and determining the intensity of these diseases using a variety of machine-learning algorithms. 900 images of grape leaves were gathered for this purpose from farmers and other sources. The core concept of the suggested approach is the consideration of complicated back-grounded images of grape leaves taken in an uncontrolled setting. Additionally, compare the results of the four machine learning algorithms PNN, BPNN, SVM, and Random Forest for classifying the various diseases and isolating the background from disease patches. Different diseases including anthracnose, powdery mildew, and downy mildew have an impact on the images of grape leaves. In several instances, multiple types of diseases were present on the

leaves. The suggested method uses features from Random Forest and GLCM (Gray-Level CoOccurrence Matrix) to reach the greatest classification accuracy of 86%.
In this study, Panigrahi, Kshyanaprava Panda, et al.[58] emphasize supervised machine learning methods for disease diagnosis in maize plants using plant images, including Naive Bayes (NB), Decision Trees (DT), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), and Random Forests (RF). To choose the most accurate model for predicting plant diseases, the aforementioned classification strategies are examined and other the analysis, they find the RF algorithm to the other classification methods, achieving the best accuracy of 79.23%. The collection for maize plant disease has four class labels and 3.823 total images. The following information is specific to the dataset for maize disease class labels: Healthy is 513, common rust is 1192, northern leaf blight is 985, and grey leaf spot is 1162.

Saha et al.[59] proposed an automatic system that can detect the main three types of rice leaf diseases that are Bacterial leaf blight, leaf blast, and brown spot respectively by the random forest decision tree classifier. In this study, they worked on, a total of 352 images images of three types of rice leaves, from the two-third data sets, are used for training purposes, and the rest rice leaves were used for testing. In this study, they used a young dataset. Because such types of leaves are more suitable for detecting diseases at an early age rather than old leaves. The used online dataset [64] contains three types of rice leaf diseases that are a blast, blight, and brown spot. Here researcher considerate intensity moment to get effective results for feature extraction After the classification process the applied proposed system calculates 91.47% of overall accuracy at the early stage.

Performance Evaluation Metrics

4. Summary and discussion

Many assessment metrics, including precision, recall, and F1-score, are used to analyse the experiment's results. Following are the computation equations. These can be understand by following figure.

ACTUAL VALUES

		Positive (+VE)	Negative(-VE)
PREDICTED VALUES	+VE	**TP**	**FP**
	-VE	**FN**	**TN**

Figure 4 – Confusion Matrix

Abbreviations: True (T), False (F), Positive(P), Negative(N)

Precision:-Precision is defined as the ratio of correctly anticipated positive observations to all predicted positive observations.

Precision = $TP / (TP + FP)$

Recall:-Recall is the number of accurately predicted positive observations to all of the actual class's observations.

Recall = $TP / (TP + FN)$

F1-score:- The F1-score is the weighted average of Precision and Recall. As a result, when determining this score, both false positives and false negatives are considered.

$F1 - score = 2 * Recall * Precision / (Recall + Precision)$

5. Conclusion and future Direction

This survey aims to provide a comprehensive review of machine learning methods. For plant leaf disease identification, including its various approaches, and latest findings. A brief statistics and available leaf dataset is discussed. Researchers can contribute with more clear images and by dataset creation of commercially used leaf images. It would be more beneficial for disease identification. Tradition disease detection technique with their flow and pros and cons are discussed and explained, why machine learning based techniques more useful. These predictions and advices can be performed any time in the year with higher accuracy. Plant leaf disease identification mechanism presents an overview to identify the disease at a generic model level. These models are broadly applied on different parameter combination and find results. A thorough review of major machine learning techniques are presented and discussed. We can lead to some more better results for more leaves.

6. References

1. https://tradingeconomics.com/india/rural-population-per cent-of-total-population-wb-data.html https://www.downtoearth.org.in/news/agriculture/agri-share-in-gdp-hit-20-after-17-years-economicsurvey-75271
2. Kempf V.A., Trebesius K., Autenrieth I.B. Fluorescent in situ hybridization allows rapid identification of microorganisms in blood cultures. J. Clin. Microbiol. 2000;38:830–838. [PMC free article] [PubMed] [Google Scholar]
3. Hijri M. Plant Pathology. Springer; Berlin, Germany: 2009. The use of Fluorescent in situ hybridisation in plant-fungal identification and genotyping; pp. 131–145. [PubMed] [Google Scholar]
4. DeLong E.F., Wickham G.S., Pace N.R. Phylogenetic stains: Ribosomal RNA-based probes for the identification of single cells. Science. 1989;243:1360–1363. doi: 10.1126/science.2466341. [PubMed] [CrossRef] [Google Scholar]
5. Cai H., Caswell J., Prescott J. Nonculture molecular techniques for diagnosis of bacterial disease in animals a diagnostic laboratory perspective. Vet. Pathol. Online. 2014;51:341–350. doi:
6. 10.1177/0300985813511132. [PubMed] [CrossRef] [Google Scholar]
7. L ́opez M.M., Bertolini E., Olmos A., Caruso P., Corris M.T., Llop P., Renyalver R., Cambra M. Inno- vative tools for detection of plant pathogenic viruses and bacteria. Int. Microbiol. 2003;6:233–243. doi:
8. 10.1007/s10123-003-0143-y. [PubMed] [CrossRef] [Google Scholar]
9. Lievens B., Brouwer M., Vanachter A.C.R.C., Cammue B.P.A., Thomma B.P.H.J. Real-time PCR for detection and quantification of fungal and oomycete tomato pathogens in plant and soil samples. Plant Sci.
10. 2006;171:155–165. doi: 10.1016/j.plantsci.2006.03.009. [CrossRef] [Google Scholar]
11. Schaad N.W., Frederick R.D. Real-time PCR and its application for rapid plant disease diagnostics. Can.J. Plant Pathol. 2002;24:250–258. doi: 10.1080/07060660209507006. [CrossRef] [Google Scholar] [10] Van der Wolf J., van Bechhoven J.R.C.M., Bonants P.J.M., Schoen C.D. Plant Pathogenic Bacteria. Springer; Berlin, Germany: 2001. New technologies for sensitive and specific routine detection of plant pathogenic bacteria; pp. 75–77. [Google Scholar]
12. Clark M.F., Adams A. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J. Gen. Virol. 1977;34:475–483. doi: 10.1099/0022-1317-34-3-
13. 475. [PubMed] [CrossRef] [Google Scholar]
14. Gorris M.T., Alarcon B., Lopez M., Cambra M. Characterization of monoclonal antibodies specific for Er-winiacarotovora subsp. atroseptica and comparison of serological methods for its sensitive detection on potato tubers. Appl. Environ. Microbiol. 1994;60:2076–2085. [PMC free article] [PubMed] [Google Scholar]

15. L´opez M.M., Llop P., Cubero J., Penyalver R., Caruso P., Bertolini E., Penalver J., Gorris M.T., Cam- bra M. Plant Pathogenic Bacteria. Springer; Berlin, Germany: 2001. Strategies for improving serological and molecular detection of plant pathogenic bacteria; pp. 83–86. [Google Scholar]
16. Chitarra L.G., van den Bulk R.W. The application of flow cytometry and fluorescent probe technology
17. For detection and assessment of viability of plant pathogenic bacteria. Eur. J. Plant Pathol. 2003;109:407–
18. 417.doi: 10.1023/A:1024275610233. [CrossRef] [Google Scholar]
19. Burling K., Hunsche M., Noga G. Use of blue-green and chlorophyll fluorescence measurements for differentiation between nitrogen deficiency and pathogen infection in winter wheat. J. Plant Physiol.
20. 2011;168:1641–1648.doi: 10.1016/j.jplph.2011.03.016. [PubMed] [CrossRef] [Google Scholar]
21. Kuckenberg J., Tartachnyk I., Noga G. Temporal and spatial changes of chlorophyll fluorescence as a basis for early and precise detection of leaf rust and powdery mildew infections in wheat leaves. Precis.
22. Agric.2009;10:34–44. doi: 10.1007/s11119-008-9082-0. [CrossRef] [Google Scholar]
23. Chaerle L., Lenk S., Leinonen I., Jones H.G., Van Der Straeten D., Buschmann C. Multi-sensor plant imaging: Towards the development of a stress-catalogue. Biotechnol. J. 2009;4:1152–1167. doi:
24. 10.1002/biot.200800242.[PubMed] [CrossRef] [Google Scholar]
25. Cs´efalvay L., Gaspero G.D., Matous K., Bellin D., Ruperti B., Olejnickova J. Pre-symptomatic detection of Plasmoparaviticola infection in grapevine leaves using chlorophyll fluorescence imaging. Eur.
26. J. Plant Pathol.2009;125:291–302. doi: 10.1007/s10658-009-9482-7. [CrossRef] [Google Scholar] [19] Scholes J.D., Rolfe S.A. Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance: A phenomics perspective. Funct. Plant Biol. 2009;36:880–892. doi:
27. 10.1071/FP09145. [CrossRef] [Google Scholar]
28. Delalieux S., van Aardt J., Keulemans W., Schrevens E., Coppin P. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. Eur. J. Agron. 2007;27:130–143. doi: 10.1016/j.eja.2007.02.005. [CrossRef] [Google Scholar]
29. Kobayashi T., Kanda E., Kitada K., Ishiguro K., Torigoe Y. Detection of rice panicle blast with multispectral radiometer and the potential of using airborne multispectral scanners. Phytopathology.
30. 2001;91:316–323. doi:10.1094/PHYTO.2001.91.3.316. [PubMed] [CrossRef] [Google Scholar]
31. Zhang M., Qin Z., Liu X., Ustin S.L.Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. Int. J. Appl. Earth Observ. Geoinf. 2003;4:295–310. doi: 10.1016/S0303-2434(03)00008-4. [CrossRef] [Google Scholar]
32. Oerke E., Steiner U., Dehne H.W., Lindenthal M. Thermal imaging of cucumber leaves affected by downy mildew and environmental conditions. J. Exp. Bot. 2006;57:2121–2132. doi: 10.1093/jxb/erj170.
33. [PubMed] [CrossRef] [Google Scholar]
34. Chaerle L., Leinonen I., Jones H.G., Van Der Straeten D. Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. J. Exp. Bot. 2007;58:773–784. doi:
35. 10.1093/jxb/erl257. [PubMed] [CrossRef] [Google Scholar]
36. Lindenthal M., Steiner U., Dehne H.W., Oerke E.C. Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography. Phytopathology.
37. 2005;95:233–240. doi: 10.1094/PHYTO-95-0233. [PubMed] [CrossRef] [Google Scholar]

38. https://silks.csb.gov.in/jhansi/diseases-and-pests-of-food-plants/
39. Hallau L, Neumann M, Klatt B (2017) Automated identification of sugar beet diseases using
40. smartphones. Plant Pathol 67(2):399–410. https:// doi. org/ 10. 1111/ ijlh. 12426 \newline{}
41. WenzhuYangSileWangXiaolanZhaoJingsiZhangJiaqi Feng (2015) Greenness identification based on HSV decision tree https://www.sciencedirect.com/science/article/pii/S2214317315000347
42. Kaur S, Pandey S, Goel S (2018) Semi-automatic leaf disease detection and classification system for soybean culture. IET Image Process 12(6):1038–1048. https:// doi. org/ 10. 1049/ iet- ipr. 2017. 0822 [30]Kaur K, Marwaha C (2017) Analysis of Diseases in Fruits using Image Processing Techniques. In: Precedings of International Conference on Trends in Electronics and Informatics ICEI 201.
43. pp 183–189
44. [31]Yao Q, Guan Z, Zhou Y, Tang J, Hu Y, Yang B (2009) Application of support vector machine for detecting rice diseases using shape and color texture features. 2009 International Conference on Engineering Computation, ICEC 2009 pp 79–83. https:// doi. org/ 10. 1109/ ICEC. 2009. 73 Bai X, Li X, Fu Z, Lv X, Zhang L (2017) A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. Comput Electron
45. Agric 136:157–165. https:// doi. org/ 10. 1016/j. compag. 2017. 03. 004
46. Asfarian A, Herdiyeni Y, Rauf A, Mutaqin KH (2013) Paddy diseases identification with texture analysis using fractal descriptors based on fourier spectrum. Proceeding - 2013 International Conference on Computer, Control, Informatics and Its Applications: Recent Challenges in Computer, Control and
47. Informatics, IC3INA 2013. pp 77–81. https:// doi. org/ 10. 1109/ IC3INA. 2013. 68191 52 [34]Bhargavi, K., and S. Jyothi. "A survey on threshold based segmentation technique in image processing." International Journal of Innovative Research and Development 3.12 (2014): 234-239.
48. Lomte, S. S., and A. P. Janwale. "Plant leaves image segmentation techniques
49. Sannakki SS, Rajpurohit VS, Nargund VB, Kulkarni P (2013) Diagnosis and Classification of Grape Leaf Diseases using Neural Networks. Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013, pp 1–5
50. Sharma, Sakshi, Vatsala Anand, and Swati Singh. "Classification of Diseased Potato Leaves Using Machine Learning." 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2021.
51. https://support.echoview.com/WebHelp/Windows_and_Dialog_Boxes/Dialog_Boxes/Variable_propert ies_dialog_box/Operator_pages/GLCM_Texture_Features.html
52. Nandhini, M., V. S. Pream, and M. S. Vijaya. "Identification and classification of leaf diseases in turmeric plants." International Journal of Engineering Research and Applications 6.2 (2016): 48-54. [40]Islam, Monzurul, et al. "Detection of potato diseases using image segmentation and multiclass support vector machine." 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE). IEEE, 2017.
53. N.R.Bhimte and V. R. Thool, “Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier,” 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 340–344.
54. Jaisakthi, S. M., P. Mirunalini, and D. Thenmozhi. "Grape leaf disease identification using machine learning techniques." 2019 International Conference on Computational Intelligence in Data Science (ICCIDS). IEEE, 2019.
55. Panigrahi, Kshyanaprava Panda, et al. "Maize leaf disease detection and classification using machine learning algorithms." Progress in Computing, Analytics and Networking. Springer, Singapore, 2020. 659669.

56. Sharma, Sakshi, Vatsala Anand, and Swati Singh. "Classification of Diseased Potato Leaves Using Machine Learning." 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2021.
57. Sabrol, H., and K. Satish. "Tomato plant disease classification in digital images using classification tree." 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2016. [46]Ahmed, Kawcher, et al. "Rice leaf disease detection using machine learning techniques." 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI). IEEE, 2019.
58. Rajesh, B., M. Vishnu Sai Vardhan, and L. Sujihelen. "Leaf disease detection and classification by decision tree." 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184). IEEE, 2020.
59. Wan, Long, et al. "Hyperspectral Sensing of Plant Diseases: Principle and Methods." Agronomy 12.6 (2022): 1451.
60. D'hondt, Liesbet, et al. "Applications of flow cytometry in plant pathology for genome size determination, detection and physiological status." Molecular Plant Pathology 12.8 (2011): 815-828. [50]Hashim, I. C., et al. "Application of thermal imaging for plant disease detection." IOP Conference Series: Earth and Environmental Science. Vol. 540. No. 1. IOP Publishing, 2020.
61. Amalraj, Augustine, and Sreeraj Gopi. "Medicinal properties of Terminalia arjuna (Roxb.) Wight &Arn.: a review." Journal of traditional and complementary medicine 7.1 (2017): 65-78.
62. Klimek-Szczykutowicz, Marta, Agnieszka Szopa, and Halina Ekiert. "Citrus limon (Lemon) phenomenon—a review of the chemistry, pharmacological properties, applications in the modern pharmaceutical, food, and cosmetics industries, and biotechnological studies." Plants 9.1 (2020): 119.
63. https://www.researchgate.net/publication/304274138_Jamun_A_traditional_fruit_and_medicin e
64. https://www.researchgate.net/publication/301330310_Antioxidant_and_Anti-
65. inflammatory_Activities_of_Platanus_orientalis_An_Oriental_Plant_Endemic_to_Kashmir_Pla nes
66. https://www.medigraphic.com/cgi-bin/new/resumenI.cgi?IDARTICULO=34540
67. Ramesh, Shima, et al. "Plant disease detection using machine learning." 2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C). IEEE, 2018.
68. Sandika, Biswas, et al. "Random forest based classification of diseases in grapes from images captured in uncontrolled environments." 2016 IEEE 13th international conference on signal processing (ICSP).
69. IEEE, 2016.
70. Panigrahi, Kshyanaprava Panda, et al. "Maize leaf disease detection and classification using machine learning algorithms." Progress in Computing, Analytics and Networking. Springer, Singapore, 2020. 659669.
71. Saha, Sristy, and Sk Md Masudul Ahsan. "Rice disease detection using intensity moments and random forest." 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). IEEE, 2021.
72. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 2012, 118, 259–272.
73. Yamamoto, K.; Guo,W.; Yoshioka, Y.; Ninomiya, S. On plant detection of intact tomato fruits using image analysis and machine learning methods. Sensors 2014, 14, 12191–12206.
74. Al Muqarrabun, L. M. R., et al. "Medicinal uses, phytochemistry and pharmacology of Pongamia pinnata (L.) Pierre: A review." *Journal of ethnopharmacology* 150.2 (2013): 395-420.
75. 2020[64]A. K. G. Francisco, “GitHub,” 2020 [Online]. https://github.com/aldrin233/ Rice

76. Diseases DataSet. (Accessed August 2020)
77. Al-Amri, Salem Saleh, and Namdeo V. Kalyankar. "Image segmentation by using threshold techniques." *arXiv preprint arXiv:1005.4020* (2010).
78. Wang, Yu, Qian Chen, and Baeomin Zhang. "Image enhancement based on equal area dualistic subimage histogram equalization method." *IEEE transactions on Consumer Electronics* 45.1 (1999): 68-75**.**