JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

An Automated Plant Leaf Diseases Classification using AKMC and AKNN Machine Learning Techniques

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 Science (PG),
Dr.SNS Rajalakshmi College of Arts & Science, Coimbatore-49.
spriyanatrajan@gmail.com

ABSTRACT
Agriculture is the largest determinant of a country's economic development. Approximately 70 % of the Indian populace is reliant on farming. Caused by a multitude of climatic circumstances, plants become susceptible to severe disease outbreaks. Those diseases begin with the plant leaves and then spread to the rest of the plant, affecting both the quantity and quality of the crops that may be grown. A person's eye can't distinguish and diagnose the disease on each crop in the region because of the enormous quantity of plants. As a result, it's critical to properly identify the individual plant to prevent the illness from spreading. As a result of this research, we have developed a system for detecting and diagnosing plant leaf diseases using "Machine Learning (ML)" techniques, and then suggest the type of disease to cure on time. Our strategy is geared toward raising agricultural yields by focusing on agricultural production. In general, most methods employ a set of fundamental procedures, including such as data acquisition, preprocessing, segmentation, extraction of features, selection of optimal features, and classifications. In this research, we have proposed an "Advanced version of K-Means Clustering (AKMC)" approach for segmenting leaves image and an advanced version of the "K-Nearest Neighborhood (AKNN)" approach for classifying leaf diseases. The AKNN classification model has been used to recognize "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves" in the leaf image. Using a nonparametric procedure, this classifier can make classifications. Data for this research is obtained from the "Manu's Disease data set" from openly available resources. The programming tool "MATLAB R2017" with a system configuration "Intel Core i7 CPU with a 64-bit Operating System" was used to evaluate the simulations. The experimental parameters such as "Accuracy", "Recall (Sensitivity)", "Precision", and "F-measure" are compared and tabulated for the proposed "AKNN" and the existing "Bacterial Foraging Optimization based Radial Basis Function Neural Network (BRBFNN)". The performance of the proposed method surpasses the existing method according to the results.
Keywords: Leaf disease, K-Means Clustering, K-Nearest Neighbor, Neural Network

1. Introduction

hroughout the context of computational intelligence, the agriculture industry has emerged as an important and novel context [1]. The basic objective of agriculture is to generate a broad variety of useful and significant plants and crops. Plant diseases may reduce the quality and quantity of a crop, thus they should always be dealt with earlier on. In current tradition, agricultural experts have concentrated on illnesses of crops and plants [2]. Numerous approaches for identifying and categorizing pathogens in plants and vegetables were developed by specialists [3]. As a consequence, it is essential to prevent infected crops from causing chaos in the quantity and quality of products.

Physical examinations are commonly employed to diagnose illness, but there are several drawbacks to this method, along with the timing, cost, and accessibility constraints, combined with the ability for several attempts. The leaves and fruits surfaces are the primary sites where many fungal and bacterial infections appear. It is hard to comprehend infections because of their complex forms. The longevity of bacteria, on the other side, is shorter and quicker. They divide a solitary cell into two in sequence to reproduce. Fibers and Genes are found in viruses, however, there are no specialized proteins to be found in these minuscule particles.

Although in recent decades many researchers had created an autonomous method based on "Image Processing (IP)" capable of rapidly identifying and recognizing illnesses in horticultural [4]. Researchers have employed IP to locate, color, shape, size, and define the borders of the affected region. Modern IP approaches are used for pre-processing and disease segmentation. Many IP approaches were used in 1970 to address the problems of inspecting crops with the naked human eye. Numerous prevailing techniques in the farming sector take into account the segmentation of separate crop elements such as fruits, stems, and leaves, extraction, and identification of various illnesses and stressed regions generated thereby. In the absence of a large amount of training data, more conventional IP techniques work effectively. The merging of background and antecedent areas, similarities, and forms produced during extracting features are among the difficulties associated with "Leaf Disease Detection (LDD)". Deep techniques have so far been successful in addressing a large number of these problems [5].

Segmentation process, extraction of features, and pattern classification are some of the most often used methodologies for detecting and evaluating leaf disease [6]. Using conventional methods, it is hard to retrieve features from a huge dataset. For big datasets, a "Deep Learning (DL)" approach extracts the most usable and important data processing from deep layers [7]. A further downside of adopting deep architecture is the presence of duplicate feature data. The "ReLu", "Max-Pooling", "Convolutional", "Softmax", and "Fully Connected" are some of the layers of the "Convolutional Neural Network (CNN)" architecture used for classification in DL [8]. For training a CNN architecture, you'll need to have a lot of data that hasn't been segmented. A CNN architecture often takes the raw information as input. The "VGGNet", "ResNet", "AlexNet", "Inception V3", and "EfficientNet" are some of CNN's highest prominent and also most sophisticated designs.

In required to practice the models, researchers use the "PlantVillage" dataset, which consists of "54,000" images of leaves divided into "36 groups" [9]. There will never be enough information for training a huge number of models. As a result, the "Transfer Learning (TL)" approach was rarely used in their research. Based on that database, they re-trained the model using this procedure with almost the original variables. The TL approaches outperform the CNN architectures in terms of the number of features that can be retrieved. Several applications benefit greatly from the TL method with DL methods, notably signals and systems, facial detection and recognition, the identification of road cracks, and the assessment of medical data. In particular, similar methods have had encouraging results in agriculture [10].

This problem statement was handled in this research on how to handle large leaf datasets. Many datasets are required to build an effective DL framework. Data sources, on the other side, are costly and complex to annotate. The preceding DL approaches, on the other side, are still only intended to classify specific plants or a breed of various plants for a specific illness. Plant leaves in the environment are frequently affected by many diseases. Furthermore, as compared to standard ML techniques, none of the aforementioned approaches have suitable combinations of DL techniques.

Developing an "Automated LDD" model to effectively detect the disease in the leaf is the major contribution of this research. There exists a necessity for consistent, precise, quick, powerful, and economical enhanced LDD approaches to identify the pathogens within leaf at a preliminary phase for agricultural advantages, considering the existence of several conventional technologies. To meet the needs of the nation's population boom, timely detection of plant pathogens is extremely important for enhancing agricultural output. Owing to their low cost and non-destructive nature, we applied classic IP and ML methodologies with gentle modifications. While the LDD procedure is relatively similar in all computing systems, there may be a few differences. The leaves images mostly from the database are subjected to some processing processes, involving pretreatment, segmentation, extraction of features, and feature selections before being classified. We had proposed an advanced version of the traditional "K-Means Clustering (AKMC)" approach in the segmentation module for segmenting leaves image and an advanced version of the "K-Nearest Neighborhood (AKNN)" approach in the classification module for classifying the type of leaf diseases. This proposed AKNN classification model has been used to recognize "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves" in the leaf image.

The paper was categorized as follows: Section 2 lists some recent publications related to LDD with advanced computing techniques, Section 3 briefs about the existing and proposed methodologies module by module, Section 4 gives a research finding of the classification module with a comparative of both existing BRBFNN and proposed AKNN classifiers, and Section 5 concludes the article with future scope.

2. Related Works

The "United-CNN" for disease identification among grape plants is discussed in [11]. This model's ability to accurately identify grape leaf infections is strengthened by its capability to fuse high-level features, enabling it to gain a sustainable competitive advantage. Classification and disease detection in leaf tissue by employing the "F-CNN" and "

S-CNN" models uses complete and fragmented images. As a result, their trained CNN network becomes higher accurate while deployed to segmentation results rather than entire images.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

For testing and training, the researchers of [12] employed certain DL approaches for pre-processing the frequency of imbalanced images. They employed "Gamma Correction" and "Decorrelation" to extend the color. The finding indicated the greater precision of having a balanced set of data. This approach uses "Color Distribution" to convert the images into ones that may be used to identify illnesses.

The researchers of [13] make use of the "Image Histogram-Transformation" approach. This approach discovered a multitude of diseases that induce crop damage, resulting in a crop shortfall. Datasets from several classifications are categorized with "Plant-Village" using a range of methodologies, with an overall accuracy of 91%.

Using an optimization technique "Whale Optimization," the researchers in [14] provide a "Hybrid-PCA" approach for extracting features and assessing the information in respect of precision and effectiveness. The two most common diseases that influence the quality and production of coffee plants are "Rust" and "Cercospora". They were able to obtain a 95% accuracy rate.

Researchers in [15] used "k-means" and a "Thresholding Segmentation" method to extract textual data from the olive plant and then used texture assessment to find the relationship between both the infected and healthy parts. They were able to reach a 93% accuracy rate.

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]. The general framework of the proposed LDD methodology is shown in Figure 1.

The flow of the proposed LDD methodology is described below:

- The infected and healthy plant leaves image is taken as input for this research.
- The foremost step is image preprocessing. The "K-means Singular-Value-Decomposition-Discrete-Wavelet Transform (K-SVD_DWT)" method is considered to preprocess the image for denoising.
- The second step is "Image Segmentation". This is performed by the "Advanced version of K-Means Clustering (AKMC)" for segmenting the diseased part from the preprocessed leaf image based on clustering. The clusters are grouped as infected and non-infected leaves. The "Region of Interest (ROI)" is selected for the infected region from the image segmentation phase.
- Then feature extraction is executed by "Principal Component Analysis (PCA)".
- The next step is feature selection and it is performed by "Particle Swarm Optimization (PSO)".
- The classification rule is applied by BRBFNN and AKNN Classifier algorithm for leaf disease classification in the proposed LDD model.
- The output is obtained and compared with the parameters "Accuracy", "Precision", "Recall", and "F-measure".

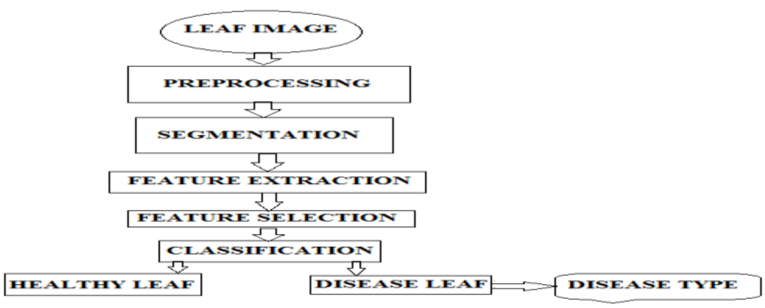

Figure 1: General structure of the proposed LDD Model

3.1 Preprocessing Leaf Image

Pre-processing eliminates unnecessary distortion and noise while enhancing the number of features required for subsequent computation. The "K-SVD DWT" is used as preprocessing technique in this research to make a more

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

comprehensive dictionary scheme by improving image denoising, and also we use this method to achieve faster image denoising than standard K-SVD. Here it is important to see how to apply the grayscale method to color pictures. To solve this challenge, implement the K-SVD method to each of the three channels R, G, and B separately. Color objects result from 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 better-preprocessed image will be produced for the segmentation process.

3.2 Segmenting Leaf Disease Region

Segmentation of image strategies 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 an image produces a series of regions that together comprise the whole image, amongst 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, although in this case would be the leaf, along with its background. This section covers the segmentation strategy for the implementation of an LDD model utilizing the AKMC method.

3.2.1 Traditional "K-Means Clustering" (TKMC)

The "K-means clustering" approach was a method of unsupervised classification that has been required if unlabelled information is provided, that is, information that might not be divided into categories, classes, or groups. The main aim of this approach is to find classes or groups of arbitrarily distributed data particles; the range of classes is determined by the vector "K" throughout clustering. This strategy is applied by several iterations to assign each set of information to each of the predetermined "K" groups depending on the applicable features. The readily accessible data pixels were clustered towards respective groups based on feature similarity. Furthermore, using the K-mean clustering method, the corresponding effects are obtained:

- The centroids including their various "K clusters", are primarily used for labeling received data points.
- Labels for training results, including each data point that is associated with a particular and distinct cluster.

Rather than just creating multiple classes or groups with little or no details, clustering allows it to explore the groups and also construct spontaneously. The "Choice of K" is crucial in determining the total of classes that will be generated. Every single centroid of a respective cluster has been made up of a set of feature values that will describe their respective groups. It will discover more about the kind of class also every cluster reflects for the datasets by looking at the centroid function weights. As previously said, this method employs repeated refinement to achieve the desired result. The set of data and the number of clusters are all necessary inputs for this process (K). In this case, a dataset is regarded as a set of extracted features for data in a repository. The earliest approximations again for specified centroids are the foundation of this technique. This could be generated from its input databases or selected randomly. Additionally, this technique can replicate itself through 2 stages: data allocation and centroid updates.

The biggest ***disadvantage*** of TKMC technique outcomes was that they were highly dependent on initialization. As the director of the centroid shifts, the efficiency alters as well. Alternatively, in the TKMC technique, randomized quantities of clusters are selected and centers are distributed at randomized locations resulting in weak inappropriate segments.

3.2.2 Advance version of "K-Means Clustering" (AKMC)

In comparison to TKMC, the proposed AKMC strategy produces stronger segmentation results. To address a weakness in the TKMC technique the procedure is updated in this proposed AKMC in terms of deciding the optimal quantity of clusters and respective means of obtaining appropriate segments. During the initial step, the preprocessed image is used to calculate the localized maximum and minimum values. Throughout this framework, an iterative approach is used to generate the optimum significance of original k- centroids by minimizing a specified objective function. Figure 2 depicts the work structure of the proposed AKMC technique.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

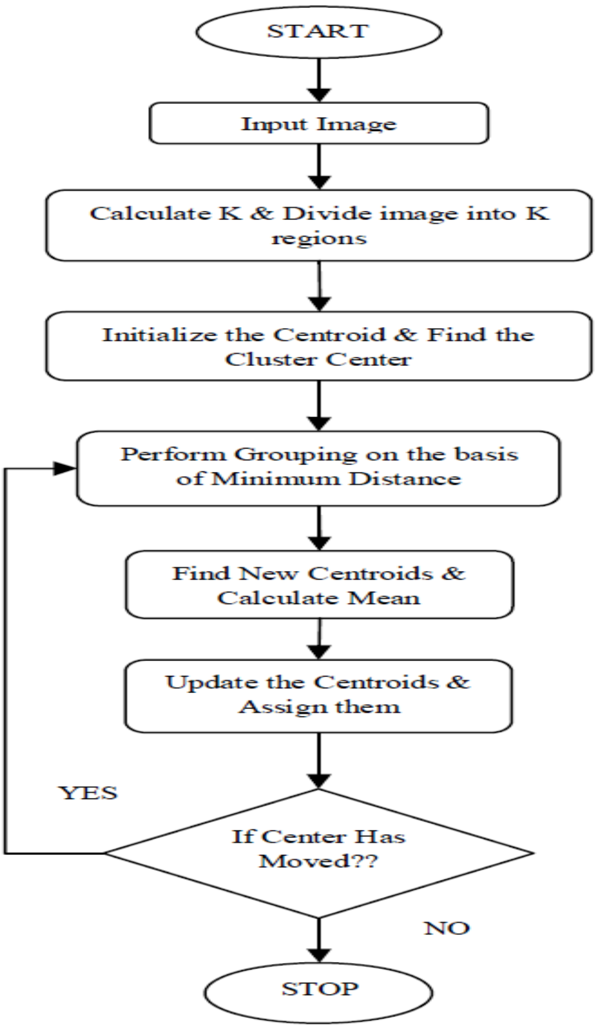

Figure 2: Proposed AKMC Segmentation Work Flow

The "Objective Function" in Equation (1) is used to measure the initial K Centroids.

$$d = \sum_{m=1}^{C} \sum_{n=1}^{N} \partial(r_{min}, r_{max}) \qquad \text{Eq} \rightarrow 1$$

The objective function for AKMC is defined in Equation (1), where "d" is the Euclidean-Distance among both localized minimal pixel "r_{min}" value and localized maximal pixel "r_{max}" value, "N" becomes the sum of pixels in the image, and "C" has been the cumulative quantity of clusters throughout the segmented image.

$$Y(C) = 0.5 * \sum_{i=1}^{N} \sum_{C(j)=i} \sum_{C(k)=i} \left\| r_j - r_k \right\|^2 \quad \text{Eq} \rightarrow 2$$

Or

$$Y(C) = \sum_{i=1}^{N} Num_i \sum_{C(j)=i} \left\| r_j - \mu_i \right\|^2 \qquad \text{Eq} \rightarrow 3$$

Where,

$$\mu_i = \frac{\sum_{k:C(k)=i} r_k}{Num_i} \;\; for\ i = 1,2, \dots, N \quad \text{Eq} \rightarrow 4$$

And

$$C(j) = \min\|\gamma_{j-\mu_i}\|^2 \text{ for } j=1,2,\ldots,Num \quad Eq \rightarrow 5$$

In Equations (2) to (6) the "μ_i" and "Num_i" denote the mean and pixel numbers in the i^{th} cluster, respectively. The above calculations are used to measure "μ_i".

The proposed AKMC system's processing time is reduced during the convergence process making it faster and more efficient. The subsequent measures are used to locate new centers:

Step-1: The mean significance for all of such fresh centers is determined after the fresh centers were estimated utilizing new allocated components.

Step-2: The actual distance value between both the present center and then the following center would then be calculated.

Step-3: If the actual value is larger than that of the mean value determined in Step-1, the fresh center value has been revised; otherwise, the present center and then the subsequent center's mean values are applied to the fresh center value.

Step-4: For each center, those 3 steps are replicated.

This segmented output from the AKMC method consists of only relevant disease parts, from this segmented image the features can be extracted for further process.

3.3 Feature Extraction from Segmented Leaf Image

The mechanisms of "Principal Component Analysis (PCA)" were utilized in this research to extract certain features from segmented leaf images. PCA is a "statistical technique" that converts a series of measurements of potentially connected factors into a system of principles of uncorrelated quantities termed principal components using an orthogonal transformation. The quantity of initial variables is smaller than that or equivalent to the set of principal components. This transition is established ensuring that the very first principal component seems to have the maximum variance possible, and then each subsequent component seems to have the maximum variance possible while remaining orthogonal to the subsequent components. Again, when mean converging the information to every attribute, PCA may be performed using "Eigen Value Decomposition" of both a "Data Covariance Matrix" or "Singular Value Decomposition" of a "data matrix". Through this essential features were extracted and further given to the feature selection process.

3.4 Feature Selection from PCA Extracted Features

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 being used. "Particle-Swarm-Optimization (PSO)" is gaining popularity in the field of selecting features. Throughout this research, the selection of features was being 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 PCA features and it passed to the classification process.

3.5 Leaf Disease Classification

Classification is a method for estimating outcomes based on feedback. In this, the training set consists of a set of properties that can be used to forecast the outcome. It decides the association between all the properties in able to forecast the performance. In this research, the classification results are the disease prediction in the leaf. The prediction set is made up of almost identical properties to the training set. The existence or unavailability of Leaf Disease, and also the category of disease, is the predictive factor. The BRBFNN and AKNN methods will be analyzed individually in the subsequent subsections for classifying the leaf into it following appropriate classes such as: "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves" as an output.

3.5.1 BRBFNN

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

The researchers in existing already have presented a technique called "Bacterial Foraging Optimization based Radial Basis Function Neural Network (BRBFNN)" for autonomously identifying and classifying plant leaves diseases. Also with the support of BFO, they've trained 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 leaves 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.

Disadvantages:

- In comparison to "Multilayer Perceptron," RBF system training is quicker, whereas classifying is slower. Attributed to the reality that throughout classification, each node inside the hidden layer must calculate the RBF polynomial again for the input feature vector.
- As a result, high-dimensional vector areas have a problem with high-performance computation due to the obvious amount of duplicative data incorporated by their high-dimensional vectors.

3.5.2 AKNN

a. Traditional KNN Classifier

KNN is one of the most prominent classification strategies and is more widely referred to as a non-parametric classification model. Hodges and Fix were introduced the KNN in 1951. The basic function of this classifier is an uncomplicated and efficient algorithm. Both classification and regression functions may be done through this. This technique takes the principle of generalization and adds a new variant to the class. It's a structured learning process, which classifies the outcome of the training set query into the nearest majority subclass. The simple explanation of this classifier is that a new entity is classified based on the attributes and training samples. The demerits of this classifier are:

(i) It was generated only with the support of the samples of training which doesn't use any additional details. This classification of the KNN results was mostly dependent on the training set if a re-calculation is required in the training set.

(ii) Training samples can be determined depending on the similarities to locate the samples of the closest neighbor K. In any case, the classifier for classification is no longer suitable because the training set includes only fewer samples.

b. Proposed AKNN Classifier

In comparison, the TKNN classifier measures the correlations by using extra time, if the training set includes more sample numbers. These problems can be overcome by proposed AKNN in upcoming forms:

- Using fewer collections of data
- Minimizes the feature size
- Using an optimization
- Every training data is continuously accounted whereas the weight of all samples is not needed to compare.

On arrival of the new sample, it uses the basis of similarity measure "Euclidean-Distance (ED)", for finding the neighbors closest to the new sample from the training segment. In this research, the ED was used for the proposed AKNN classifier model.

Euclidean Distance

This proposed AKNN method classifies the relevant information obtained from the nearest training samples. 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

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

class. The ED in this proposed AKNN classification calculates and expresses the distance from the testing data vector to all the training vectors based on euclidean which was given in the following Equation (6).

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}$$

Eq → 6

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.

Algorithm Steps:

The following steps are taken in the AKNN classification:

- Choose the number of neighbors that are nearest through parameter K.
- For finding the nearest K individuals from the training dataset that are identified with each case in the target dataset, which is the set to be expected. Euclidean-Distance is used to assess how near the training member is to the target-row were examined.
- Sorting the distance in terms of the minimal Kth distance and settling the closest neighbor.
- Assembling the neighbors that are nearer through classifying.
- Using the basic neighbor's class as the current query instance it predicts the attribute.
- The complete process will be repeated till the target-set form.

The method is quite basic, which tends to scan the closest neighbors to the test sample and transfer its class label on a majority of the neighbors' labels.

Classification of Leaf's Disease Using AKNN

The AKNN classifier used in this research to detect the classes as "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves" for the classification of the leaf image. This classifier does a classification based on a non-parametric process. If a new training feature subset is applied to an established training collection, there is no need for previous information regarding the layout of the feature subsets gained by PSO in training sets (No retraining is required). The proposed AKNN technique performance can be used as an eternity of the input pattern of a given class. While 'k' rises, predicting confidence will be increase.

The classifier AKNN here is composed of 2 phases:

(i) Training Phase:

In this research, the classes are named "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves". While training the subset of features are labeled and grouped in these classes.

(ii) Testing Phase:

In this research, the subset of features is not labeled initially in the testing phase. The AKNN performs the classification by processing the feature subsets and labels it by finding the K-Nearest datapoint then classifying it by comparing it with trained classes "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves".

For this, the AKNN follows as:

- Creating a list of gathered subsets with optimal features.
- Using ED for computing the distances among the saved subset of features (trained-set) with the subset of features that are unknown (tested-set) for classification.
- Identifying the K nearest by using the closest adjacent class-labels "Alternaria Alternata", "Anthracnose", "Bacterial Blight", "Cercospora Leaf Spot", and "Healthy Leaves" to identify the class-labels of the undefined record by voting has a higher majority.

Figure 3 shows the methodology of the proposed AKNN for classifying Leaf Disease.

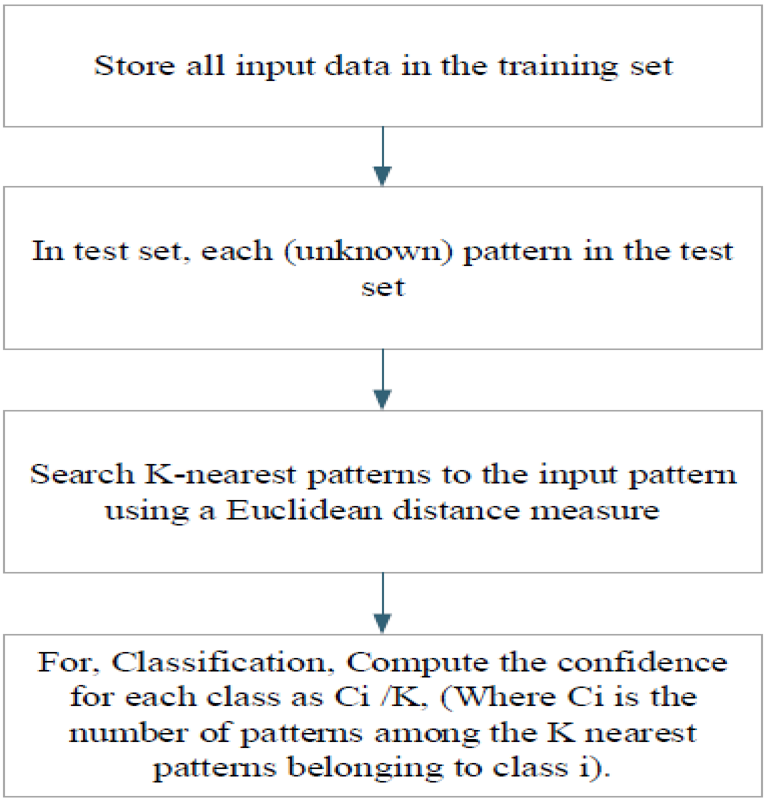

Figure 3: Proposed AKNN Methodology

By using the above decision rule, the AKNN classifier is employed to identify the type of leaf disease with less computational time.

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 experimental parameters such as "Accuracy", "Recall (Sensitivity)", "Precision", and "F-measure" are compared and tabulated for the proposed AKNN and the existing BRBFNN classifiers.

Dataset: The leaf images 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 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 being 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 (7).

$$"Accuracy = \frac{TP+TN}{TP+TN+FN+FP} * 100" \qquad \text{Eq} \rightarrow 7$$

Table 1: Accuracy Comparision

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

Table 1 and Figure 4 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 and proposed AKNN classifiers to evaluate their accuracy rate. The findings prove that the accuracy level of the proposed AKNN classifier is higher for all the classes while comparing it with the existing BRBFNN classifier.

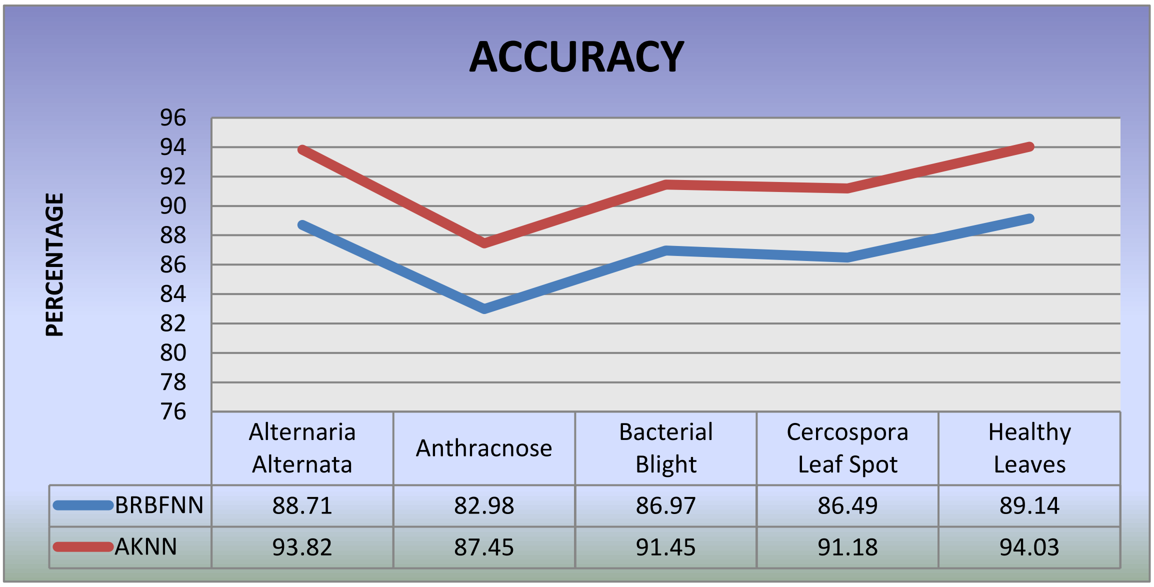

Figure 4: Accuracy Comparision

4.2 Recall/Sensitivity

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

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 (8) essentially calculates and specifies the percentage for Recall.

$$"Recall = \frac{TP}{TP+FN}" \qquad \text{Eq} \rightarrow 8$$

Table 2 and Figure 5 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 and proposed AKNN classifiers to evaluate their Recall rate. The findings prove that the Recall level of the proposed AKNN classifier is higher for all the classes when comparing it with the existing BRBFNN classifier.

Table 2: Recall Comparision

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

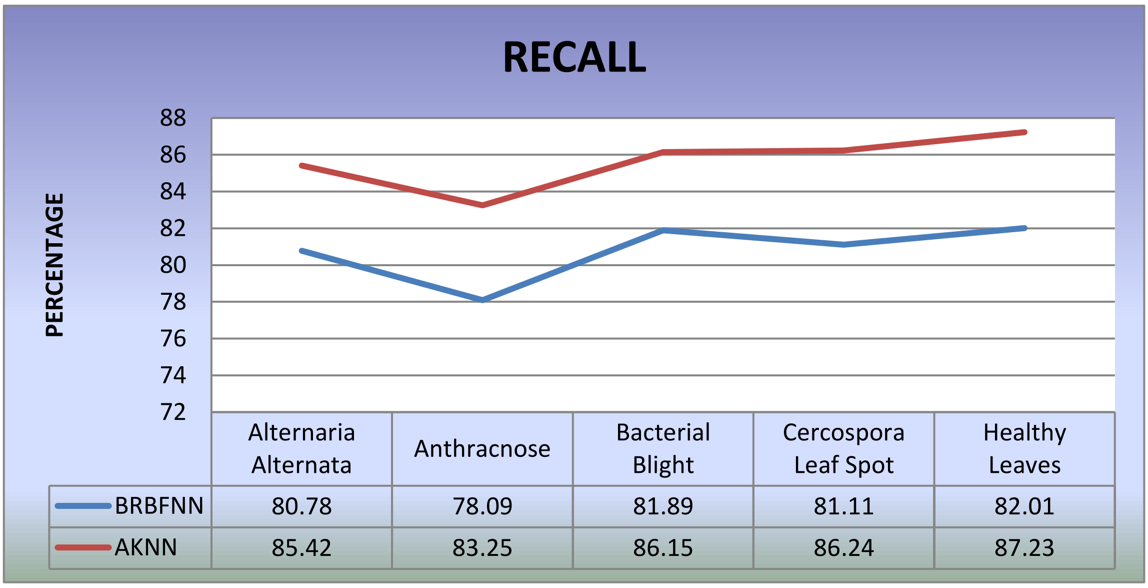

Figure 5: Recall Comparision

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 form 95% of the time. This also describes the data points in consideration. The Confusion-Matrix, as seen in Equation 9, is used to quantify the precision level.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

$$Precision = \frac{number\ of\ true\ positives}{no\ of\ true\ positives + false\ positives} = {TP}/{TP + FP} \quad \text{Eq} \rightarrow 9$$

Table 3: Precision Comparision

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

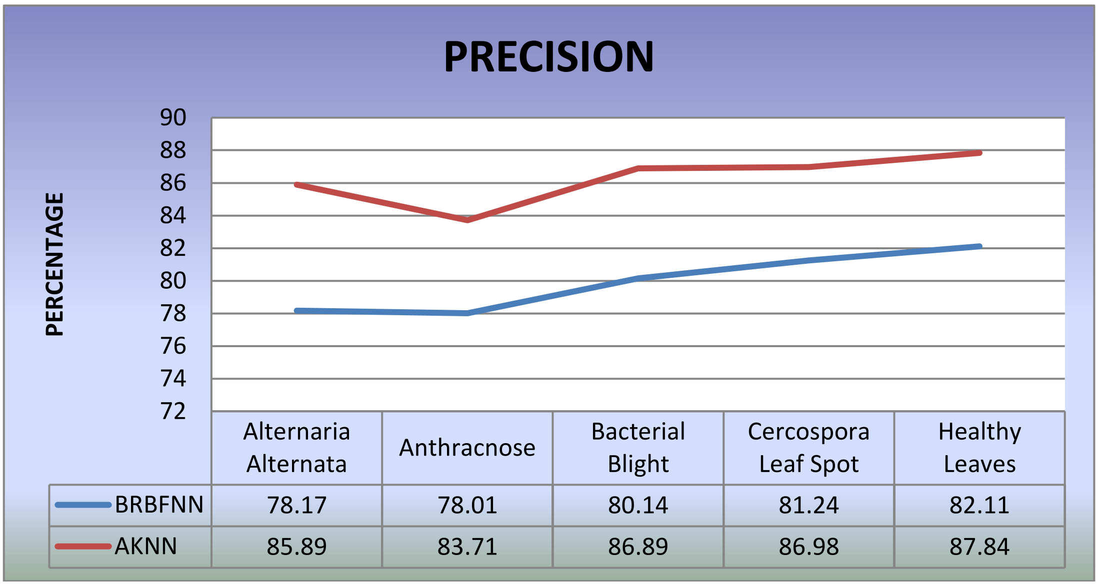

Figure 6: Precision Comparision

Table 3 and Figure 6 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 and proposed AKNN classifiers to evaluate their Precision rate. The findings prove that the Precision level of the proposed AKNN classifier is higher for all the classes when comparing it with the existing BRBFNN classifier.

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 10.

$$F_1 = 2.\frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2.\frac{precision.recall}{precision + recall} \quad \text{Eq} \rightarrow 10$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 3129-3142
https://publishoa.com
ISSN: 1309-3452

Table 4: F-Score Comparision

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

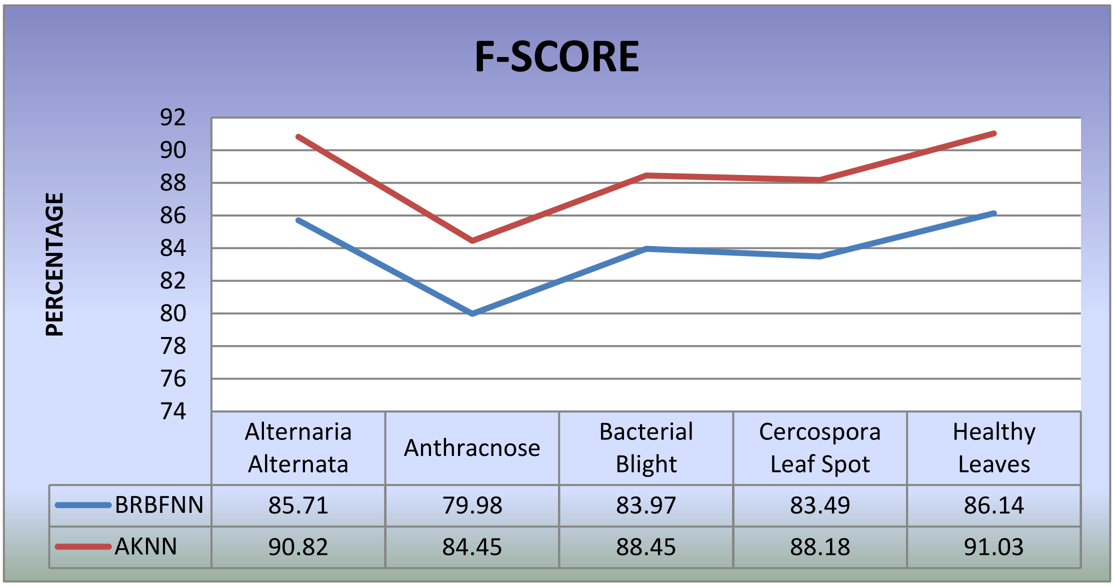

Figure 7: F-Score Comparision

Table 4 and Figure 7 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 and proposed AKNN classifiers to evaluate its F-Score rate. The findings prove that the F-Score level of the proposed AKNN classifier is higher for all the classes while comparing it with the existing BRBFNN classifier.

5. Conclusion

We could increase the production and grade of plants and their substances by limiting the biotic variables that cause significant crop damage. Recognition system, categorization, particle extractor, and other computer vision-based ML methods have all surpassed traditional methods in the combat against plant leaf disease. The categorization of plant leaves affected with fungal infections as "Alternaria Alternata", "Anthracnose", "Bacterial Blight", and "Cercospora Leaf Spot", and normal leaf as "Healthy Leaves" is made easier with the help of our new model, called AKNN classifier. The existing classifier RBFNN and the proposed classifier AKNN were implemented in our LDD model and the classification performance was calculated using the essential metrics such as "Accuracy", "Precision", "Recall", and "F-measure". The technique has also been effectively applied and consistency checked on a publicly available Manu's leaf disease collection of leaf image data. As compared to the existing BRBFNN with the proposed AKNN, the metrics levels of the AKNN are higher for all leaves in this overall research work. In the future, we plan to focus on feature extraction and a novel approach to classification to enhance the accuracy rate.

References

[1]. Kaur, P.; Gautam, V. Plant Biotic Disease Identification and Classification Based on Leaf Image: A Review. In Proceedings of 3rd International Conference on Computing Informatics and Networks, Delhi, India, 29–30 July 2020; Volume 167, pp. 597–610.

[2]. Maddikunta, P.K.R.; Hakak, S.; Alazab, M.; Bhattacharya, S.; Gadekallu, T.R.; Khan, W.Z.; Pham, Q.-V. Unmanned Aerial Vehicles in Smart Agriculture: Applications, Requirements, and Challenges. IEEE Sens. J. 2021, 21, 17608–17619.

[3]. Hang, J.; Zhang, D.; Chen, P.; Zhang, J.; Wang, B. Classification of Plant Leaf Diseases Based on Improved Convolutional Neural Network. Sensors 2019, 19, 4161.

[4]. Basha, S.M.; Rajput, D.S.; Janet, J.; Somula, R.S.; Ram, S. Principles and Practices of Making Agriculture Sustainable: Crop Yield prediction using Random Forest. Scalable Comput. Pract. Exp. 2020, 21, 591–599.

[5] . Mary, N.A.B.; Singh, A.R.; Athisayamani, S. Classification of Banana Leaf Diseases Using Enhanced Gabor Feature Descriptor. In Inventive Communication and Computational Technologies; Springer: Singapore, 2021; Volume 145, pp. 229–242.

[6]. Hossain, S.M.M.; Tanjil, M.M.M.; Ali, M.A.B.; Islam, M.Z.; Islam, M.S.; Mobassirin, S.; Sarker, I.H.; Islam, S.M.R. Rice Leaf Diseases Recognition Using Convolutional Neural Networks. In Advanced Data Mining and Applications. ADMA 2020; Springer: Cham, Switzerland, 2020; Volume 12447, pp. 299–314.

[7]. Atila, Ü.; Uçar, M.; Akyol, K.; Uçar, E. Plant leaf disease classification using EfficientNet deep learning model. Ecol. Inform. 2021, 61, 101182.

[8]. Nagaraju, M.; Chawla, P.; Upadhyay, S.; Tiwari, R. Convolution network model based leaf disease detection using augmentation techniques. Expert Syst. 2021, e12885.

[9]. Xiong, Y.; Liang, L.; Wang, L.; She, J.; Wu, M. Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset. Comput. Electron. Agric. 2020, 177, 105712.

[10]. Chen, J.; Chen, J.; Zhang, D.; Sun, Y.; Nanehkaran, Y. Using deep transfer learning for image-based plant disease identification. Comput. Electron. Agric. 2020, 173, 105393.

[11]. Sharma, P.; Berwal, Y.P.S.; Ghai, W. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. Inf. Process. Agric. 2019, 7, 566–574.

[12]. Oyewola, D.O.; Dada, E.G.; Misra, S.; Damaševiˇcius, R. Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing. PeerJ Comput. Sci. 2021, 7, e352.

[13]. Abayomi-Alli, O.O.; Damasevicius, R.; Misra, S.; Maskeliunas, R. Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. Expert Syst. 2021, 38, e12746.

[14]. Gadekallu, T.R.; Rajput, D.S.; Reddy, M.P.K.; Lakshmanna, K.; Bhattacharya, S.; Singh, S.; Jolfaei, A.; Alazab, M. A novel PCA–whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. J. Real-Time Image Process. 2020, 18, 1383–1396.

[15]. Sinha, A.; Shekhawat, R.S. Olive Spot Disease Detection and Classification using Analysis of Leaf Image Textures. Procedia Comput. Sci. 2020, 167, 2328–2336.

[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.