Voting Classification Method with Clustering Method for the Plant Disease Detection

Dishant Sharma[1]
Research Scholar,
Computer Science and Engineering,
Chandigarh University, Gharuan, Mohali
(Punjab)
dishantsharma44.ds@gmail.com

Nitika Kapoor[2]
Assistant Professor,
Department Of Cse,
Chandigarh University, Gharuan, Mohali
(Punjab)
nitika.cse@cumail.in

Dr. Parminder Singh [3]
Associate Professor,
Department of Information Technology,
Chandigarh Engineering College, Landran,
(Punjab)
singh.parminder06@gmail.com

***Abstract*—The economy of the country is largely dependent on the production of the agricultural field. Thus, it is essential to detect the diseases of plant at primary phase for maximizing the agriculture yield. The automated methods are assisted in detecting the diseases at initial phase and providing more accuracy. The disease is detected after the starting of appearance of symptoms on the leaves of the plant using automated methods. This work introduced an automated method for detecting the disease of plant on the basis of 4 tasks such as to pre-process the image, segment an image, extract the features and classify the disease. The literature survey conducted on diverse methods, which the researchers suggested already, is also considered in this paper. The symptoms of infected leaf are analyzed using GLCM algorithm and the voting classifier is presented to classify the disease. In this classifier, DT, SVM and K-NN techniques are integrated that will lead to enhance the accuracy for detecting the diseases in advance**

Keywords—plant Disease, glcm, k-mean, svm, decision tree, k-nearest neighbor, voting classifier

I. Introduction

Developing potential to diagnose and resist plant diseases helps prevent harvest losses. Plant diseases pose a serious challenge to crop protection worldwide, contributing to 10–16% of the yield loss worldwide every year. The diseases of plants are less exacting for controlling in their premature phases. However, cultivators usually become unsuccessful in the detection of appearing plant infections as frequent variation are usually not apparent. This setback in disease detection would result in substantial crop loss, which would reduce plant yields and lead to economic repercussions. Although the challenge presented by plant infections is lethal to agricultural productivity [1], involving excessive physical effort in plant disease recognition is a trash. It will take a long time to learn diagnostic skills and parameters and will require significant time of practice to upscale accuracy. Thus, a machine will help in this matter. Prior to the introduction of a machine for plant disease detection, the diagnosis of plant disease was dependent on the specialist agronomists or phytopathologists to a great extent. They developed criteria specification, as per the research in the field and observations of plants, to heal the illness prior to it divagates. Nevertheless, owing to the great diversity of disease symptoms and the enormous divergence in the same symptoms among different plants, even specialists with specialized visual device may be unsuccessful to recognize diseases in the early phase.

The area of IPT and MLA in detecting the disease and detection is a research hotspot that has huge potential to deal with issue of timely and accurate recognition of pests and infections [2]. Machine Learning Algorithms are emerged in several domains namely ImageNet at which the accuracy has crossed the human level of perception. Obtaining this kind of efficacy in identifying insect and infection in plants is the main aim of current exploratory work. Image processing typically defines the computer-specific manipulation and examination of captured photographs with a comprehensive range of sensors, such as light cameras, and sensors functioning in various electromagnetic spectrum bands. In fact, numerous encouraging researches have been done in the field of detecting the plant disease via hyper-spectral methods. However, hyperspectral equipment is pricey and not readily available to common agriculturalists and employed workers [3]. Therefore, the image processing techniques are highly useful in insect and disease detection by means of coloured photographs. Symptoms of diseases of plants are appeared on the infected plant in most cases. IPTs are designed for diagnosing such cases in a faster, more accurate and cost-effective way using normal digital pictures. Figure 1 depicts the image processing-based process to detect the diseases occurred on plants.

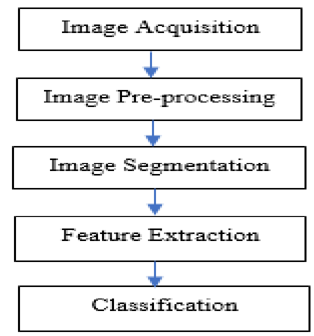

Figure 1: Image Processing based Plant Disease Detection

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

a. Image Acquisition: This task is concerned with the acquirement of different plant imagery from different datasets [4]. One such dataset is 'The Plant Village Dataset. It is an open-access repository containing 54,323 images in total. Multiple classes are selected per specie. The images are generally captured in controlled environmental conditions. This might lead to model bias. A test dataset comprising a number of images can be obtained from Google as well to access this. These images comprise supplementary plant anatomy, in-field background data and variable disease stages.

b. Image Pre-processing: This stage is executed for deciding the performance of a model. It is quite challenging to differentiate viral, bacterial, and fungal disorders, and generally, an overlap of signs appears. These signs are present in the form of any quantifiable variation in color, shape, or function that occurs as the plant reacts to the pathogen. This criticality might be overcome by using RGB images. It generates clear, de-noised images that may consume more time as compared to the greyscale image in training. Different pre-processing tasks are considered in order to make an object or image noise-free. The leaf image is cropped in image clipping to obtain an inclined region of the image. The image is smoothened using the Smoothing Filter.

c. Image Segmentation: This stage emphasizes on segmenting the recognized image into diverse sub-regions. The attributes available in every area are different from each other and the internal attributes of every area are similar in some context. The existing techniques to segment the image are called edge-based technique, threshold-based method and region-based method [5].

d. Feature Extraction: To extract the attributes is the major stage in the process of recognizing the image. This stage is implemented to extract the information about an image which is varied due to the zooming, illumination and abstraction of an image into some particular vector description. This stage categorized the image attributes in two parts: global and local in accordance with the diverse extraction ranges. The global attribute in an image is an attribute in which the overall information regarding an image is involved. The local attributes are considered as the partial attributes of an image available in a specific region [11].

e. Image Classification: It is often referred as a procedure to take an input and provide an output in form of class or a probability of having an input as a specific class. While detecting the diseases in plant, this stage concentrates on classifying the input plant image in two classes: normal or diseased. After detecting an image with disease, some techniques are implemented for further classifying these images into diverse infections. Various ML algorithms are introduced by several researchers over the years in order to classify the images. The most common ML techniques to detect the disease in plant are KNN, SVM, and ANN which are described as follow [6]:

i. KNN (K Nearest Neighbour): It is a ML technique based on instance. It is a technique that avoids all computations after classification for lazy learning. This technique is planned according to the concept that in case, a sample having k most equivalent neighbours in the feature space, the majority of samples must be a member of a particular class. Hence, the sample itself refers to this class. In general, classification task can use voting method, that is, the most visible category label in k samples is chosen as the predictive outcome. The regression is performed using averaging method, that is, the average of the actual values of k samples of the output label is used as the prediction outcome

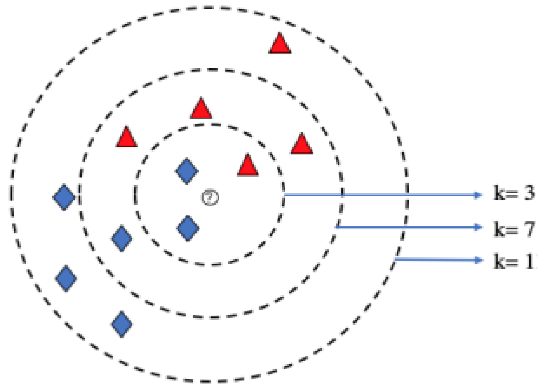

Figure 2: K-nearest Neighbour

Figure 2 demonstrates a schematic diagram of K-Nearest Neighbour, here, k is a major element. When the values of k are different, distinct outcomes are obtained in the classification process. In this, k = 3 is represented the prismatic sample judgment result, k = 7 is the outcome which is obtained as a triangle and k = 11 is used to indicate the prismatic result. Several methods of computing distance are implemented to acquire diverse NNs which provide distinct results for classifying the data. The distance can be measured for investigating the K-Nearest Neighbours [7].

The Nearest Neighbour Rule is generated when the K-Nearest Neighbour algorithm is formulated theoretically. For a scenario, let x as a point which is labelled and the discovery of point present nearby x is done and y is utilized to represent this discovered point. Nearest Neighbour Rule is responsible for creating a possible error when y is labelled same as to x and defined mathematically as:

$$P^* \leq P \leq P^* \left(2 - \frac{c}{c-1} P^*\right) \dots \dots . . (1)$$

This equation contains P^* to illustrate the Bayes ER, the number of classes is represented with c and P is ER of NN [16]. For huge number of points, a lower ER of the NN is obtained which is two times smaller than the Bayes ER. Since, the best possible optimal assignment is found lower, the double of that ER is also found minor. The availability of data points in large volume leads to offer higher probability that the label of x is same as y. The given equation consists of simpler KNN. Hence, the ER of this algorithm is calculated smaller in comparison the ER of BN algorithm.

ii. Support Vector Machine (SVM): SVM is significant ML algorithm. This algorithm is inspired from the idea of statistical learning. SVM has gained wide-ranging fame as a classification algorithm because of its exclusive benefits in managing issues with small samples, nonlinear and high dimensional data sets. It is also possible to use this algorithm for analysing data, identifying patterns, and performing regression analysis. SVM performs non-linear mapping. SVM is referred as a 2nd-class classifier. Its standard algorithm is described as the linear classification algorithm with the highest interval in feature space, which means the learning approach of SVM focusses on to maximizing the interval [8].

In the data space, hyperplanes are classified by defining the spatial classification model of points. Initially, equation 1 is used to define the classification function:

$$f(x) = w^t x + b \tag{2}$$

In equation 1, the parameters of the hyperplane to be determined are stored in w^t, while b represents the bias. Many may support vectors may be required to determine the t, w parameters which in turn are used for determining the classification of the hyperplane. Generally, the accuracy obtained after predicting the classification is a distance amid a point and the hyperplane. SVM aims to make this distance maximum. In real time, linear inseparable examples generally occur. Hence, mapping of data features to a high-dimensional space is essential. A linear inseparable mapping at high-dimensional space may enlarge the size for which its computation is made complex. This issue can be solved using the kernel function.

The kernel function is measured at lower dimensions prior to converting to a higher dimensional space, and a substantial classification impact is defined at this location. This approach may prevent complex calculations once this transformation is performed. In contrast to other kernel functions, lesser number of parameters are required by Gaussian kernel function and it also shows more flexibility. Hence, following Gaussian kernel function is used for mitigating the inseparable issues:

$$K(x, x_i) = exp\left(-\frac{\|x - x_i\|}{2\sigma^2}\right) \tag{3}$$

An SVM is generally applied to process data noise with stack variables as noise contaminates data.

iii. Artificial Neural Networks: Endeavours to artificially reproduce the biological cycles which give rise to intelligent conduct concluded in the building of ANN. It is a numerical system based on the natural neural organizations. An interconnected set of counterfeit neurons is contained in this algorithm that performs the data processing through a connectionist way [9]. Mostly, ANN is considered as an adaptable model that changes its structure dependent on outer or inside data that undergoes the network during the learning phase. Moreover, ANNs are non-linear statistical information instruments to model data. They can be utilized for the modelling of intricate connections among input and outputs or to discover patterns in data. It is a nearby replica of the natural sensory system. In this model, a neuron is utilized to multiply the inputs with loads, compute the total, and implement a threshold. The outcome of this calculation is communicated to ensuing neurons later on. Fundamentally, the Artificial Neural Network is summed up to:

$$y_i = \left(\sum_k w_{ik}\, x_k + \mu_i\right) \tag{4}$$

In this expression, x_k represents inputs to the neuron I, while w_{ik} denotes weights associated with the inputs. Also, threshold, transfer function, and the output of the neuron are represented by μ_i, f (•), and y_i respectively [10]. The transfer function f (•) may be linear, tangent hyperbolic and polynomial function. A portion of the renditions of Artificial Neural Network is contingent upon which algorithm is utilized at the abstract phase such as PNNs, GRNNs and MLP-NNs. The most usually utilized model of ANN is FFBP algorithm.

II. Literature Review

Fatma Marzougui, et.al (2020) established a CAS on DL model relied on ANN [11]. In this, CNN algorithm namely ResNet was implemented to detect the diseases of plant at initial phase. This algorithm offered higher accuracy on an augmented dataset in which images of normal and infected leaves wee comprised. The images were classified as: having disease or healthy using the established mechanism. The results depicted that the established mechanism performed well proved more efficient as compared to others for detecting the diseases occurred on plants and.

Sunil C. K., et.al (2022) suggested a U 2 -Net for eliminating the unwanted background of an input image for which multi-scale attributes were chosen [12]. The EfficientNetV2 algorithm was implemented to present a technique of detecting the disease of cardamom plant. The suggested approach was evaluated in the experimentation and compared with the traditional techniques. The experimental outcomes demonstrated that the suggested approach yielded the accuracy upto98.26% while detecting the plant diseases. The future work would aim to gather the images of cardamom diseases with nutrition deficiency to improve the dataset and to expand the suggested approach for recognizing the severity of the disease.

V V Srinidhi, et.al (2021) developed the DCNN algorithms known as Efficient Net and Dense Net for detecting the infections occurred on apple plant from the images of leaves of apple plant and classifying them into four kinds [13]. The technique of detecting the Canny Edge, Blurring and Flipping methods deployed for improving the dataset. The results indicated that the first algorithm yielded accuracy up to 99.8% and the second one offered 99.75%. The developed algorithms were useful to deal with the limitations of CNN (convolutional neural networks).

Abdul Hafiz Bin Abdul Wahab, et.al (2019) designed an image processing technique on the basis of AI to detect the diseases of a Chilli plant via images of leaves [14]. The major intend of this algorithm was to implement KMC to segment the image and compare it with SVM (Support Vector Machine) algorithm. This algorithm was adopted for extracting the computed images and classifying them into classes. Different SVM algorithms were evaluated using several metrics and kernel functions. The designed algorithm was capable of distinguishing the normal plant from the diseased one and its accuracy was found superior.

Hui Fuang Ng, et.al (2021) established a mobile application for detecting and classifying the plant infection via DL (deep learning) technique [15]. This application made the implementation of R-CNN (Region based Convolutional Neural Network) withInception-v2 backbone network for the diagnosis of diseased plants robustly and effectively. The experimental results obtained on dataset of grape disease images validated that the introduced application offered the accuracy up to97.9%. Furthermore, this application was adaptable to detect and control the plant diseases at primary phase which resulted in mitigating the losses and preventing the further spread of the disease.

Kiran Kumar Gurrala, et.al (2019) investigated a novel method to diagnose the diseases of plant for the plants on the basis of IPT and SVM algorithm [16]. The segmented image was utilized to generate the attributes to diagnose the diseases subsequent to process the image of leaves having disease. The modification of CPDA was exploited to extract the attributes. A dataset consisted of 100 images of infected leaves was applied to train SVM for recognizing the diseases such as anthracnose, leaf spot, leaf blight, scab. The results revealed that the investigated technique was performed well as compared to the classic method to detect the plants infections.

Rajiv Kumar, et.al (2021) projected a system in which a classic smart phone for predicting the diseases of plant with the implementation of ML technique [17]. This system concentrated on gathering the data as images of plant disease and a dataset was generated for detecting several diseases occurred on plants and crop. The projected system was effective to detect the diseases at premature phases for preventing the productivity. The plant diseases and the crop kinds were detected by training the NN (Neural Network) framework. The accuracy attained from the projected system was calculated 96.78% to detect the diseases.

Ajay Kumar, et.al (2022) discussed that it was essential to diagnose and cure the diseases occurred on crops at initial phases for maximizing the crop production [18]. CNN (Convolutional Neural Network) Adam optimization algorithm was intended for diagnosing and determining the diseases in plants on the basis of their leaves. Several metrics including batch size, dropout, and the number of epochs were utilized to compute the efficacy of the intended algorithm. The results confirmed that the intended algorithm yielded a superior accuracy up to96.77% as compared to other models.

Eisha Akanksha, et.al (2021) recommended an effective approach to diagnose the diseases of maize plants automatically [19]. Initially, this approach aimed to convert the images into colored format and to remove the noise in images. Subsequently, the R band was exploited for extracting the attributes. The classification algorithm named OPNN employed the chosen attributes for classifying the image as healthy or diseased. AJO (artificial jelly optimization) algorithm was implemented to enhance the PNN. Eventually, the segmentation phase deployed the Northern leaf blight disease leaf images for separating the infected region of a leaf. Several metrics such as accuracy, sensitivity, and specificity were considered to compute the recommended approach. The recommended approach offered the accuracy up to 95.5%.

M. Nikhitha, et.al (2019) formulated tool for assisting the farmers in diagnosing the diseases in advance [20]. The fundamental goal of this tool was to generate a tool to identify the level of the ailment and grade them according to it. The CNN made the implementation of Inception algorithm to classify the disease. The TL (transfer learning) method was employed for retaining it. The results indicated that the formulated DNN (deep neural network) algorithm was applicable in detecting the disease in vegetables and plants and completely effective for the agricultural industry. Kapoor N. et.al., (2021) have discussed decision tree & Random Forest algorithms that we used in predicting heart disease. The author used hybrid Classifier to predict the heart disease [21]. Chawla N. et. al., (2020) have discussed Deep Learning and machine Learning model for diagnosis of Musculoskeletal abnormalities. They have tested their proposed scheme on MURA data set [22].

III. Prepare Your Paper Before Styling

The concept to detect the plant disease is planned on the basis of the recognition of the diseased portion of image of leaf utilized for input. This paper introduced a mechanism for detecting the plant disease in which distinct stages are included. These stages are discussed as:

1. Pre-processing: - The primary phase for detecting the diseases occurred on plant is to pre-process the image. The system makes the deployment of a dataset which is generated through a reliable data source for input. This dataset is known as plant village which is an official website consists of information related to the plants and their diseases. The images of potato leaves are comprised in this dataset. There are 3 portions of dataset in which healthy leaves, leaves infected with early blight disease and the leaves infected due to late blight disease. The transformation of existing RGB format image is done into gray scale to perform additional processing.

2. Segmentation: - This stage is executed to divide the image into various segments. The image is segmented for identifying the objects or acquiring the information from the images of plants. The complexity of analyzing an image is mitigated using this procedure. The boundary line of pictures and objects is investigated. A label is assigned to every pixel in an image for which distinct features of the pixels contained in same label are shared. The objects are classified according to a set of features using KMC. This stage focuses on diminishing the sum of the squares of the distance amid the object and the equivalence cluster for classifying the object. Three is considered as a higher value of k in input. The value of K assisted in generating the segments of images. The required area is chosen after predicting the disease from the input leaf.

3. Feature Extraction: - The attained results are presented as ROI. Hence, this procedure is executed for extracting the attributes from this area. This process is executed to extract a group of values or features from an image. The considerable information about an image is acquired using the extracted features. Consequently, the further process becomes easy. The color, texture, etc. are common features for recognizing the

diseased portion of plants. In recent times, several methods for extracting the features are present. These techniques are useful for generating a system. Gray Scale Co-occurrence Matrix is a statistical technique utilized with the objective of extracting the texture feature.

The symptoms are detected at initial phase after extracting the features such as contrast, IDM, Entropy, etc. These attributes are defined as: -

Contrast: It is utilized for evaluating the local variations within an image.

$$Contrast = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j) \right\}, |i-j| = n \qquad (4)$$

This component assists in supporting inputs from $P(i,j)$ ahead of the diagonal, i.e., $i \neq j$

Homogeneity: It is a component of the nearness of distributing the elements in the Gray Level Co-occurrence Matrix to the GLCM diagonal

$$\sum_i \sum_j \frac{P_d[i,j]}{1+|i-j|} \qquad (5)$$

Local Homogeneity, Inverse Difference Moment (IDM):

$$IDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} P(i,j) \ldots.(6)$$

The homogeneity of the image impacts the IDM due to the least contributions given to the weighting factor $(1+(i-j)^2)^{-1}$ IDM from inhomogeneous areas $(i \neq j)$. Consequently, the IDM value is found lower for images without homogeneity, and a relatively superior for images having homogeneity.

Entropy: This measure is related to the informative content. The randomness of distributing of intensity distribution is quantified using entropy. The 1st order entropy is least in inhomogeneous scenes and greater for a homogeneous scene.

$$-\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \times \log\big(P(i,j)\big) \ldots.(7)$$

Correlation: This attribute is employed for computing the Gray level linear dependence among the pixels at the particular locations proportionate to one another.

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P \frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \ldots.(8)$$

Sum of Squares, Variance:

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-\mu)^2 P(i,j) \qquad (9)$$

This attribute emphasizes on assigning higher weights on the elements which can be differentiated from the average value of $P(i,j)$.

4. Classification of Data: - The last phase is to create a framework to detect the disease of plant. The training and test set are two sections of dataset. The 60% of the data is employed in the initial section and the rest of the data is employed in the latter section so that the disease is classified. KNN (K-Nearest neighbour) algorithm is planned on the basis of instance. This classification algorithm deploys similarity functions to relate the indefinite samples to the unknowns at individual level. This algorithm is unable to learn quickly. The analysis and development of this approach can be performed simultaneously. The k-nearest centres are taken into consideration to allocate the superior portion of class to the indefinite case. This algorithm is simple. The data is classified in accordance with the majority vote and its k neighbours. This algorithm is suitable to deal with the issues related to classify the data. The RF is an ML approach that performs flexibly and quickly. The tree predictors are incorporated in this algorithm. The outcomes attained from this algorithm are found promising in all of the cases. Diverse kind of data present in numeric, binary and nominal form is handled effectively using this algorithm. RF concentrates on generating diverse trees. The integration of these trees offers outcomes at higher precision. To classify the data is the major task in ML. The hyper-parameters having similarity with DTs are comprised in it. The RTs (random tees) are overlapped in this algorithm, and it is easy to perform analysis. To illustrate, seven RTs contains the information regarding some attributes. 4 of these trees are agreed and rest trees are not. The majority voting is considered to generate the ML (machine learning) algorithm. This algorithm generates outcomes with more accuracy via random subset of attributes on the dataset. The output achieved from the RF and KNN is taken as input in voting classifier for voting amid 2 classification algorithms and providing optimal results.

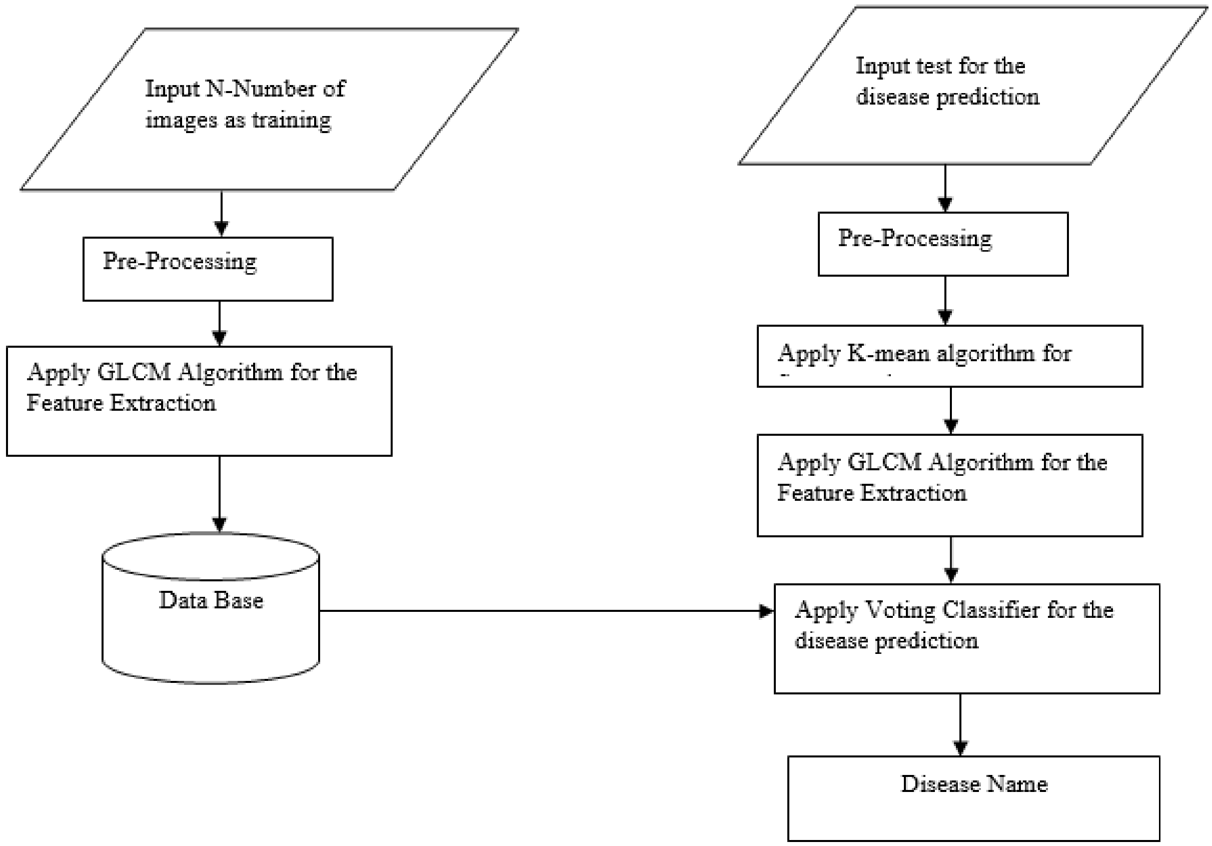

Figure 3: Flowchart of Proposed Approach

IV. Result and Discussion

Plant Village dataset is employed to implement the introduced technique. It is public dataset in which information about plants and their infections is comprised. A specific label is allocated to every image in this dataset that can represent the kind of disease. The images of potato leaves are utilized in the testing phase. This dataset helps in classifying the images in three classes: images having early blight disease, images infected with late blight disease and the healthy images. The performance is evaluated according to various metrics such as recall, accuracy, and precision. These metrics are discussed as:

A. Accuracy: It is a ratio of no. of samples that are categorized correctly to the total samples of a program i. The parameter is mathematically expressed as:

$$A_i = \frac{t}{n} \cdot 100 \qquad 10$$

In which, t is total samples that are classified accurately and n is the total samples

B. Precision: This metric is defined as TP cases divided by the total number of cases declared as positive.

Precision= TP/TP+FP...(11)

C. Recall: This metric is found by dividing the TPs with the total number of positive cases.

Recall= TP/TP+FN...(12)

Figure 4: Input pictures

Figure 4 represents the deployment of a plant image as input to detect the plant infections. This image is taken from the database. The diseases of plants are detected in 4 stages, such as to pre-process an image, segment the image, extract the attributes and classify the picture.

Input Image (a)

Segmented Image (b)

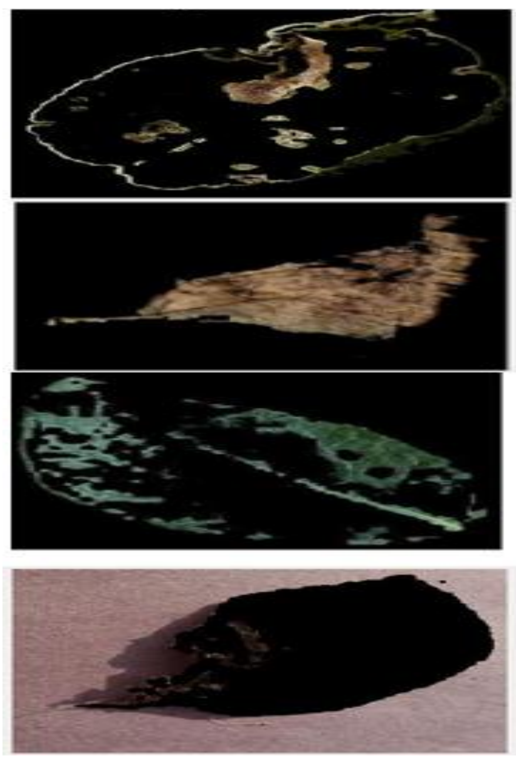

Figure 5: Segmentation of Images

Figure 5 illustrates the images which are employed for input to detect the disease. The picture is segmented via the KMC algorithm. The output, obtained after segmenting the image, is taken as ROI (region of interest) which is chosen to perform the classification.

Figure 6: Input image of potato leaf the output image is the disease affected with early blight

Figure 7: Input image of potato leaf the output image is the disease affected with early blight

Figure 8: Input image of potato leaf the output image is the disease affected with late blight

The attributes are extracted from the image with the implementation of GLCM algorithm. This algorithm is effective to extract the textural attributes of the image and these attributes are considered as symptoms of the disease. The figure number 5, 6, 7 represented that the input image used for input and Gray Level Co-occurrence Matrix algorithm is considered for extracting the attributes of the image in order to detect the diseases such as early blight or late blight. The voting classifier is adopted to detect the kind of disease in which SVM, KNN and DT algorithms are integrated. This classifier method will capable of enhancing the accuracy by 95% in comparison with the existing techniques namely SVM.

TABLE I. OUTPUT OF SVM CLASSIFIER

Disease Sample	No. of images in training	No. of images in tests	Accuracy	Precision	Recall
Early Blight	6	4	84.32	84.18	85.0
Late Blight	6	4	84	85	84.87
Healthy	6	4	85.67	84.78	85.56

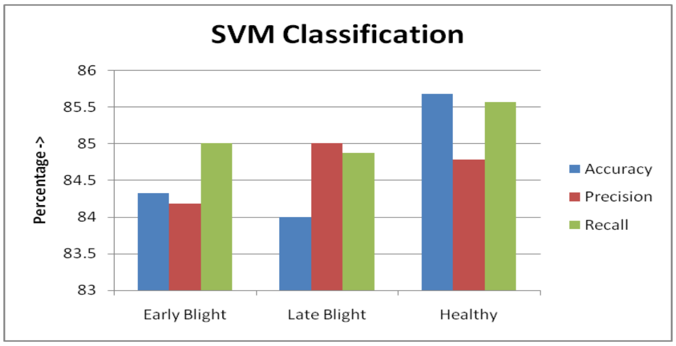

Figure 9: SVM Classification Results

The figure 9 displayed that SVM (Support Vector Machine) algorithm is adopted for predicting the disease. The table 1 and figure 10, represented the results concerning accuracy, precision and recall of every class known as early blight, late blight and healthy.

TABLE II. OUTPUT OF PROPOSED SCHEME

Disease Sample	No. of images in training	No. of images in tests	Accuracy	Precision	Recall
Early Blight	6	4	95.12	95.10	95.0
Late Blight	6	4	95	95	95.17
Healthy	6	4	95.20	95.10	95.20

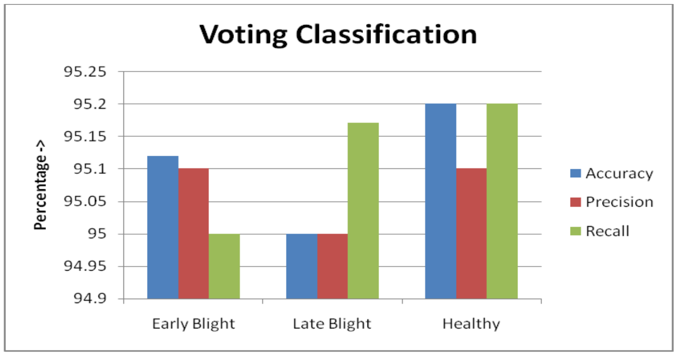

Figure 10: Voting Classification Results

TABLE III. OVERALL COMPARISON RESULTS

Parameters	SVM Classifier	Voting Classifier
Accuracy	85 percent	95 percent
Precision	84.78 percent	95.05 percent
Recall	84 percent	95 percent

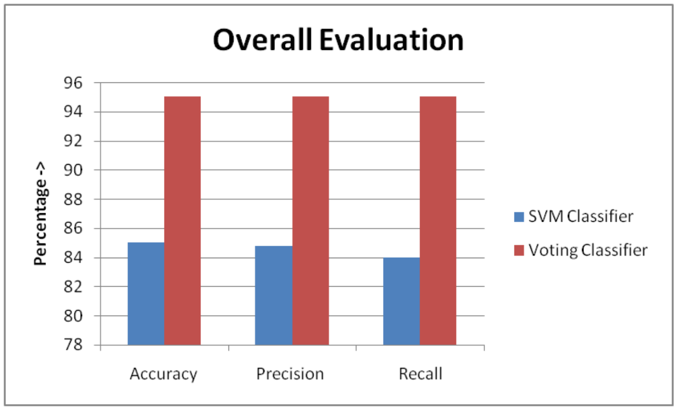

Figure 11: Result Evaluation

The figure 11 demonstrates the comparison of results acquired from SVM (Support Vector Machine) and the voting classification algorithm. For this, precision, recall and accuracy are considered. The computation of outcomes is done for every class such as early blight, late blight and healthy. The voting classifier enhances the efficacy up to 10% in contrast to other algorithm to detect the diseases of plant.

V. CONCLUSION

This work is emphasized on detecting the plant diseases. A generic model to detect the plant disease detection is introduced in which 4 stages are contained such as to pre-process the image, extract the attributes, segment the image and classify the disease. The infections of plants were detected using microscopes in past decades. However, this procedure is impractical because it is not easy task to monitor the leaves of plants using such technique. The existing method employs GLCM to extract the attributes. The input images are segmented using KMC (K-Means Clustering). A voting classification algorithm is adopted in the introduced technique instead of SVM (Support Vector Machine). The introduced technique is evaluated with respect to 3 parameters. The outcomes depict that the introduced technique enhances the accuracy and FPR around 10% as compared to the traditional technique.

REFERENCES

[1] P. K. V, E. G. Rao, G. Anitha and G. K. Kumar, "Plant Disease Detection using Convolutional Neural Networks," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1473-1476, doi: 10.1109/ICOEI51242.2021.9453045.

[2] S. V. Militante, B. D. Gerardo and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2019, pp. 579-582, doi: 10.1109/ECICE47484.2019.8942686.

[3] B. Chellapandi, M. Vijayalakshmi and S. Chopra, "Comparison of Pre-Trained Models Using Transfer Learning for Detecting Plant Disease," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 383-387, doi: 10.1109/ICCCIS51004.2021.9397098.

[4] M. Sardogan, A. Tuncer and Y. Ozen, "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm," 2018 3rd International Conference on Computer Science and Engineering (UBMK), 2018, pp. 382-385, doi: 10.1109/UBMK.2018.8566635.

[5] S. S. Kumar and B. K. Raghavendra, "Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 313-316, doi: 10.1109/ICACCS.2019.8728325.

[6] U. B. Korkut, Ö. B. Göktürk and O. Yildiz, "Detection of plant diseases by machine learning," 2018 26th Signal Processing and Communications Applications Conference (SIU), 2018, pp. 1-4, doi: 10.1109/SIU.2018.8404692.

[7] N. Prashar, "A Review on Plant Disease Detection Techniques," 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), 2021, pp. 501-506, doi: 10.1109/ICSCCC51823.2021.9478132.

[8] K. Ahmed, T. R. Shahidi, S. M. Irfanul Alam and S. Momen, "Rice Leaf Disease Detection Using Machine Learning Techniques," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 2019, pp. 1-5, doi: 10.1109/STI47673.2019.9068096.

[9] J. K. Kamble, "Plant Disease Detector," 2018 International Conference On Advances in Communication and Computing Technology (ICACCT), 2018, pp. 97-101, doi: 10.1109/ICACCT.2018.8529612.

[10] L. S. Puspha Annabel, T. Annapoorani and P. Deepalakshmi, "Machine Learning for Plant Leaf Disease Detection and Classification – A Review," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0538-0542, doi: 10.1109/ICCSP.2019.8698004.

[11] F. Marzougui, M. Elleuch and M. Kherallah, "A Deep CNN Approach for Plant Disease Detection," 2020 21st International Arab Conference on Information Technology (ACIT), 2020, pp. 1-6, doi: 10.1109/ACIT50332.2020.9300072.

[12] S. C. K., J. C. D. and N. Patil, "Cardamom Plant Disease Detection Approach Using EfficientNetV2," in IEEE Access, vol. 10, pp. 789-804, 2022, doi: 10.1109/ACCESS.2021.3138920.

[13] V. V. Srinidhi, A. Sahay and K. Deeba, "Plant Pathology Disease Detection in Apple Leaves Using Deep Convolutional Neural Networks : Apple Leaves Disease Detection using EfficientNet and DenseNet," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1119-1127, doi: 10.1109/ICCMC51019.2021.9418268.

[14] A. H. Bin Abdul Wahab, R. Zahari and T. H. Lim, "Detecting diseases in Chilli Plants Using K-Means Segmented Support Vector Machine," 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), 2019, pp. 57-61, doi: 10.1109/ICISPC.2019.8935722.

[15] H. F. Ng, C. -Y. Lin, J. H. Chuah, H. K. Tan and K. H. Leung, "Plant Disease Detection Mobile Application Development using Deep Learning," 2021 International Conference on Computer & Information Sciences (ICCOINS), 2021, pp. 34-38, doi: 10.1109/ICCOINS49721.2021.9497190.

[16] K. K. Gurrala, L. Yemineni, K. S. R. Rayana and L. K. Vajja, "A New Segmentation method for Plant Disease Diagnosis," 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), 2019, pp. 137-141, doi: 10.1109/ICCT46177.2019.8969021.

[17] R. Kumar, "Deep Learning to Detect Plant Diseases," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 674-677, doi: 10.1109/ISPCC53510.2021.9609389.

[18] A. Kumar, V. Bali and S. Pandey, "Improving the efficiency of Plant-Leaf Disease detection using Convolutional Neural Network optimizer-Adam Algorithm," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022, pp. 362-367, doi: 10.1109/Confluence52989.2022.9734132.

[19] E. Akanksha, N. Sharma and K. Gulati, "OPNN: Optimized Probabilistic Neural Network based Automatic Detection of Maize Plant Disease Detection," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1322-1328, doi: 10.1109/ICICT50816.2021.9358763.

[20] M. Nikhitha, S. Roopa Sri and B. Uma Maheswari, "Fruit Recognition and Grade of Disease Detection using Inception V3 Model," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 1040-1043, doi: 10.1109/ICECA.2019.8822095

[21] Kapoor, N., & Singh, P. (2021). Heart Disease Prediction Using Decision Tree and Random Forest Classification Techniques. In *Applications of Big Data in Large-and Small-Scale Systems* (pp. 234-259). IGI Global.

[22] Chawla, N., & Kapoor, N. (2020). Musculoskeletal Abnormality Detection in Humerus Radiographs Using Deep Learning. *Rev. d'Intelligence Artif.*, *34*(2), 209-214.