Utilizing Machine Learning to find Plant Leaf Disease

Dr.J.Savitha[1], Manikandan.N[2], SajithRamana.D[3],Vasanth.R[4]

[1]*Professor, Department of Information Technology, Dr. N.G.P Arts and Science College*

[2,3,4]*Student, Department of Information Technology, Dr. N.G.P Arts and Science College*

***Abstract*—In contrast to the agricultural area, every other field has benefited from new technologies in some way. Studies from the past indicate that plant leaf diseases are the only reason of the 42% loss in agricultural production, which is increasing. By detecting a disease from the input photographs, our method for detecting plant leaf disease can solve our big issue. Feature extraction, picture segmentation, and image pre-processing were the processes in this procedure. The K Nearest Neighbor (KNN) classification is then applied to the outcomes of these three phases. The proposed implementation has a 98.56% prediction accuracy for plant leaf diseases. Additionally, it offers extra information regarding a leaf-affecting disease, such as the Affected Area, Disease Name, Total Accuracy, Sensitivity, and Elapsed**

***Index term*—Image Segmentation, Machine Learning, Plant Leaf Disease Detection**

I. INTRODUCTION

As new technologies are created, the sector of agriculture is growing in popularity since it serves a range of additional uses in addition to providing food for a big population. Because they serve as a source of energy and contribute to the fight against global warming, plants are vital to human existence. There are many diseases that affect plants today, some of which have terrible consequences for the economy, society, and environment. Consequently, it is essential to quickly and properly identify plant diseases. Whether or not an infectious agent is the principal cause of a plant disease might be used to define it broadly.

The method used makes use of digital image processing tools to produce the desired result. The extent of the ailment cannot be adequately determined by the human eye because the symptoms are subjective in nature. In order to assess the severity of diseases in a manufacturing region, observations made with the naked eye are typically used. The image processing has made a big contribution to agriculture. Several neural network techniques, including Back Propagation and Principal Component Analysis (PCA), have been used to identify the fungus sickness.to identify plant leaf disease by increasing the necessary rate in the classification method. The identification of the fungi is done from their morphology, due to the effects done on the reproductive structure directly.

The multi-class classification method utilized is linear SVM, which is relatively inefficient and reduces classification accuracy by merely classifying the data into two groups.

This paper's major objective was to suggest enhancements to the existing classification methods for plant leaf disease detection using machine learning. The research of numerous plant leaf illnesses, as well as the investigation and evaluation of several methods for disease detection in plant leaves utilizing image processing techniques, were the main goals of this paper.

In our work, additional classifiers like KNN will be employed to divide the data into more than two categories. Our system identified seven disorders in all. 75 photos made up the data collection. Furthermore, this technology identifies the disease name, accuracy percentage of the affected area, sensitivity, and time since detection. In this essay, diseases such Down Mildew, Early Blight, Mosaic Virus, Leaf Miner, and White Fly are explained. The section is organized as follows: Section II is a survey of the literature that includes descriptions of all previous research and activity as well as open-ended questions for future investigation. Section III: Methodology describes the process behind our strategy. Implementation Section IV provided a detailed explanation of each phase. The experiment setting is described in depth in Section V Experimentation, and the results are presented in Section VI and Section VII, which provide the paper's overall conclusion.

II. LITERATURESURVEY

This paper presents research on various order approaches that might be used for identifying plant leaf diseases. Numerous grouping systems exist, including the k-Nearest Neighbor Classifier, the Probabilistic Neural Network, the Genetic Algorithm, the Support Vector Machine, and Principal Component Analysis. The numerous characterization approaches used for plant leaf malady order are reviewed in t0l0his article [1]. The created preparation plan consists of four main steps, starting with an information shading change structure. A RGB image is created, and since RGB is for shading, this RGB is converted to HSI.

In the end, the surface insights are calculated from SGDM lattices using age and his for the shading descriptor.[2] This paper addresses this issue with the aim of developing picture preparation calculations that can detect yield issues from pictures, in light of shading, surface, and shape to naturally identify diseases or other conditions that may affect harvests and provide the rancher with the prompt and accurate answers via SMS. [3]. They provide a thorough analysis of how to identify diseases in plant leaves and group them using picture-making approaches in this paper. In this way, photo-editing has been linked to the confirmation of plant diseases. The paper has been divided into two categories that are considered to be standard, namely identifiable proof and collecting of leaves.[4].

In the current work, a method for accurately and quickly diagnosing plant diseases using various picture-handling techniques and artificial neural networks (ANN) is proposed. The goal of the current work is to develop a fundamental infection recognition framework for plant diseases. Taking the photos is the first step in the process. split and separated using the Gabor channel. According to test results, order execution by ANN using a list of capabilities is superior and has a 91% accuracy rate.[5] While using principal component analysis (PCA), the components of the component data were not diminished, but the perfect off- formation results for grape ailments were obtained because the fitting precision and the desire accuracy were both 100%, and those for wheat diseases were obtained because the fitting accuracy and the figure precision were both 100%. While using PCA to reduce the components of the component data, the perfect affirmation result for grape diseases was obtained because the fitting accuracy was 100% and the desire precision was 97.14%, and it was obtained for wheat diseases because the fitting accuracy and conjecture accuracy were both 100%.[6] Test findings show that SVM's ability to execute a plan of action is superior to neural frameworks'. Cucumber illness confirmation rate based on SVM of shape and surface segment is superior to that of just using the shape highlight.[7] Its purpose is to demonstrate a clever method for preventing region-specific evidence that is subject to image processing. This system is operational 24/7. Various issues are present throughout the middle of the day.

To differentiate citrus unpleasant painful from leaf-establishment, a general ulcer sore descriptor is initially utilized. Thirdly, AdaBoost is mastered in feature assurance and classifier learning [9] and a two-level different tiered area structure is created to recognize the contamination harm. This investigation's primary goal is to create a model system for diagnosing the paddy contaminations known as Blast Disease (BD), Brown-Spot Disease (BSD), and Narrow Brown-Spot Disease (NBSD). The approach includes image acquisition, converting RGB images into a two-dimensional image using modified thresholding based on the neighboring entropy far-thest point and Otsu method. Accordingly, the accuracy rates for the paddy illnesses are about 94.7 percent when employing the age rule technique [10]. The illnesses considered are Grape natural product can become over powdered by powdery or downy mildew. Genuine turns, minor center points, and other leaf characteristics are removed in order to identify contamination and are then handed to the classifier for organization.[11] The best data collection precision (96.7%) was attained by the model utilizing 14 selected HSI surface features, which indicated that it is ideal to use a reduced shade, submersion, and power surface rundown of abilities to[12] differentiate citrus strip ailments. A robustness assessment of the gathering model's typical plan precision and standard deviation were 96.0% and 2.3%, respectively, showing that the model is viable for asking new natural item tests as indicated by their strip criteria.

Due to direct effects on the reproductive structure, the morphology of the fungi is used to identify themToday, a variety of technologies are used to identify plant diseases in distant areas. K-means division computation and SVM Classifier are used in conjunction with each other as a plant disease diagnosis tool. The Agri-Guide v1.0 recommends safety measures and offers therapy for plants with a response time of 1ms. The Agri-Manage v1.0 offers 83% accuracy precision and 73% accuracy to the unidentified patterns of monocot and dicot plant diseases. Somewhere, a technique is used to specify the botanical tissues of the fruits and vegetables and to control their quality. This method is called Electrical Impedance Spectroscopy, and it is used to investigate the composition and electrical characteristics of plant tissues. In this study, it is suggested that.

When the requirement for a simple plant leaf disease detection system is surveyed in order to promote improvements in agribusiness, some significant factors emerge. To identify plant diseases, a variety of methods are used, including BPNN, SVM, K-implies bunching, and SGDM. This study makes the suggestion that every original image is an RGB image since it combines primary colors like red, green, and blue. Due to image enhancement, the improved image has higher quality and clarity than the original image. The term "disease" is typically only used to describe the destruction of live plants. The paper presents a crucial technique that must be created for identifying plant diseases. . The approach investigations aim to cure plant diseases more precisely and with more throughput than human expectations. The CIELAB color model's results are unaffected by background, leaf type, type of disease spot, or camera flash, and it properly identifies disease.

[19] SVM classifier that has been used to identify plant diseases. As seen in the outcomes of one image, the location is 5.56Background and Black pixels are both divided in the initial advance. This analysis includes a fresh piece of work that aims to identify plant contamination zones.

[20] The prevention of plant diseases is the next point that is becoming more and more important since it promotes the affordable development of the entire nation. Irresistible disease corrupts the nature of nutrition and causes yield misfortunes in the agricultural sector. It broadens the rural item's scope and character. Inthis For the purpose of determining the irresistible disease in the plants, paper, leaf photographs are used. To characterize the infection in the plants, an ANN technique has been employed, including self-regulating highlight maps, support vector machines, SVMs, etc. This paper[21] examines early detection and acknowledgment tactics that are suggested for the discovery of the rust in the plants. Soft c-infers For the extraction of implanted highlights in the wheat leaves, a clustering calculation, infection detection, acknowledgement of its sort, and ID calculation have been developed. Under the guidance of artificial neural networks (ANN), this is carried out.

III. METHODOLOGY

A. Pre-ProcessingStage

This process is a preparation step for taking photos of plant leaves. With pixel resolutions of 568x1020, the RGB plant leaf images were taken with a digital camera. There have been 75 information tests gathered. There are five different forms of plant diseases included in it. The model input is a plant leaf image that has been further transformed to grayscale using the MATLAB Picture Preparation Library. Figure 1 depicts a block schematic of the stages employed in the proposed study.

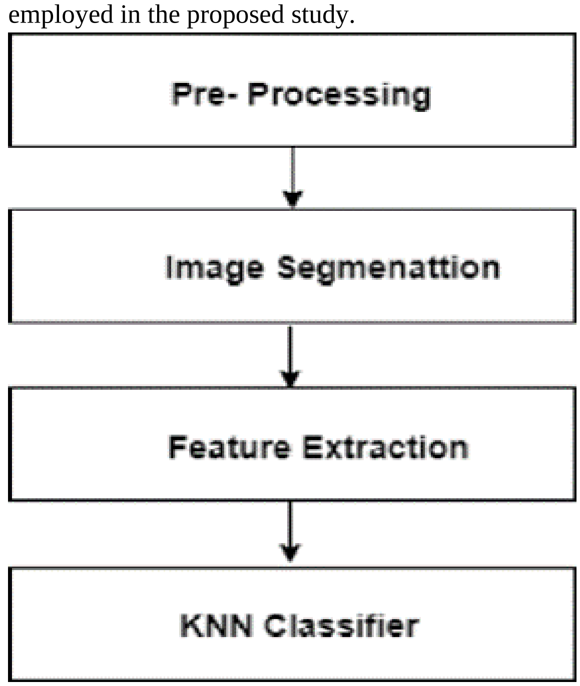

Fig.1.BlockDiagramofallthesteps

B. ImageSegmentation

Applying the image segmentation approach will separate the images based on their characteristics. The two main types of image segmentation techniques are region-based segmentation and threshold-based segmentation. The valuable pieces must then be extracted as the following step Not every area includes a sizable amount of information. As a result, for additional analysis, only the portion that comprises more than 50% of the data is considered. A generic block diagram showing the many stages of the process is shown in Figure 1. In this work, the pictures are segmented using the more noise-resistant region-based k-mean segmentation technique. For picture segmentation, we will offer a data set, and the result will be cluster data.

C.FeatureExtraction

The distinctive characteristics of the leaves are portrayed through a variety of traits. A few leaves have a unique shape, some have noticeable surface characteristics, and some are distinguished by a combination of these qualities.

1) GLCM Algorithm: This approach, first put forth by Haralick in 1973, is still one of the most widely used strategies for surface investigation [11]. The main concept behind this method is to create highlights based on gray level co-occurrence matrices (GLCM). A gray-level co-occurrence matrix is used to extract co-occurrence features. Since the GLCM algorithm only accepts 13 features at a time, only 13 features were used throughout the entire project. The process is outlined in the formula below.

- In the matrix where the data is saved, total up all the pixels.
- Place the pixel counts in the matrix P[I,j].
- Use the histogram approach to examine the similarity of the pixels in the matrix.
- The pixels must be divided in order to normalize the g elements.

C. Classifier

The co-event highlights for the leaves are extracted, contrasted, and then compared in a similar way to be stored in the feature library during the classification stage. Although there are many other classifiers, only two have been discussed in this paper: the KNN classifier that has been proposed and the Support vector machines (SVMs) classifier that is now in use. Many related supervised learning techniques are used in SVM classifiers for both classification and reversal. In a high- or large-dimensional space, a support vector machine creates a hyperplane or set of hyperplanes that can be used for grouping, reversal, or other tasks.The K-NN also serves as the classifier for calculations using supervised learning. Although we are aware of the objectives in supervised learning, the target pathway is unknown.

The best model for understanding AI closest neighbors structures. Give us a moment to consider the fact that there are various categories of named tests. The concept of similar objects being grouped or gathered together is homogenous in nature. Currently, if something that isn't labeled needs to be named under one of the namesake labels. Currently, the best calculation for ranking things is to use their K-closest neighbors, which has a record of every class that is currently usable and can accurately place new items into classes depending on the most votes they receive. Therefore, KNN is one alternative to classify an unlabelled object into a recognized category. The K value chosen serves as the primary metric for evaluating the accuracy and competence of the k-NN calculation. Additionally, KNN's positive stance is that it uses an unbiased algorithm and makes no assumptions about the data being considered.

IV.LITERATURE REVIEW

Fig. 2 provides a thorough overview of the proposed study. The proposed work focuses on three essential elements to complete the entire task.

A. TrainingDataset

The image is sent as information and enters the pre-processing phase, where the quality of the leaf image is improved. Unaffected areas are also used to indicate defective leaves, as seen in figure 3. The GLCM computation is then connected. It is a technique where an image's texture is taken into account, transformed to grayscale, and then presented with remarkable features. The segmentation algorithm k-mean, which is based on textural features, is related. The centered point from which the Euclidian separation is calculated and the information is clustered by similarity is determined in the k-mean clustering technique. Here, clusters in that situation do a partition each time. In any case, K-means clustering separates the leaf image into four portions or clusters, each of which may include one or more disease-causing cells, indicating that the leaf is diseased with several different diseases.After providing the group number, the KNN classifier will return a group that contains harmful parts. We obtain the name of the disease and

the area it affects via the KNN classifier.
The Euclidian separation is calculated, and the data is grouped according to their shared characteristics. We will enter the number of people in the group where our ROI (region of interest) is. We will use the KNN classifier to come up with a specific defective disease term after entering the cluster number.

V.EXPERIMENTATION

Every experiment is run with MATLAB. There were 75 samples of plant leaves utilized as input data for the disease, of which 20 were used for testing and 20 were used to train the dad.Out of the several clusters we tried to run on this leaf, cluster 4 was the one that we discovered to be the most accurate for identifying the leaf's ailment. For instance, a leaf in figure 3 is afflicted with four distinct diseases, but the final ailment from which the leaf is suffering is from the Mosaic Virus, and its affected area is 18.58%, which is higher than all of the other four affected areas. As a result, one area's % out of the four afflicted areas was the largest, and as a result, that percentage disease will be the one from which our disease ultimately suffers, as shown in Figs. 3 to 8. Figure 3 depicts a diseased leaf, while Figure 4 is a cluster 1 with a 15.21% afflicted leaf area percentage. also known as Leaf Miner, is a disease. Fig. 5 shows a cluster 2 with a percentage of the leaf impacted at 18.48%, and the illness is known as White Fly like Ways. The other two clusters are not affected. Fig. 6 shows a cluster 3 in which 18.58% of the leaf's surface area is impacted. Also known as the Mosaic Virus Fig. 7 shows a cluster 4 with a 15.21% damaged leaf area percentage. Early Blight is the name of the illness. The illness name for Figure 8 is Mosaic Virus, which was chosen based on the largest percentage area.

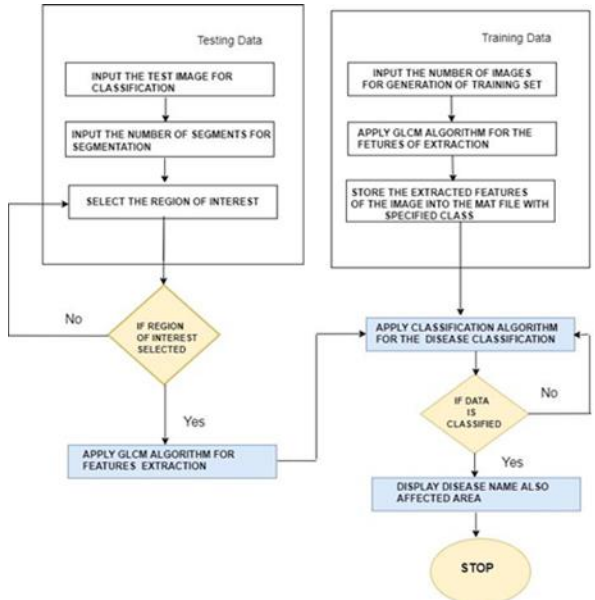

Fig.2.FlowChart

Fig.3.Input plant disease leaf Image

B.Testing Dataset

V.RESULTS

Diseases that are taken into considerations are Mosaicvi us, In this case, the data may be trained or untrained. In the event that information is prepared, it will lawfully enter the classification following element Extraction. Additionally, we must enter the number of segments for segmentation if the data is untrained. K-mean clustering, which is a textural-based approach, is then calculated. The concentrated point in the k- mean clustering is established from Which Powdery Mildew, DownyMildew, Miner, White Leaf Fly displays visuals that produce erroneous results and is unable to correctly only superior for the detection of two diseases, not for any others.

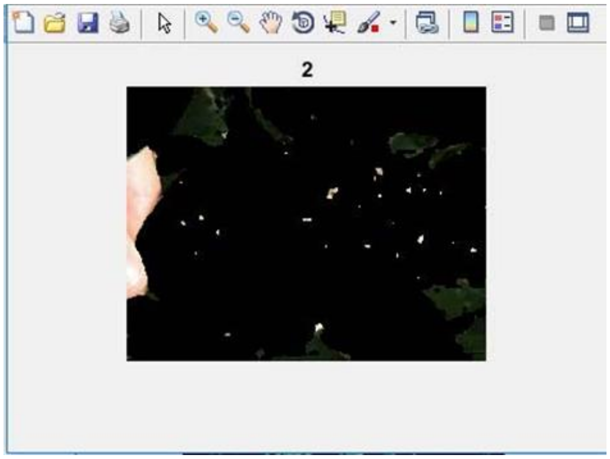

Fig.4.Cluster1 Percentage of affected Area:15.21%
Disease Name Leaf Miner

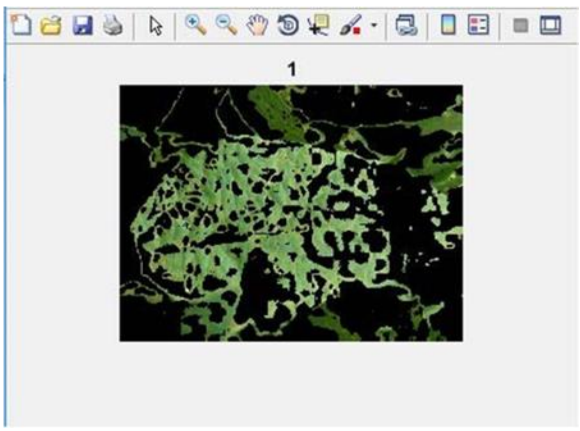

Fig.5.Cluster2 Percentage of affected Area:18.48
Disease Name White Fly

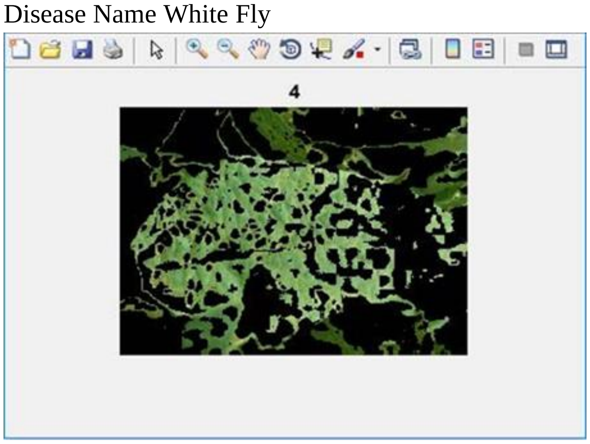

Fig.6.Cluster3 Percentage of affected Area:18.58%
Disease Name Mosaic Virus

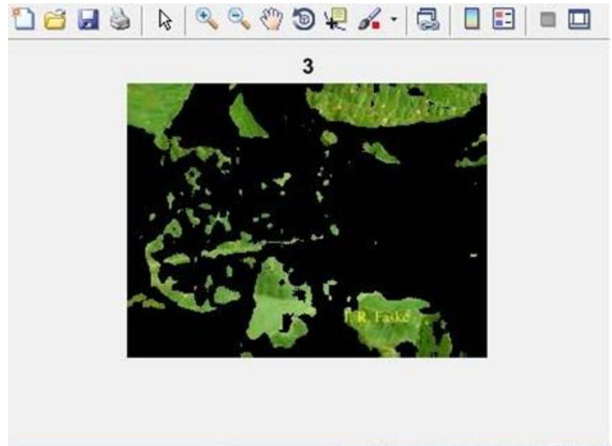

Fig.7.Cluster4 PercentageofaffectedArea:15.21%
Disease Name EarlyBlight

more than that. KNN classifier may also identify a variety of disorders. Even so, it demonstrates the extent of each disease's damage to the leaf's surface. Table 1 shows that the average accuracy of the suggested technique is 98.56% while the average accuracy of the existing technique is 97.6%.

BLEI COMPARISON BETWEEN EXISTING ALGORITHM AND PROPOSED ALGORITHM

Algorithm Used	Number of Images used for dataset	Name of the Diseases	Detection Accuracy
Linear SVM Classifier (Existing Algorithm)	150	Sun Burn Disease, Fungal Disease, Early Scorch, Bacterial Leaf Spot, Frog eye leaf Spot	97.6%
KNN Classifier (Proposed methodology)	75	Mosaic virus, Powdery Mildew, Downy Mildew, Leaf Miner, White Fly	98.56%

VI.CONCLUSIONANDFUTUREWORK

- The KNN classifier outperformed the SVM classifier in this paper's main experiment to demonstrate how machine learning can be used to improve existing classification algorithms for plant leaf disease identification.
- Another objective was to investigate several leaf illnesses that have not before been investigated, so the proposed algorithm was tested on five distinct ailments that affect plants, namely: Early Blight, the mosaic virus, down mildew, the white fly, and the

leaf miner.
- According to the experimental results, the accuracy of the suggested algorithm is 98.56%, compared to 97.6% for the present system.
- use more datasets and focus on improving accuracy.

REFERENCE

[1] Sanjay N. Ghaiwat, ParulArorar, Detection and Classification of Plant Leaf Diseases Using image processing Techniques: A Review, InterNational Journal of React advance in Engineering Technology, ISSN (online):2347-2812, Volume 2 Issues 3 2014

[2] "Deep learning-based image recognition for plant diseases" by Hughes, David P., and Marcel Salathe. (2016) Link: https://arxiv.org/abs/1604.03169

[3] "Detection of plant diseases using convolutional neural networks" by Quds's, Md. Abdul, and Yongcheng Gao. (2017) Link: https:// ieeexplore.ieee.org/abstract/document/8100079

[4] "A review on deep learning techniques for image and speech recognition" by LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. (2015) Link: https://www.nature.com/articles/nature14539

[5] "Deep learning for plant disease detection and diagnosis" by Sladojevic, Srdjan, et al. (2016) Link:https://www.researchgate.net/publication/310499512_Deep_Learning_for_Plant_Disease_Detection_and_Diagnosis_A_survey

[6] "PLANT DISEASE CLASSIFICATION USING A DEEP NEURAL NETWORK" by Ferentinos, K. P. (2018) Link: https://www.researchgate.net/publication/325207926_PLANT_DISEASE_CLASSIFICATION_USING_A_DEEP_NEURAL_NETWORK