Journal of Research Administration
Society of Research Administrators International

MACHINE LEARNING APPROACHES INCORPORATING EPCA AND ESVM FOR AUTOMATIC CLASSIFICATION OF PLANT LEAF DISEASE

Mrs.R. Dhivya
Research Scholar (FT), Department of Computer Science, Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore-49, divgopal2@gmail.com

Dr. N. Shanmugapriya
Associate Professor & Head, Department of Computer Applications (PG), Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore – 49, spriyanatrajan@gmail.com.

ABSTRACT

Farming ensures that all people will have enough to feed however if the world's population suddenly expands. It's also advisable to anticipate vegetation infections in their preliminary phase in the sector of agriculture is crucial to accommodate the fruits and vegetables to the wider public. Though it is problematic to identify the infections at the preliminary phase of the plants. The purpose of this study is to educate farmers on modern methods that may be used to lessen the prevalence of plant-leaf diseases. In this research, an automated "Leaf Disease Detection (LDD)" framework is developed with novel approaches of "Image Processing (IP)" and "Machine Learning (ML)" for identifying the type of diseases in the leaf. The proposed LDD uses a direct image of a leaf as its input. After the source image has been preprocessed to eliminate unwanted noise, the denoised image of the leaf was transmitted for the segmentation process where the "Region of Interest (RoI)" is segmented, then the segmented leaf image is fed into the feature extraction module. The primary objective of this Feature Extraction is to extract leaf characteristics from image data to transform it into a format that enables similarities amongst leaf images. In this research, we propose an "Enhanced Principal Component Analysis (EPCA)" for extracting the features from the segmented leaf image, that uses respectively statistical and directional features to identify disease categories in a leaf image. After extracting the features it moves on into the feature selection module to select optimal features, then the selected feature subset will move into the classification module. For classification, we propose an "Enhanced Support Vector Machine (ESVM)" framework with a "Weighting Kernel" to classify the leaf images into the following categories "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The performance tests for both the proposed "ESVM" classifier as well as the current "Bacterial Foraging Optimization based Radial Basis Function Neural Network (BRBFNN)" and "Advanced K-Nearest Neighborhood (AKNN)" classifiers have been evaluated. The parameters include "Accuracy", "Recall (Sensitivity)", "Precision", and "F-measure". Based on the findings, the proposed classifier outperforms the existing classifiers.

Keywords: LDD, PCA, SVM, AKNN, BRBFNN

1. INTRODUCTION

Among the most essential activities in agriculture has been the timely diagnosis of plant diseases [1]. Leaves sometimes get biotic infections from a wide variety of species, including but not limited to "Bacteria", "Viruses", "fungus", "Nematodes", "Beetles", "Spiders", and "herbicides". Abiotic infections, on the other hand, are spurred around by issues like climate (dryness, warmth, freezing, heavy winds, floods), water (not enough), and nutrients (not enough) [2]. Whenever the thresholds for such severity of these diseases are surpassed, they may cause large losses in agricultural quality and production which can have a major financial impact [3].

Nonetheless, genetics had already gradually created accessible genotypes with enhanced disease resistance over time. The climate changes cause biotic and abiotic pathogens to co-occur, and the problem of production and reliability declines is still critical on a worldwide scale. Manually, visual feedback assessment is still widely used for most assessments, however, it could be difficult to determine the exact kind of disease present [4].

Farm owners do seem to be using their bare eyes to check their crops, a task that requires keen eyesight, specialized knowledge, and years of training. A few among them include recommendations backed by fundamental principles and supporting resources like photos or documents which may help anyone recognize signs and illness patterns that let help figure out the distinction among both abiotic and biotic illnesses, as well as the likely cause and cure to adopt. Some farmers may need expert assistance in making a proper, thorough assessment [5].

Every one of the above approaches takes a significant amount of skill and money, making them impractical for big farms and prohibitively costly for small ones. During fertilization during the growth of the initial 4 or 6 leaves, i.e., precisely where it might be most appropriate to counteract them, identifying leaves of any morphology, such as wide leaves or grasses, is challenging [6].

Because of this difficulty, LDD seems to be more crucial than ever, and scientists are scrambling to develop tools that provide more precise diagnoses [7]. In particular, innovative, budget, and sustainability techniques and methods to assist farmers throughout their everyday work are required due to rising community awareness for nature stewardship as well as the requirement for better productive farming to deal with a population boom and declining land availability. It is precisely in this environment that ML approaches have the potential to engage together in the transformation toward the prompt reduction of microorganisms hazardous to plants, hence limiting the need for costly and environmentally unfriendly pesticide remedies as well as other types of treatment [8].

It is now widely accepted that ML approaches will play a crucial role in facilitating this transformation. There has been a lot of work done on this issue over the years, with many different approaches centered on the utilization of specialized imaging systems like a thermos and stereoscopic images, colored and intensity images, or either optical imaging combined alongside arbitrary image analysis processes [9].

Those cutting-edge imaging techniques could one day allow precise farmers to make educated choices about the cultivation of certain, high-value harvests. Conventional RGB imagery may be preferred, although, for the widespread implementation of perception strategies for combating diseases in plants even within limited resources and lower socioeconomic regions of the globe. Multiple analysis directions have been inspired by advances in ML including their superior classifying skills on common images [10].

Problem Statement: Traditional ML methods have had greater experience in detecting and evaluating plant diseases, but these methods are restricted to picture segments, extraction of features, and pattern matching. Features cannot be extracted from a massive volume of data through conventional methods, also including subjective identification and ML. Standard IP techniques work effectively when there aren't a large amount of training data to analyze. Convergence of surrounding and target areas, determining similarities, and retrieved shapes as negative aspects of LDD. Therefore, fresh methods are required to successfully solve the existing challenges.

Paper Contribution: To identify and categorize leaf diseases, this research proposes a novel ESVM method. The stages of image acquisition, noise removal, image enhancement, image segmentation, feature extraction, data classification, and statistical modeling are all accounted for in the proposed model. For feature extraction, we propose an EPCA approach for extracting the features from the segmented leaf image, that uses respectively statistical and directional features, and for classification, we propose an enhanced version of SVM with weighting kernels to classify the leaf images into the following categories "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves".

Paper Organization: Section 2 lists some recent publications related to LDD with advanced computing techniques, Module-by-module explanations of the existing and proposed methods are presented in Section 3, and Section 4 gives a research finding of the classification module with a comparative of both existing BRBFNN, AKNN and proposed ESVM classifiers, and the paper comes to a conclusion in Section 5, which also discusses future direction.

2. RELATED WORKS

The researchers of [11] proposed a "Neural Network (NN)" based online automatic leaf illness edge detection technique. Their proposed architecture had 4 "Radial Basis Functions (RBF)". Mango leaves have been captured first instantaneously employing an internet camcorder. Next, features have been retrieved from the images after they had been preprocessed with a "Scale-

Invariant Feature Transform (SIFT)" method. Finally, the training of the NN was therefore optimized by making the most of its highly distinguishable features using the "Bacterial Foraging Optimization" method. RBFNN then was employed to remove the affected areas from the images of mango leaves. The results of the tests showed that their proposed approach was very accurate in classifying the different types of fungal illnesses.

The researchers of [12] suggested a NN with ML to improve the speed and accuracy with which infections are diagnosed. The images of diseased leaves were used as training information which was gathered through a classifier model. To diagnose mango plant disease, an ML based model was developed to automatically import and compare new images of the diseased leaf using a training set. Their proposed approach was capable of accurately identifying and classifying the tested disease with quite a classification performance of 80%. By using an ML model instead of a purely manual one, the innovation has helped in disease identification even without the intervention of farmers, hence increasing efficiency. In addition to increasing mango production and satisfying consumer demand throughout the globe, this method also would effectively treat the affected mango plant diseases.

In an attempt to combine the positive characteristics of the "Inception Module (IM)", the researchers of [13] suggest a "V2IncepNet". Higher dimensionality feature retrieval, as well as image classification, are performed by the IM, while the "VGG-Net" layer is responsible for the basic extraction of features. IM performs higher dimensionality feature retrieval as well as image classification, while the "VGG-Net" layer isolates basic features. The usage of specific color characteristics is widespread throughout this work. It was determined that the proposed model has a precision of 92% or higher in classifying the degree to which mangoes leaves are infected with Anthracnose infection. Their proposed framework was simple, but it worked well.

To detect premature leaf tissue disorders with microscopic spots, which are only visible in greater resolution pictures, in [14] the researchers designed an ANN approach. An image enhancement approach was used to segregate all contaminated objects throughout the entire database. Then they implemented a wrapper relying approach for selecting features relies on dual metaheuristics, they first generated a set of candidate features for measuring the blobs, and afterward choose features from that set according to its influence upon that model's efficiency. Inputs consisting of the identified factors have been transmitted to the ANN. Their dataset contains images including both normal and sick leaves. Their proposed framework seems to have greater accuracy in classification than existing approaches.

To identify the mangoes Anthracnose infection, the researchers [15] used a novel CNN framework. An actual database has been acquired across farmland in "Maharashtra, Karnataka, and Delhi" for the testing process. There are images including both diseased and healthy mangoes leaves. It was determined that their proposed model has a precision of 95%.

3. METHODOLOGIES

The automated identification and accurate intervention of leaf disease play a major significance in increasing the development and sustainability of numerous plants using IP and ML technologies. In this research, a multitude of IP approaches is used to identify illnesses in plant leaves. These include preprocessing techniques, segmenting the disease region, extracting the statistical features, selecting the optimal features, and finally classifying the disease type [16]. Figure 1 illustrates the framework structure of the proposed LDD technique.

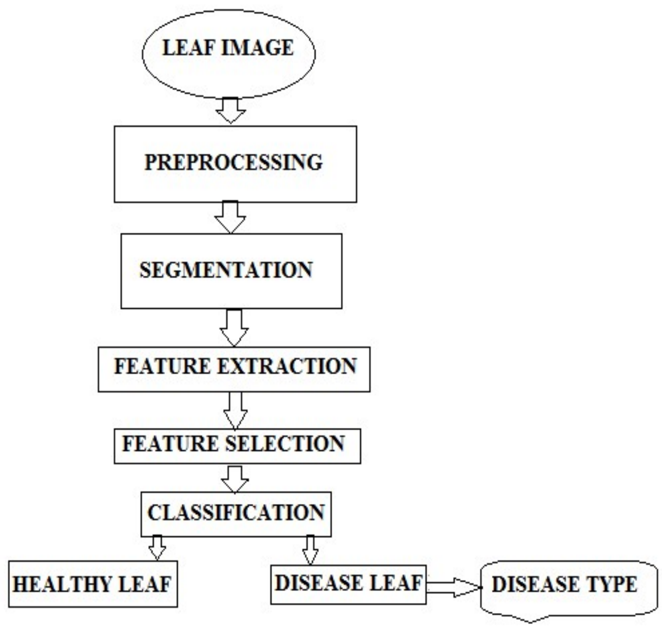

Figure 1: Proposed LDD Model's Structure

3.1 PREPROCESSING

Performing a form of pre-processing over an image before attempting to compute it could help reduce some of the distortions and noises. The "K-SVD DWT" is used as preprocessing technique in this research to make a more comprehensive dictionary scheme by improving image denoising, and also we use this method to achieve faster image denoising than standard K-SVD [17]. Here it is important to see how to apply the grayscale method to color pictures. To solve this challenge, we implement the K-SVD method to each of the three channels R, G, and B separately. The color objects are the results of this simplistic approach. They are attributed to the assumption that there is still a strong association between channels in realistic pictures. The method would be extended using DWT to column vectors, which have been the summation of the R, G, and B values, in an attempt to achieve the right colors. The dictionary can be updated more often as a result of the method's ability to learn color channel correlations. Through this, a preprocessed image will proceed with the segmentation process.

3.2 SEGMENTATION

Segmentation of leaf image operates by identifying features and borders (paths, circles, and so on) and applying labels to each pixel so that pixels with about the identical label have similar visual features. Segmentation of a preprocessed leaf image produces a series of regions that together comprise the whole leaf image, thus every pixel in a region is identical in terms of a feature or derived properties including color, density, or texture. When it comes to almost similar features, regions that seem to be similar to one another vary greatly. Following the pre-processing step, segmentation has been implemented to distinguish the ROI of the leaf, along with its background. In this research, for the segmentation process, we proposed an AKMC "Advanced K-Means Clustering" technique.

3.3 FEATURE EXTRACTION

The EPCA method is proposed in this research to scale down a dataset while keeping its uniqueness unchanged. Knowledge is often found in data sources as vectors of experimental parameters having "values (integer, binary, or real numbers)". Each vector of 3 parameters, that are oriented with each of the 3 "coordinate-axis (x, y, and z)", could be used to characterize a geometrical position in a 3-D area.

It is possible to use a vector of variables to characterize a sampling precisely. The dimension of the set was established by the total parameters, which in turn establishes the lengths of the vectors inside the sampling. In addition, the spectrum of values with which every parameter tolerance might be defined by specifying a continuity of variation on every parameter.

The 3 parameters representing Cartesian coordinates, for instance, it was constrained to "[-1/2, 1/2]", assuming the data collection contains 3-D points contained inside a cube of side 1, with its center located at "(0, 0, 0)". This range characterizes the extent of heterogeneity for all 3 parameters.

EPCA's ultimate goal is to discover data that has hidden patterns and transform it into a form that one's similarities and differences become apparent. Once the patterns were identified, the knowledge could well be processed as a hierarchy of components, also with the lowest relevant points being dropped out first.

EPCA would reduce the dimensionality of a dataset whenever the descriptive parameters used to characterize the information are sufficiently related. To even further illustrate this concept, let's examine again the scenario with the 3 parameters defining the 3 locations of points within a cube of the segmented leaf image. For instance, when some of the points inside the set are on the right plane, then we could conclude that the 3 parameters have all been connected.

Because of the potency of EPCA throughout this scenario, 1 of the original 3 parameters was already changed into a null value. Thus, merely 2 parameters would be needed to characterize the new space's points, making it more compact than the original. Points that exist within a 2-D space become relevant, therefore information concerning rejected 3-D becomes irrelevant. This approach simplifies the problem quite simply. Throughout the subsequent sections, we would go more extensively into the EPCA procedure.

To convey that the segmented leaf image for evaluation contains points having coordinates "(-2, -1), (-1, 0), (0, 1), (1, 2), (2, 3)" is to suggest that the segmented image contains data from a 2-D area. The "x" may take on a range of values between "(-2) and (2)", while "y" can take on a similar range of "(-1) to (-3)". Those 2 ranges characterize the "x" and "y" parameter variations. It is possible to demonstrate a link between these two aspects. The x-coordinates increase as the y-coordinates increase, and there's a direct line connecting all the points. Therefore, it is possible to acquire another coordinate when only one among the 2 is known.

The purpose of EPCA is to transform specific parameters that were not related. It would enable the higher probability parameters to be prioritized, reducing the overall segmented leaf image size. Variability continues to be most strongly associated which was termed as "Principal Components (PC)". These PCs are typically ranked in a sequence of increasing variability, the initial PC is viewed as reflective of the findings. Within regression analysis, it doesn't matter if a PC of smaller order may reflect fewer variations inside the set.

3.3.2 EPCA's Statistical Extraction of Features

Finding a simpler way to represent information is a prominent application of EPCA. This is unique in two ways. While it processes the information, first it continuously reduces the units of reference to a more reasonable and precise range. Next and primarily, it extracts distinctive features from source information in such a way that dimensionality is decreased. In every case, the essential feature values were still there, and they were able to be utilized to determine whether source information was valid.

The "Covariance Matrix (CM)" might also be generated from matrices by using a set of features. This CM could then be used to determine the Eigen-values. The Eigenvectors have been beneficial in their purpose of expressing comprehensive sets. Only a few minor Eigen-values have been determined to be significantly preferred and more significant, meanwhile, the rest are significantly relatively modest and exhibit a low sensitivity to data fluctuations. Therefore, the preferable and higher variable solutions are effectively preserved by calculating the weighted sum of its sample more even with corresponding Eigen-vectors through its corresponding Eigen-values.

Here are some broad factors used in EPCA analysis:

- Applying this Equation (1) to the given data, find the CM.

$$\sum V = \frac{1}{Num}\left\{\left(diag(m) - \overline{diag}(m)\right)\left(diag(n) - \overline{diag}(n)\right)^{T}\right\}$$

Eq→1

Such that, "1≤ m, n ≤ Num".

- Based on the "Eigen-vector Matrix (V)" as well as "Diagonal Matrix (D)" of the derived Eigen-values, it could conclude as per Equation (2):

$$V^{-1}\sum V = D$$

Eq→2

- Its PC characteristic is obtained by sorting the Eigen-vectors into descending manner of the magnitude of their corresponding Eigen-values.
- Finally, PCs are generated from information by calculating the internal product from the input mostly with appropriate Eigen-vectors.

The Vector "v" linked mostly with Category "y" could have its EPCA determined by transferring it into the sub-regions in a specific particular way. Line up with both the lengths or separations of the "e' Eigen-vectors" which belong well to "e' Eigen-values" there at the start of the "Auto-Correlation Matrix (R)" in descending order, whereby "e'" is less than "e". The result of the preceding operation is a matrix containing "e' coefficients c1,..., ce'". The vector "v" is described using a linear representation of the Eigen-vectors and their associated weights "c1,..., ce'".

These have segmented leaf images with a resolution of "256x256" as forwarded from the segmentation module in this research. It sometimes gets "32x32" sub-bands of resolution here. Such sub-band images are most often used in EPCA to extract EPCA values. Hence the proposed EPCA function was able to determine the 13 "Statistical Features" such as "Mean", "Standard Deviation", "Entropy", "Root Mean Square", "Variance", "Smoothness", "Kurtosis", "Skewness", "Inverse Difference Moment", "Contrast", "Correlation", "Energy", and "Homogeneity".

The extraction of the features' initialization process is shown in Figure 2. The EPCA's 13 statistical features are shown in Figure 3. The 13 statistical features from the segmented image are convoluted in this research by EPCA to 3 PC template models, as shown in Figure 4.

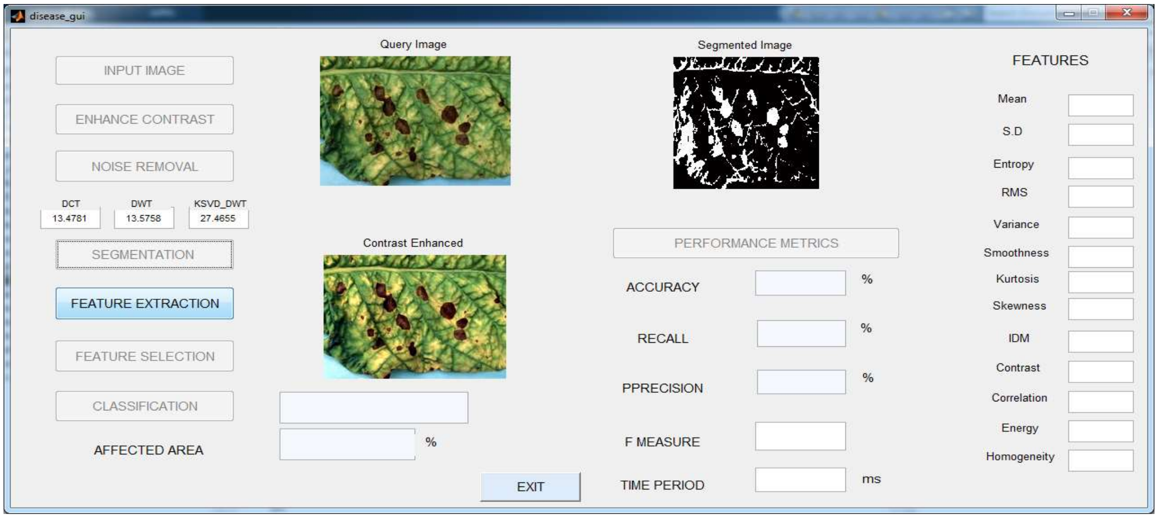

Figure 2: Feature Extraction Initialization

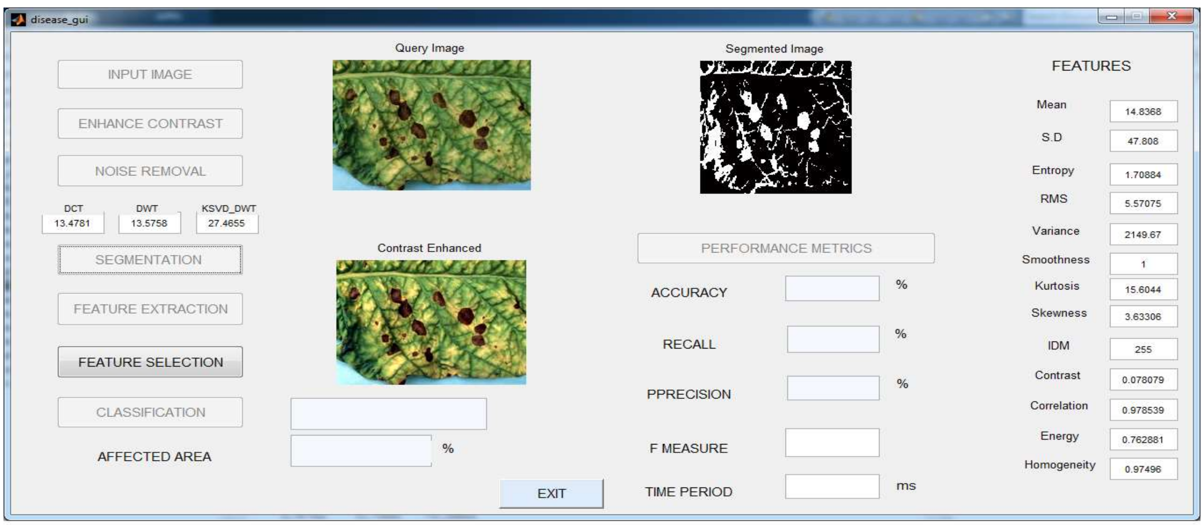

Figure 3: Statistical Features from EPCA

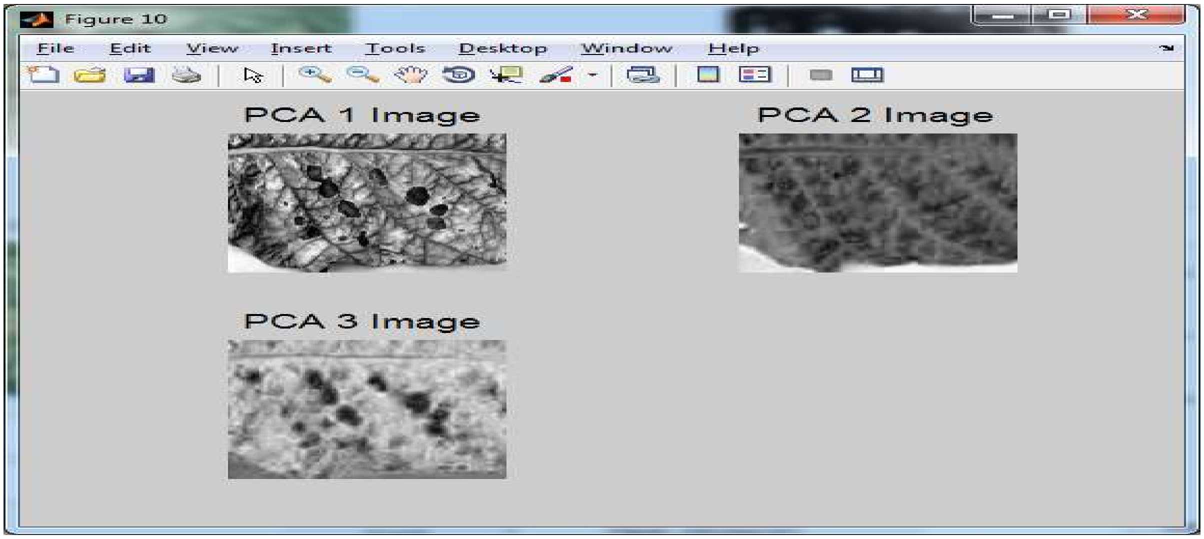

Figure 4: Principal Component Template Models

3.4 FEATURE SELECTION

Patterns are typically interpreted as a vector of feature values in traditional pattern identification approaches. The selection of the features will also have a big effect on the subsequent classification method's capability. The primary goal of selecting the features would be to lessen the number of features involved throughout classification thereby ensuring a high level of recognition rate. Lower biased features are removed, allowing a subset of the initial features with enough detail to distinguish between groups. For selecting the optimal features, many selection methods were used. "Particle-Swarm-Optimization (PSO)" is gaining popularity in the field of selecting features. In this research, the selection of features was included to minimize dimensionally and obtain the best solution thus reducing computation time. Here the PSO will choose the optimal feature subset from the overall EPCA features and it passed to the classification process.

3.5 CLASSIFICATION

3.5.1 BRBFNN CLASSIFIER

The researchers in existing already presented a technique called "Bacterial Foraging Optimization based Radial Basis Function Neural Network (BRBFNN)" for autonomously identifying and classifying plant leaves diseases [18]. Also with the support of BFO, they've trained in how to identify the diseased area by analyzing plant leaves. Using reference to proximity from the central, the RBFNN linear model has a specific competency that rises and falls steadily over time with distances. That's also effective for dealing with the complexities of the impacted area on plant leaf images. By employing the "Region Growing" approach, which searches for seed points and groups those with comparable qualities, the RBFNN's effectiveness has been significantly improved, which benefits mostly in extracting the features. Specifying the weights of RBFNN is made easier using BFO's imitating capacity and multi-optimization function. Similarly, developing connectivity that is capable of quickly and accurately identifying various areas of a plant leaves image.

3.5.2 AKNN CLASSIFIER

The AKNN "Advanced K-Nearest Neighborhood" method was proposed earlier for classifying the type of leaf diseases through the relevant information obtained from the nearest training samples [19]. The data sets are initially differentiated into the "training dataset" and the "testing dataset". The closest "K" training set elements are identified on every row in the test set and the categorization of test data with indiscriminately having shattered connections is evaluated. On either side, when the nearest Kth vector has links, all instances are taken into the class. The "Euclidean Distance (ED)" in this AKNN classification calculates and expresses the distance from the testing data vector to all the training vectors based on euclidean. It insists on taking the test instance "x" and finds in the training data for the closest K-neighbour and assigns "x" to the class which is most popular among K-neighbours. This method requires two processes: training and testing. The training process stores the vectors of the feature subset and names the samples for the training. The features of the research sample whose class is unclear are determined in the classification stage as previously. By measuring the distance again from the new vector from all stored vectors, the nearest 'K' samples are chosen. The new item is predicted, and the AKNN algorithm is recognized as the nearest training set. Big "K" values reduce the classification influence of noise. The ED used in this AKNN model is to quantify correlations between neighbors.

3.5.3 ESVM CLASSIFIER

When it comes to ML techniques that rely on classifying patterns, SVM has emerged as the go-to option. In 1995, Vapnik introduced the SVM. Typically referred to as quantitative learning approaches, this strategy provides a platform for locating information, generating predictions, and settling on a course of action. It helps in picking the best hyperplane again for the task at hand. The fundamental idea behind the method is to project the asymmetric distinguishable samples over to a higher dimensionality space using a multitude of kernel operations.

This classification algorithm can differentiate between classes by using a hyperplane inside the feature set which is modulated by a kernel factor. The isolation of extracted features towards accurate knowledge is usually problematic in the environment of higher dimensionality data.

Therefore, we suggest a fine-tuned ESVM using a kernel approach based on weights to implement a nonlinear differentiation approach. This technique translates the original information into higher dimensionality, and their translation into one whole results in uniform dispersion.

When comparing the distance between each point of data as well as the delineation points across classes, the ESVM appears to be the most optimal choice.

The proposed ESVM training as well as categorization uses a kernel characteristic for training the ESVM and statistically significant characteristics for the categorization phase, both of which have an impact on the computing distance. Sequencing these rounds improves the method's accessibility although tackling problems like massive datasets in the testing phase as well as a wide variety of characteristics, most of which are vector-based.

3.5.1 Methodology of ESVM

ESVMs are constituted fundamentally as binary classifications because of the nature of the data they processed. Moreover, it's possible that it was developed to identify the many categorization problems often associated with LDD research. Both training images and testing images are utilized to feed input into the ESVM classification algorithm. Two distinct classes emerged throughout the training stage: "Disease part (c1)" as well as "Non-disease part (c2)".

3.5.2 Algorithm for ESVM

- Mostly during the training phase, the training set including associated labeling has been delivered as feedback.
- The averaged relevance from the trained results could be subtracted from the training instances to achieve standardization.
- Finally, the decision result for each label is calculated, which yields the trained relevance.
- Applying this procedure, the decision component could be evaluated as per Equation (3):

$$d(v) = w_i L_i v_i^T v + n$$

Eq→3

- The normalized test results as well as the margin result have been computed.
- Those decision results are subsequently calculated to the qualifying values that use the following Equation (4) to establish the correct label:

$$v \in \begin{matrix} C1, & d\,v > 0.5 \\ C2, & \text{otherwise} \end{matrix}$$

Eq→4

- The results are then presented as the associated classification label.

3.5.3 Classification of Leaf's Disease Using ESVM

The SVM classification model is being utilized as the classifier in this research to classify leaf images into the following categories: "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The PSO technique has been used to build a feature subset by using the MATLAB toolbox. In a feature vector, a histogram of feature subset instances is created.

The following are the steps for using the ESVM method to characterize the leaf image:

Step-1: Create a collection of image category sets

The image has been separated into training and test samples in this step, and the images are retained for processing. It is simple to manage a vast volume of information.

Step-2: Create a subset of features

Under this step, an optimal function from the feature dictionary is chosen from represented images from every leaf category to construct a feature subset.

Step-3: ESVM may be used to train an image classifier

For this research, a multi-class ESVM classification model has been utilized to train to leverage binary ESVM with error-correcting output code. The image has been encoded using the training set. For the feature subset, the histogram is becoming a finalized feature subset.

Step-4: Classify any image or a group of images into categories

During this step, the leaf image has been labeled. This classification procedure is divided into two phases

They are:

(i) Training phase
(ii) Testing phase

Mostly in the training phase, the classifiers are trained upon extracted features to construct a classification model, with 3 being the predicted significance and "3*3" being the support vectors or chosen features extracted from most of the training images. During the testing phase, the above model has been employed to classify the test images further into predefined categories.

(i) Training phase

There are two modules in the training phase. They are:

(a) Features Finalizing

(b) Generating a Model

(a) Features Finalizing

The leaf features from its feature subset are finalized using multi-class. There are three key stages to putting this phase into action.

(i) From the feature subset, detect and extract leaf local features.

(ii) Construction of a codebook dependent on leaf characteristics.

(iii) Use the codebook to define an image in a histogram.

The process of extracting final features began with the detection and extraction of localized features. Feature identification is the method of identifying a salient or prominent image area inside a spatially confined image. After that, a distinguishing feature is extracted that characterizes a series of key points in an image.

(b) Generating a Model

These were fed into the proposed ESVM classifier, which builds the model. The model would then be tested during the training phase to guarantee that the highest potential accuracy level is achieved for each class in the database. Labels for feature subsets include "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". As a result, the training datasets are split into two parts. Eighty percent of the training data had been used to construct classification models, and indeed the residual 20 percent was being used to create test images for the developed model's assessment. The ESVM model development for the proposed LDD is shown in Figure 5.

(ii) Testing Phase

The classification model is built mostly in the testing phase, and the chosen support vectors are added to the test image inside the training phase to assign these to the predefined categories. This classification contains a range of classes, including "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The trained classification model could be effectively calculated based on this performance and recognizing which classes have been used to create each classification model.

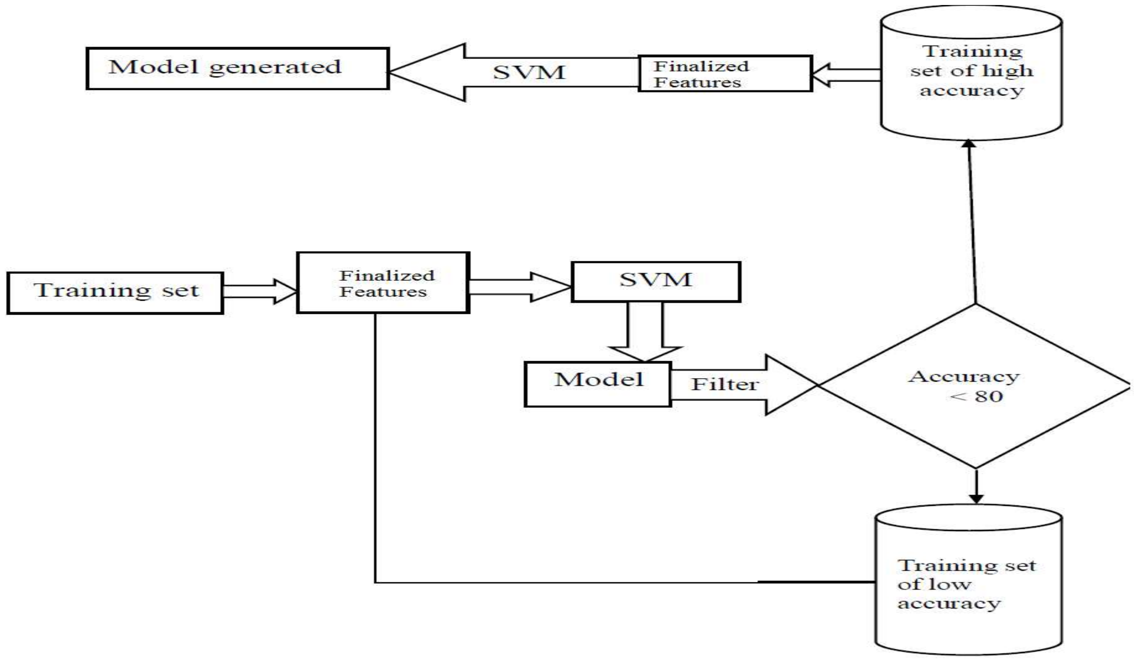

Figure 5: SVM Model Generation

4. RESULTS AND DISCUSSIONS

By selecting the appropriate measurement metric, it is necessary to evaluate the output of every computing method in an attempt to demonstrate its effectiveness, generalization potential, and accuracy of the results. The performance metrics included in this research were developed to see how consistent classification methods are with various data. Throughout this research, leaf images from all of the classes "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves" were used to evaluate the classification methods. The performance tests for both the proposed "ESVM" classifier as well as the existing BRBFNN, and AKNN classifiers have been compared and evaluated by the parameters "Accuracy", "Recall (Sensitivity)", "Precision", and "F-measure".

Dataset: The leaf image collection "Manu's Disease data set" was gathered from publicly accessible sources. These images' contrast and intensity representations vary significantly. There are numerous approaches for evaluating the efficiency of classification methods focused on the various properties of metrics.

Process: The outcomes of the existing BRBFNN method and the proposed AKNN method for the classification of leaf diseases are presented in this section. In the subsequent sections, proprietary parameters like "Accuracy", "Recall (Sensitivity)", "Precision", and "F-measure" are compared and tabulated. The simulations were developed and tested using "MATLAB R2017" and a "64-bit Windows Operating System" with a configuration "Intel Core i7 processor".

4.1 ACCURACY

Among the most popular measurement strategies for estimating system efficiency is classification accuracy. It's considered a primary standard parameter for assessing the classification system's effectiveness. The implemented system's efficiency improves as the classification accuracy improves. This metric is determined by using the Confusion-Matrix ("True Positive [TP], True Negative [TN], False Positive [FP], and False Negative [FN]") which has the advantage of simplicity as given in Equation (5).

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} * 100 \qquad \text{Eq} \rightarrow 5$$

Table 1: Accuracy Comparision

DISEASE TYPES	BRBFNN	AKNN	ESVM
Alternaria Alternata	88.71	93.82	96.53
Anthracnose	82.98	87.45	90.21
Bacterial Blight	86.97	91.45	94.21
Cercospora Leaf Spot	86.49	91.18	94.01
Healthy Leaves	89.14	94.03	96.87

Table 1 and Figure 6 show the performance of Accuracy for different classes "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The leaf images from each above class had been taken for classification and compared with the existing BRBFNN, AKNN, and the proposed ESVM classifiers to evaluate their accuracy rate. The findings prove that the accuracy level of the proposed ESVM classifier is higher for all the classes while comparing it with the existing BRBFNN, and AKNN classifiers.

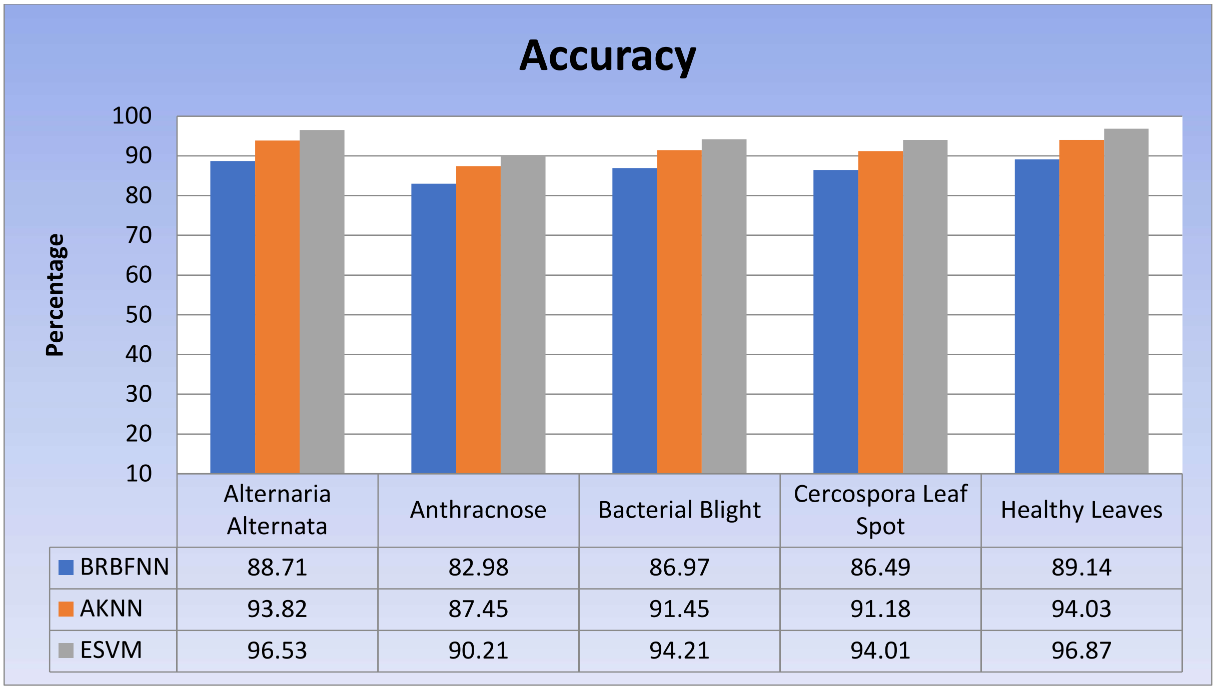

Figure 6: Accuracy Comparision

4.2 RECALL/SENSITIVITY

Using the Recall, we can estimate how many TP predictions were achieved out of all possible positive predictions. The availability of disease in the leaf image is depicted in this research analysis which has a TP value. The TN, on the other hand, shows that the disease is unavailable. The test identifies the exact results of the TP, which are proportional to the level of exposure. This research demonstrates the detection of leaf diseases, and Equation (6) essentially calculates and specifies the percentage for Recall.

$$Recall = \frac{TP}{TP+FN}$$

Eq →6

Table 2 and Figure 7 show the performance of Recall for different classes "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The leaf images from each above class had been taken for classification and compared with the existing BRBFNN, AKNN, and the proposed ESVM classifiers to evaluate their Recall rate. The findings prove that the Recall level of the proposed ESVM classifier is higher for all the classes when comparing it with the existing BRBFNN, and AKNN classifiers.

Table 2: Recall Comparision

DISEASE TYPES	BRBFNN	AKNN	ESVM
Alternaria Alternata	80.78	85.42	88.26
Anthracnose	78.09	83.25	86.12
Bacterial Blight	81.89	86.15	89.01
Cercospora Leaf Spot	81.11	86.24	89.09
Healthy Leaves	82.01	87.23	90.09

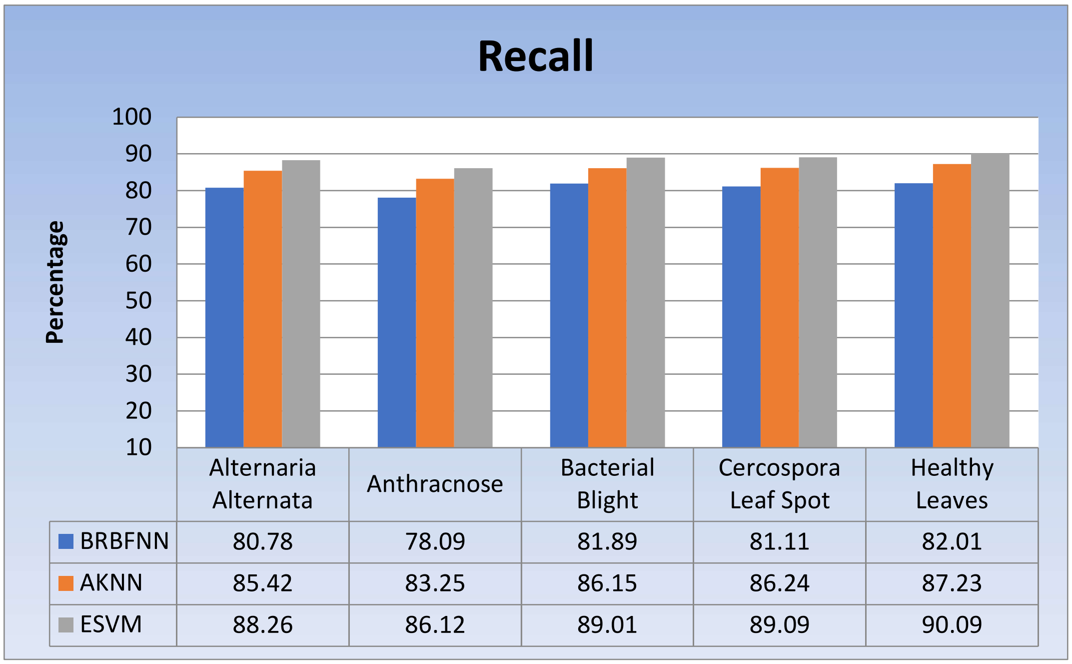

Figure 7: Recall Comparision

4.3 PRECISION

The random error measure is just what Precision is termed. TP and FP, in other terms, are also used. It's the proportion of TP's positive traits. This research compares leaf images for disease type symptoms with those leaves that are correctly classified. When the value is 0.953, it means that in a leaf image, it accurately predicts disease type 95% of the time. This also describes the data points in consideration. The Confusion-Matrix, as seen in Equation (7), is used to quantify the precision level.

$$Precision = \frac{number\ of\ true\ positives}{no\ of\ true\ positives + false\ positives} = {}^{TP}/_{TP + FP} \qquad \text{Eq} \rightarrow 7$$

Table 3: Precision Comparision

DISEASE TYPES	BRBFNN	AKNN	ESVM
Alternaria Alternata	78.17	85.89	88.67
Anthracnose	78.01	83.71	86.59
Bacterial Blight	80.14	86.89	89.71
Cercospora Leaf Spot	81.24	86.98	89.79
Healthy Leaves	82.11	87.84	90.65

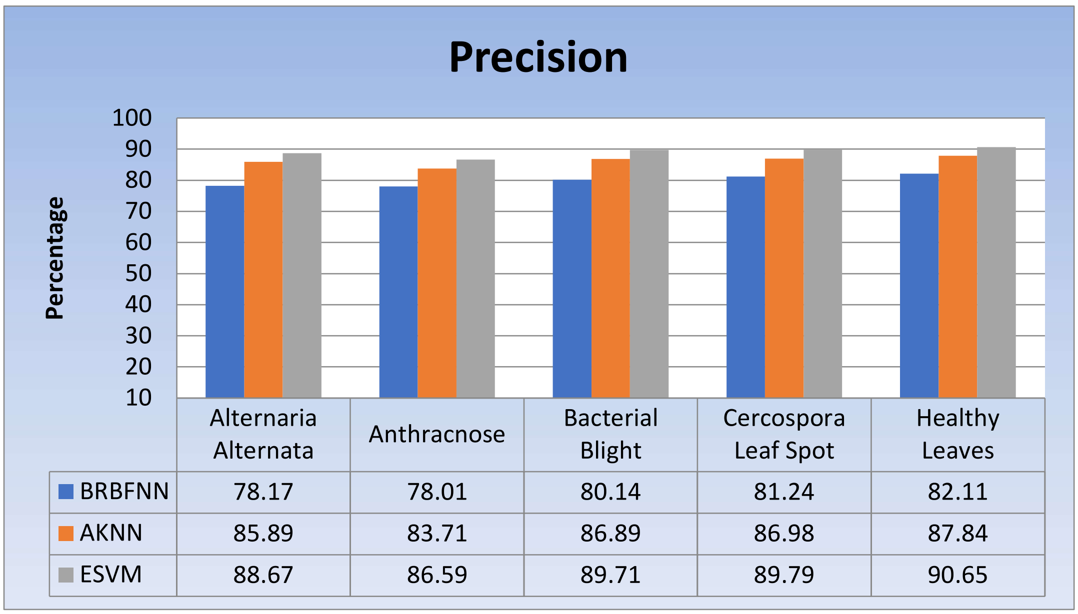

Figure 8: Precision Comparision

Table 3 and Figure 8 show the performance of Precision for different classes "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The leaf images from each above class had been taken for classification and compared with the existing BRBFNN, AKNN, and proposed ESVM classifiers to evaluate their Precision rate. The findings prove that the Precision level of the proposed ESVM classifier is higher for all the classes when comparing it with the existing BRBFNN, and AKNN classifiers.

4.4 F-MEASURE

The F-score is an indicator of the accuracy of a model in a given leaf dataset often called an "F1-score" or "F-measure". It's being utilized to test conditional structures that separate instances into "Negative" or "Positive". The F-Measure is indeed a process of combining the model of recall and precision and has been described also as a "harmonic mean of model recall and precision". The average weight of recall and precision is calculated by F1-Score. The F1-Score is at "one" or equivalent to "one" of its highest qualities. The worst meaning is found to be "zero" or similar to "zero". For the accuracy of the evaluation, the F1-score is used which encompasses all recall and precision values. The F1 score is better than one as well as worse than zero. These measures are simple and are calculated using the Confusion-Matrix as given in Equation (8).

$$F_1 = 2.\frac{1}{\frac{1}{recall}+\frac{1}{precision}} = 2.\frac{precision.recall}{precision+recall}$$

Eq →8

Table 4: F-Score Comparision

DISEASE TYPES	BRBFNN	AKNN	ESVM
Alternaria Alternata	85.71	90.82	93.68
Anthracnose	79.98	84.45	87.23
Bacterial Blight	83.97	88.45	91.23
Cercospora Leaf Spot	83.49	88.18	91.01
Healthy Leaves	86.14	91.03	93.89

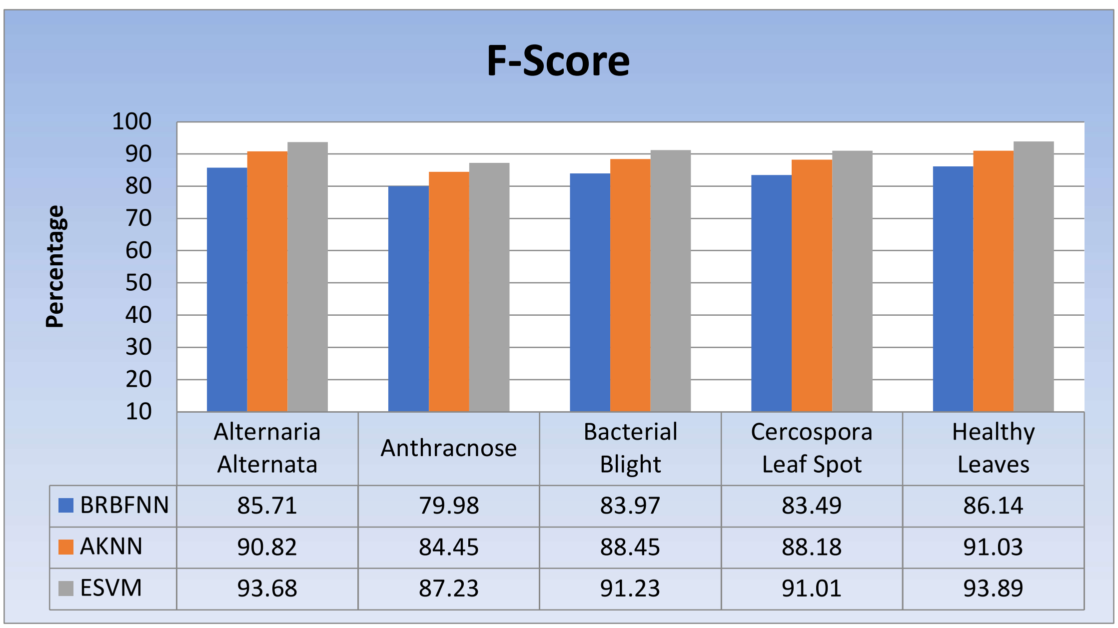

Figure 9: F-Score Comparision

Table 4 and Figure 9 show the performance of F-Score for different classes "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". The leaf images from each above class had been taken for classification and compared with the existing

BRBFNN, AKNN, and proposed ESVM classifiers to evaluate its F-Score rate. The findings prove that the F-Score level of the proposed ESVM classifier is higher for all the classes while comparing it with the existing BRBFNN, and AKNN classifiers.

5. CONCLUSION

In this research, we present a novel LDD framework by outlining its core components, which include "Preprocessing", "Image Segmentation", "Feature Extraction", "Feature Selection", and "Classification (disease prediction)". The quality of the obtained unprocessed image has been enhanced first by employing KSVD-DWT for noise removal as well as other artifacts. The images were further segmented using AKMC after they had been preprocessed. After the images got segmented, the foremost salient features have been extracted using EPCA. Then, a feature matrix is obtained by EPCA, and PSO is used to choose the best features. This information was then fed through into the identification process using a proposed ESVM approach for disease categorization. We adjusted the weights of ESVM to make it more accurate in classifying data. Both of the proposed ESVM together with the already available BRBFNN, and AKNN classification methods have their performance measured for the present study. The accuracy of the proposed novel ESVM is increased in comparison to the current methods. Altogether, the assessment indicates here that the proposed work has reached its full performance potential and so meets the requirements for LDD. Recent studies have focused on developing better methods for spotting leaf diseases so that farmers may better support their customers. By this proposed LDD model by improved extraction of features and identification using a powerful optimization strategy, it will be very useful for farmers at both lower and higher levels. This research work can be expanded in the future by a hybrid process of combining booster algorithms and the number of classification classes.

REFERENCES

[1]. Nagaraju, M., and Chawla, P. (2020). Systematic review of deep learning techniques in plant disease detection. Int. J. Syst. Assurance Eng. Manag. 11, 547–560. DOI: 10.1007/s13198-020-00972-1.

[2]. Kashef, R. (2020). Adopting Big Data Analysis in the Agricultural Sector: Financial and Societal Impacts. Singapore: Springer Singapore.

[3]. Pantazi, X. E., Moshou, D., and Bochtis, D. (2020). “Chapter 3-Utilization of multisensors and data fusion in precision agriculture,” in Intelligent Data Mining and Fusion Systems in Agriculture, eds X. E. Pantazi, D. Moshou, and D. Bochtis (Cambridge, MA: Academic Press), 103–173.

[4]. Lu, Y., and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. Comput. Electron. Agric. 178, 105760. doi: 10.1016/j.compag.2020.105760.

[5]. Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53, 5455–5516. DOI: 10.1007/s10462-020-09825-6.

[6]. Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L., and Zhao, C. (2020). A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. Remote Sens. 12, 19. DOI: 10.3390/rs12193188.

[7]. Too, E. C., Yujian, L., Njuki, S., and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. 161, 272–279. doi: 10.1016/j.compag.2018.03.032.

[8]. Liu, X., Min, W., Mei, S., Wang, L., and Jiang, S. (2021). Plant disease recognition: a large-scale benchmark dataset and a visual region and loss reweighting approach. IEEE Trans. Image Process. 30, 2003–2015. DOI: 10.1109/TIP.2021.3049334.

[9]. Ümit Atila, Uçar, M., Akyol, K., and Uçar, E. (2021). Plant leaf disease classification using efficient net deep learning model. Ecol. Inform. 61, 101182. DOI: 10.1016/j.ecoinf.2020. 101182.

[10]. Chen, J., fu Zhang, D., and Nanehkaran, Y. A. (2020). Identifying plant diseases using deep transfer learning and enhanced lightweight network. Multimedia Tools Appl. 79, 31497–31515. DOI: 10.1007/s11042-020-09669-w

[11]. Chouhan, S.S.; Singh, U.P.; Jain, S. Web Facilitated Anthracnose Disease Segmentation from the Leaf of Mango Tree Using Radial Basis Function (RBF) Neural Network. Wirel. Pers. Commun. 2020, 113, 1279–1296.

[12]. Mia, M.; Roy, S.; Das, S.K.; Rahman, M. Mango leaf disease recognition using neural network and support vector machine. Iran. J. Comput. Sci. 2020, 3, 185–193.

[13]. Nagaraju, Y.; Sahana, T.S.; Swetha, S.; Hegde, S.U. Transfer Learning based Convolutional Neural Network Model for Classification of Mango Leaves Infected by Anthracnose. In Proceedings of the IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 6–8 November 2020; pp. 1–7.

[14]. Pham, T.N.; Van Tran, L.; Dao, S.V.T. Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection. IEEE Access 2020, 8, 189960–189973.

[15]. Kumar, P.; Ashtekar, S.; Jayakrishna, S.S.; Bharath, K.P.; Vanathi, P.T.; Kumar, M.R. Classification of Mango Leaves Infected by Fungal Disease Anthracnose Using Deep Learning.

In Proceedings of the International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021.

[16]. Mrs. R. Dhivya, Dr. N. Shanmugapriya. (2021). A Comprehensive Review on the Identification and Classification of Leaf Diseases Model by their Various Stages. *Design Engineering*, Vol 2021: Issue 08, 10428-10444. Retrieved from https://www.thedesignengineering.com/index.php/DE/article/view/6102.

[17]. R. Dhivya and N. Shanmugapriya, "An Analysis Study of Various Image Preprocessing Filtering Techniques based on PSNR for Leaf Images," 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), 2022, pp. 1-8, DOI: 10.1109/ICACTA54488.2022.9753444.

[18]. S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, "Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for Identification and Classification of Plant Leaf Diseases: An Automatic Approach Towards Plant Pathology," in IEEE Access, vol. 6, pp. 8852-8863, 2018, DOI: 10.1109/ACCESS.2018.2800685.

[19]. R. Dhivya and N. Shanmugapriya. (2022). An Automated Plant Leaf Diseases Classification using AKMC and AKNN Machine Learning Techniques. JOURNAL OF ALGEBRAIC STATISTICS, 13(3), 3129–3142.