Image-Based Plant Disease Detects using Machine Learning

A.Vijay Vasanth[a], Saiyed Faiayaz Waris[b], Dr Gudipati Murali[c], T R Rajesh[d]

[a]Department of Computer Science and Engineering, SRM Institute of Science & Technology, Tamil Nadu, India
[b,d]Department of Computer Science and Engineering, VFSTR (Deemed to be University), Andhra Pradesh, India
[c]Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Andhra Pradesh, India

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract : Computational algorithms are used for the processing of numerical images. It includes a wide variety of algorithms that can prevent problems during processing, such as noise buildup and signal distortion, and that can be applied to input data. In the context of multivariate systems, digital image processing can be modeled and images are defined in two dimensions (perhaps more). To automatically identify and classify plant leaf diseases, we need to plan and develop a software solution using image processing. The image processing method has primarily four steps. The first step is the structure of the color transformation of the Red Green Blue (RGB) leaf images and it will be translated into a grayscale image after enrichment using Otsu. Using the K-means clustering technique, the images are at hand in the second step. At step 3, Gray Level Co-Occurrence Matrix computes and extracts segmented contaminated objects. To categorize the disease according to the characteristics computed and obtained from the target image, we use the Multi-Class Support Vector Machine (MCSVM). The proposed approach was tested using a baseline data set and showed greater accuracy than the use of Artificial Neural Networks (ANN). In conclusion, the proposed classification model, the MCSVM, is highly effective in identifying leaf diseases and can work well for different types of diseases.

Keywords: Image processing, Multi-class Support Vector Machine, RGB, Artificial Neural Network.

1.Introduction

In today's world, the agricultural landmass is more than just a food supply and productivity is based on the Indian economy. It involves the identification of plant diseases that play a significant role. The early stage is the use of plant diseases which is diagnostic. Automatic disease detection has benefits with this technique. In the current approach, it needs to perform the detection of plant diseases, which is essentially the unaided eye observation performed by researchers. In addition to the continued monitoring of plants, which is very expensive when it comes to large farms, a large team of experts is required. In some countries, farmers do not have enough facilities and do not have any idea of communicating with experts [1]. The frequency of consultations with experts is also significant and time-consuming. Under these conditions, the proposed method is useful for the surveillance of field crops. On plant leaves, symptoms can be seen, making automatic detection easier and cheaper. To provide automated process control, inspection, and image-based robot guidance that makes computer vision easier [2].

The process of separating or clustering an image into separate sections is referred to as image segmentation. From the base threshold approach to color image segmentation, there are many ways to perform image segmentation. Because of this, humans can readily identify and manipulate individual objects. When segmenting images, computers have no way of recognizing objects intelligently because many methods have been developed. Depending on the image segmentation mechanism, multiple features have been found in the image such as color, borders, or parts [3].

2 .Problem Definition

Plant disease visual detection is more tedious and less reliable, and it can be performed in restricted areas which can put less effort with less time and become more accurate if used as an automatic detection method. Early brown and yellow spots are general, fungal, infectious, and bacterial diseases of plants. Image analysis makes it possible to determine and assess the area affected by the disease and the color difference.

In the first stage, the color transformation structure of the RGB leaf image is built. The second phase uses the median filter to achieve a clean image and covers device-independent color space conversion and evaluates each pixel with its surrounding pixels based on the color intensity difference. Then, to reach its goal of determining the parts affected by the disease, it uses density-dependent bundling. The third stage is when the texture characteristics are segmented and infected objects are measured. The extracted characteristics are passed through a pre-formed neuronal network in the fourth stage.

Plenty of processing power and a lot of time coming online was done with the help of techniques in the current method. Therefore, propose a system to make use of the Otsu classifier, which takes the AML values to arrive at a more pronounced and detailed conclusion at a much lower calculation cost. Using Matlab, A MCSVM is used to extract the characteristics of the segmented image for ultimate disease detection and recognition.

3. Literature Survey

Five main steps are used in the identification of plant leaf diseases using image processing. It needs image enhancement and segmentation is affected by image preprocessing. The useable area is segmented in the processing frame for digital camera or web image acquisition, retrieval, and classification of features. The presence of plant leaf diseases would be recognized at last. As shown in the following step-by-step process [4]:

1) Acquiring an RGB image.
2) Convert the input representation in color space.
3) Fragmentation of components.
4) To obtain relevant segments.
5) Computing texture characteristics.
6) Classification

The image needs to be pre-processed to enhance the data that suppress undesired distortions, to increase certain image characteristics that are needed for further processing tasks and analysis. It involves color space conversion, image enhancement, and image segmentation. By converting the RGB images from the leaves the color space representation is obtained. The objective is to make it easier to specify the color in a standard and accepted way [5]. It represents the Hue Saturation Value (HSV) is the transformation of RGB images. RGB is for color generation and its descriptor. An excellent tool for color perception is the HSV model. The tint is an attribute of color that, interpreted by an observer, describes the pure color. The quantity of white light applied to the color, with the intensity of light indicated by the value is called saturation. The stained part was used for further analysis after the color space transformation process. If no further information is available, the saturation and value are reduced [6].

The region to be segmented that is found should be closed. Given the absence of edge pixels in this region-based segmentation, there would be no difference. For segmentation, the boundaries are specified. Each phase has a pixel that is considered and connected to the region. The edge flow is transformed into a vector after recognizing the change in color and texture. Then those borders for further segmentation are detected [7]. The region of interest was removed, i.e. the diseased portion after segmentation. In the next step, significant features are extracted that can be used to determine the value of a given sample. The characteristics of the image usually include color, shape, and texture. Most scientists consider leaf texture to be the most important feature of plant classification. Plant diseases are divided into different groups using the texture characteristics using the nearest K-classifier [8].

With a range of 0 to 255, it is difficult to implement applications using RGB and it converts RGB into gray frames. To improve the images of plant diseases, the equalization of the histogram attributes the force and then applied to the image [9]. As the image of the infected leaves captured is an RGB image, it can switch from RGB to Grayscale for pre-processing the Grayscale image. It contrasts the brightness of the grayscale with the luminance of the color picture. Originally we get the values of the three primary colors (Red, Green, and Blue) and using the Gamma expression encode these linear force values.

Gamma's expansion is as follows:

$$\boxed{C_{linear} = \begin{cases} \dfrac{C_{rgb}}{12.92} & C_{rgb} <= 0.04045 \\ \dfrac{(C_{rgb} + 0.065)}{1.065} & C_{rgb} > 0.04045 \end{cases}} \quad \text{-------- (1)}$$

Crgb is an RGB primary that ranges from 0 to 1 and Clinear is also a linear force value that ranges from 0 to 1. Using the weighted sum of the three linear intensity values, the luminance of the output frame is obtained, and using the function: get the conversion y=f(x).

Here, x = initial input data.
y = translated output data.

Using the weighted sum of elements R, G, and B, then f(x) function converts RGB values into grayscale.

$$f(x) = 0.2989 * R + 0.5870 * G + 0.1140 * B \text{ ------------------------- } (2)$$

It is necessary to segment the leaf while the feature is retrieved from this image. Otsu threshold and k-means clustering algorithm are used and compared in testing infected leaves. For k-means, the values extracted from the characteristics are inferred. The collection of similar values in an array or similar color maps is called Clustering [10]. The mapped regions are grouped using color mapping and form a group called grouped tagging. Here, the pixels of a color image will be grouped into different clusters and form final segmented image regions using the density-based clustering method. Compare with k-means that the other techniques are more precise to achieve the clarity of clustering. For disease classification, the RGB image is used. Green pixels are detected after implementing k-mean grouping techniques, and then variable threshold values are obtained using the otsu process.

Like Spatial, Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern, and Run Difference Method of Feature Extraction, these are different methods for extracting functions. This matrix is generated and the function calculated for texture statistics is calculated by using the GLCM function. It is a texture analysis method that considers the spatial relation of pixels in the GLCM, also called the gray-level spatial dependence matrix [11]. It computes the pixel pairs and the unique values that existed in a given image and spatial relationship. It is used to determine the texture of an image, generate a GLCM, and pull out statistics.

Decision Trees (DT), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and other classification methods that have been studied are contrasted. The DT may be used as a filter and to handle a large amount of information. It is a fundamental way of identifying datasets that can produce effective and accurate outcomes [12]. This algorithm is remarkable, which can be very easily formed and provides solid results for classification data for adjustment accuracy. Big Data now extends to all fields with rapid growth in networking and increased data storage and collection capacity.

For example, the classification of an email received is based on the examination of characteristics like originating IP address, the number of emails received from the same sender, subject line, email address itself, email body content, etc. lead to "spam" or "not spam". The algorithm is designed to take into consideration the email and characteristics that contribute to the final value [13]. The Machine Learning structure provides more examples of spam and non-spam e-mails, which makes sense, the sooner the next unknown e-mail will be predicted. The quick and precise way is to learn the decision tree. This method makes it easy to train and parallelize large datasets. It develops on individual n processors independently. By using meta-learning, learners progressively regulate them and become aware of the models of perception, interrogation, learning, and development.

The result is a tree of reduced size and another characteristic of the final growth of DT, the knots have no accuracy. For large drive sets that contribute to growing large trees, pruning is probably very important. There are different methods to prune a DT. In C4.5 it is very fast and it has been demonstrated that it generates an approach known as pessimistic size [14].

The SVM is a machine learning framework that classifies points into one of the two half-spaces that are disjunct. To convert the original data into greater dimensions, it uses non-linear mapping. The objective is to build a function that accurately estimates the class to which the new article and the old points belong [15]. In the age of Big Data, the best description behind the overall margin or gap is if we use a decision limit to classify a comparison to another, it will end up closer to one set of data sets. It happens if the data is ordered or linear, but a great deal of the knowledge is unstructured/non-linear and the dataset is inseparable and SVM kernels are used [16].

Traditional methods of classification operate weakly when they operate directly due to the huge amount of data. If SVM is used then it can avoid information representation issues. In comparison with other classification approaches, the most promising approach and the solution is the SVM. A large amount of data and a negotiation between the complexity and error of the classifier would directly balance the SVM properly and precisely. Another advantage of SVM can be generated and used for a specific problem that could be directly applied to the data without the need for a feature extraction process. There are major issues in which the functionality extraction process loses a large amount of structured data.

The SVM shall process the general instruction information in the classification technique. It should stay with large and complex data so that when there is too much noise in the data sets, the result of the SVM will be strongly affected. As it is an efficient classification model, it provides an optimized algorithm to solve the overload problem which is helpful for processing complex data. To show the highest eigenvalues and its vectors of formation data do overlap (nucleus) and covariance matrices in the quantum form SVM can use nuclei. When the training set is noisy and unbalanced, it has high reliability of the training and small mistakes of generalization, illustrating its possible problems. It takes time to scan multiple datasets and is not that flexible. So it's too expensive to build large datasets [17].

ANN includes animal brains that lead to neural networks. To process complex data inputs for many machine learning algorithms, this is not an algorithm. Normally, by considering examples without scheduling task-specific rules, these systems "learn" to carry out tasks. When considering the implementation of ANN, the signal at a connection between artificial neurons is a real number, and taking the sum of the inputs, the output is calculated by the non-linear function. Links between artificial neurons are 'Edges'. Generally, as learning progresses, the edges have a weight that may vary. A signal may only be sent if the aggregate value is above the threshold. Typically, by taking inputs these neurons can aggregate into layers and can perform various types of transformations. After passing through the layers more than once, these signals move from the first layer (input layer) to the last layer (output layer).

The main task of this method is to shape and identify targets for large amounts of inputs. Most importantly, the Hughes phenomenon or "the curse of dimensionality" is the most complex issue and may affect NN modeling. Also, ANN is known to be a black box since it contains no prior knowledge and itself is complicated. Look at the black box algorithm as a linear classifier, using all the labels each class can be separated. Let 'D' be a set of training examples. Collect elements with label D as positive examples and consider that they all remain negative for any label and then build a binary classification problem for I.

By applying possible labels as 'K' then produce k binary classifiers w1,w2,..., wk. If the k classifiers have been learned, the Winner Takes All (WTA) strategies to determine such as f(x) = arg max iw Tix. The "score" wTix(it) is possible to apply the sigmoid function on (it) can be considered as the likelihood that x has the I label. By considering graphically, three-color labels are shown as a dataset below figure. We have separated the black, blue, and green from the rest by applying binary classifiers.

When some points are not linearly separable from others with a certain label and pattern cannot be used which represents carat which is shown in Figure 1. In most cases, the expressiveness of this paradigm is affected. Learning is not always possible and cannot be dissociated. There are no theoretical reasons even if it is common and works well.

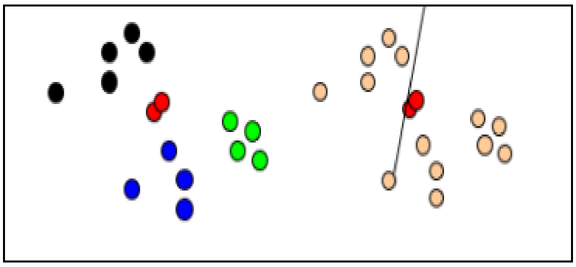


Figure 1: OvA SVM

The estimate by using a binary classifier in hypothesis space for every pair of class there is a separation. Define a binary learning problem for each pair by looking at all pairs of labels ((k2)of them). The positive are all examples with label i, and the negative are those with label j for pair (i,j). By this scenario consider k classifiers as in OvA, we have (k2) classifiers and each label gets k1 votes, and the decision is more involved, output of binary classifiers may not cohere as the result. If we take label i, if i wins on x more often than any j= 1,...,k as an option to classify the example, Figure 2 is used to making a decision. We can do a tournament alternatively. Continue with the winners and go down iteratively starting with n/2 pairs.

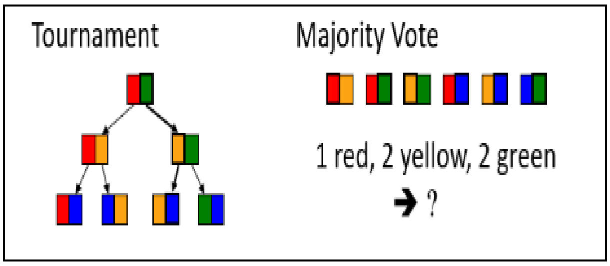


Figure 2: AvA SVM

Various types of SVM were compared and found that an MCSVM would be the perfect fit for our system to reach the desired speed and accuracy. By determining all decision functions at the same time, the question of unclassifiable regions for multi-class datasets was resolved. This SVM formulation in which the multiclass problem is solved with a single optimization problem is called an all-in-one SVM classifier. That's called an All-At-Once approach. It constructs k binary decision functions where/h g i (X) decision function separates the drive data points of class Ci from the other data points of the class.

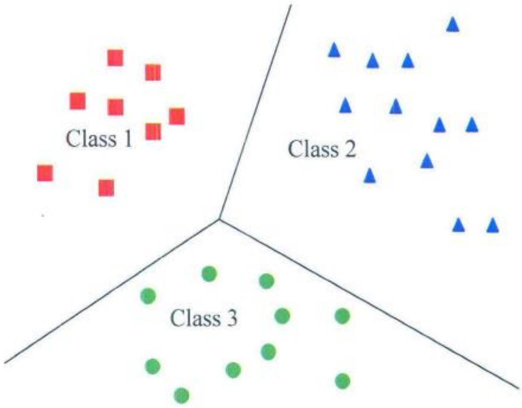


Figure 3: Multi-class SVM

Figure 3 shows the decision boundaries for a three-class classification problem using the all-at-once SVM approach.

4. Proposed System

The proposed algorithm, a diagrammatic representation of the proposed system, and the method used to carry out our system design. The identification of plant leaf diseases comes with 5 main steps based on image processing. A digital camera or internet image acquisition, image pre-processing requires image enhancement, segmentation involves extraction and classification of features for affected and functional regions as shown in Figure 4.

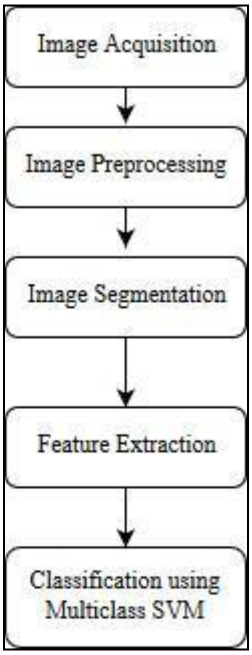


Figure 4: Leaf image processing

4.1 Acquisition of Image: Collecting the data from the public repository is the initial step and takes the image as the input for additional processing. To use any formats such as .bmp, .jpg, .gif to input our method most common image domains were taken.

4.2 Image Preprocessing: Images obtained from the real field are possible contain dust, spores, and water spots as noise. Data preprocessing purpose is to minimize the noise and to adjust the pixel values in the image as per order. This reinforces the image quality.

4.3 Image Segmentation: The third stage in our proposed methodology is image segmentation. Using the Otsu classifier, segmented images are grouped into various fields. The RGB color model is transformed into a Lab color model before the images are clustered. The Lab color model's creation is to group the segmented images easily together.

4.4 Feature Extraction: In the processing of images, the extraction of characteristics begins with the initial collection of measured data and generates derived values (characteristics) intended to be descriptive and non-redundant, enabling subsequent steps of learning and generalization. In some cases, it enhances human interpretation. Feature extraction is a technique of dimensionality reduction, where an initial set of raw variables is reduced for processing two or more manageable groups (features) and represents the original data set accurately and fully.

If an algorithm's input data is too large to be processed and is suspected of being redundant (e.g. the same measurement at both feet and meters, or the repeatability of images displayed as pixels), it can be transformed into a reduced feature array (also named a feature vector). The collection of a subset of the original characteristics is called the function's selection. The chosen features are intended to contain the required input data information so that the desired task can be accomplished using this reduced representation instead of the complete initial data.

Extraction of shape and textural features utilization is Gray Level Co-Occurrence Matrix(GCLM). The existing image shows the brightness of the pixel. It tests, calculates, and marks characteristics such as contrast, correlation, energy, homogeneity, skewness, kurtosis, mean, standard deviation, entropy, etc., which is then stored in a table for SVM training.

4.5 Classification: By using a Multi-Class SVM this technique is done. Inherently, they are two-class classifiers. The traditional way to do multiclass classification with SVMs:

- A set of class documents (positive labels) and an additional set of documents (negative labels), and a classifier are constructed for each class of training set.
- Apply each classifier separately as per the test document. There is no effect on the decisions of the other classifiers by the decision of one classifier.

In practice, the most popular technique was to create one-against-all classifiers (commonly referred to as one-against-all classification or OVA classification) and to determine the class with the greatest margin that classifies the test date. Time for training classifiers will decrease, as the data set for each classifier is much smaller, it requires building classifiers.

5. Implementation of the Proposed System

5.1 Design steps

As explained in the Design content, the design steps consist of the basic steps of Image Processing combined with the classification technique, in our case, a Multi-Class SVM, to recognize the disease and to calculate the percentage of the affected area.

The color space of CIELAB (also known as CIE L*a*b* or sometimes abbreviated as 'Lab' color space) was established by the International Commission for Illumination in 1976. For color, three numerical values were expressed as L* for lightness, and a* and b* are the components of green-red and blue-yellow. Concerning human color vision CIELAB was designed to be perceptual. At the same time, numerical change in these values corresponds to the same amount of visually perceived change.

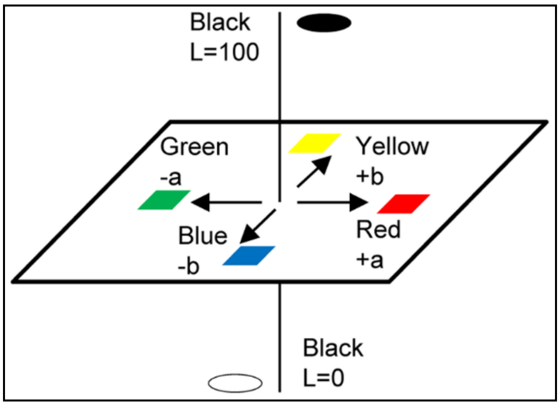


Figure 5: Lab space color model

L*a*b* consists of an 'L*' radiance layer,' a*' chromaticity layer indicating where the color falls along the red-green axis, and' b*' chromaticity layer shows where the color falls along the blue-yellow axis in the space. The above figure shows the color model for the lab room. All the color details as shown in the layers of 'a*' and 'b*'.

In the technique of image processing nextly we use Otsu's classifier. It is a method used to perform Threshold image-based clustering. Nobuyuki Otsu conducts the diminishment of the gray level image to a binary image. The image consists of two groups of pixels per algorithm. A bi-modal histogram is used to (front and background pixels). We may evaluate the optimal threshold, and also the combined distribution (intra-class variance) that is negligible or equivalent, by isolating two classes. The enhanced leaf image is occurred by this classifier. The common leaf image and the enhanced image are shown in Figures 6 (a) and (b).

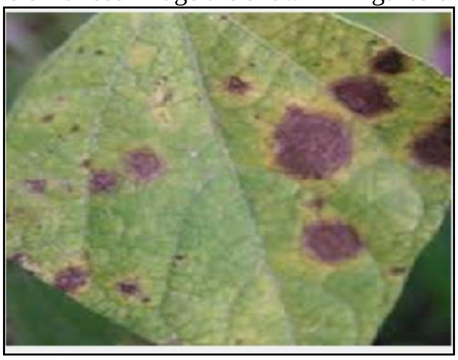


Figure 6(a): Normal leaf

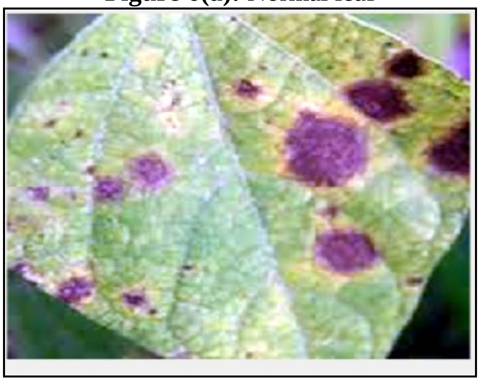


Figure 6(b): Enhanced leaf

Further, by using, K-means clustering divides the leaf into 3 clusters, i.e, 3 portions- healthy part, diseased part, and the background As shown in Figure 7. We then choose the diseased region from the 3 clusters to process it further.

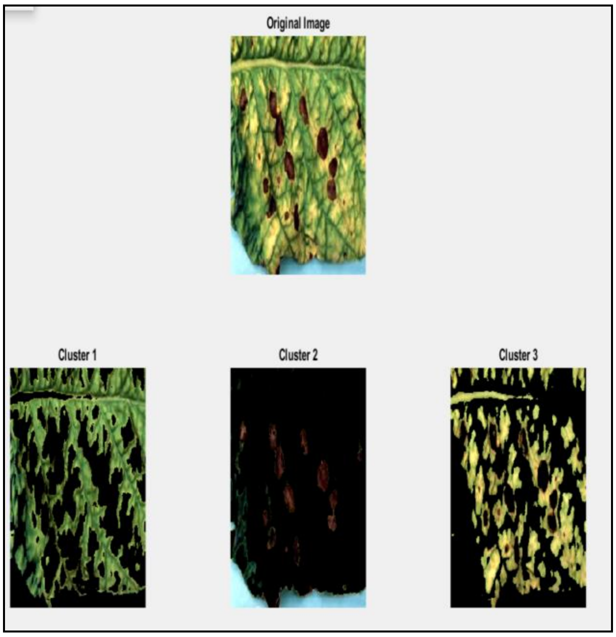


Figure 7: Clusters after Segmentation

5.2 Feature Extraction

Here the most important aspect is predicting the contaminated area. Gray Level Co-Occurrence Matrix uses the extraction of shape and characteristics of texture (GLCM). It illustrates the brightness of the image pixel. It was constructed as angles in degrees at a distance d=1, as shown in Figure 8, (0,45,90,135). Entropy, energy, contrast, correlation, etc. are the different measures offered. At various angles, these measurements are measured.

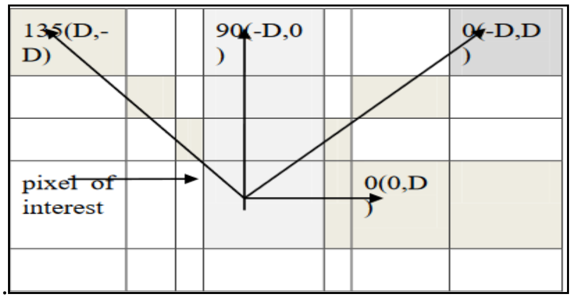


Figure 8: Construction of GLCM

At 0 degrees, first Fetchangle horizontally. Here element(1,1) has value1 in the output since there is 1 opulence in the input image where two horizontal images are similar to a 1-pixel gap and having values as 1 and 1. GLCM(1,2) has a value of 2 since there is two opulence in the input image where two pixels have values of 1 and 2 close to distance 1 horizontally. As there is no opulence in the input image, GLCM(1,3) has a value of 0, where two horizontally equivalent pixels at distance 1 have values of 1 and 3. Here, at various angles, the process is repeated for the whole GLCM matrix. The above implication is shown in Figures 9(a) and (b).

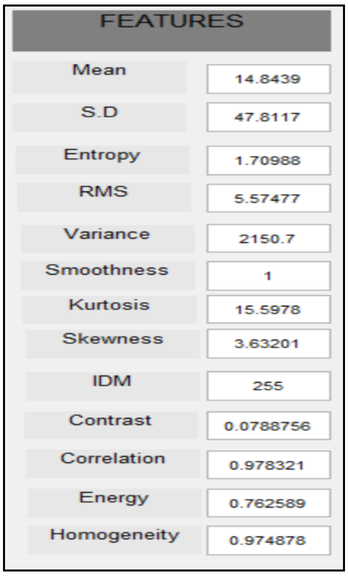
1	1	5	6	8
2	3	5	7	1
4	5	1	1	2
8	5	1	2	5

1	2	0	0	1	0	0	0
0	0	1	0	1	0	0	0
0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0
1	0	0	0	0	1	2	0
0	0	0	0	0	0	0	1
2	0	0	0	0	0		0
0	0	0	0	1	0	0	0

Figure 9(a): GLCM-1 **Figure 9(b):** GLCM-2

Extract features from the GLCM above, such as contrast, correlation, homogeneity, energy, etc. is shown in Figure 10.

5.2.1 Contrast: Contrast is referred to as CON. The other name for Contrast is Sum of Square Variance. The entire image is extended by measuring the intensity contrast that links the pixel to its neighbor.
5.2.2 Correlation: Calculating the overall correlation of the image between a pixel and its neighbor means that the linear grayscale dependence is measured on the neighboring pixels.
5.2.3 Energy: It's orderly because energy is used to do the job**.** In an image that calculates the orders, this makes it possible to use the texture. In GLCM, the number of square items is displayed. It is thought to be the square root of the ASM (Angular Second Moment) and is used as energy.
5.24 Homogeneity: It was named HOM. The value that measures the leakage within the distribution GLCM, elements are transferred diagonally across the GLCM.
5.2.5 Skewness: It is a measurement of symmetry. The dataset is identical to the left and right of the central dot, which is symmetrical.
5.2.6 Kurtosis: It's a measure of data versus a normal distribution, whether it's heavy or light. Datasets with strong kurtosis appear to have heavy tails or outliers and for low kurtosis, they tend to have light tails or a lack of outliers.
5.2.7 Average: The average pixel intensity values are Mean.
5.2.8 Standard Deviation: This function represents the deviation (or) the variance that the input frame in pixels.
5.2.9 Entropy: Describes the randomness of the grayscale distribution**.** If the gray levels of a picture are randomly distributed, entropy is high.
5.2.10 IDM: The reverse moment of the difference is a measure of the texture of the image, usually called homogeneity. The measurement of the proximity of the distribution of the elements at the GLCM diagonal is obtained by the IDM characteristics.
5.2.11 Variance: Variance is used to estimate how each pixel differs from the adjacent pixel (or central pixel) and is used to categorize it into different regions.
5.2.12 Smoothing: The relative measurement of proxies and distortions taken to make an image suitable for processing is smoothing.



FEATURES	
Mean	14.8439
S.D	47.8117
Entropy	1.70988
RMS	5.57477
Variance	2150.7
Smoothness	1
Kurtosis	15.5978
Skewness	3.63201
IDM	255
Contrast	0.0788756
Correlation	0.978321
Energy	0.762589
Homogeneity	0.974878

Figure 10: Features extracted for the query image

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.0789	0.9783	0.7626	0.9749	14.8439	47.8117	1.7099	5.5748	2.1507e+03	1.0000	15.5978	3.6320	255
2	0.4668	0.8657	0.7967	0.9592	14.1501	48.1396	1.3658	4.3136	1.6322e+03	1.0000	15.7654	3.6744	255
3	0.3676	0.9102	0.7573	0.9625	16.4441	51.4194	1.6679	5.3404	2.3050e+03	1.0000	13.7926	3.4025	255
4	0.5412	0.7510	0.5382	0.9222	17.9717	37.6635	2.5829	7.4037	1.3068e+03	1.0000	10.4951	2.5883	255
5	0.5128	0.7103	0.8947	0.9717	17.1185	35.5205	2.8432	10.4505	1.1622e+03	1.0000	27.6033	4.6820	255
6	0.6976	0.8739	0.4873	0.9104	31.5604	56.4596	2.9830	8.1140	2.8443e+03	1.0000	4.4008	1.6129	255
7	0.4886	0.9580	0.2687	0.9403	71.8528	83.0729	5.1204	11.4616	5.6827e+03	1.0000	1.8270	0.6497	255
8	0.4309	0.8966	0.7660	0.9656	17.4376	52.4639	1.8789	5.7289	2.0524e+03	1.0000	12.8361	3.2736	255
9	0.5761	0.9092	0.7104	0.9584	23.8136	60.2088	1.6734	5.4362	3.2303e+03	1.0000	6.9582	2.3349	255
10	0.7462	0.9098	0.5279	0.9007	40.0473	73.8575	2.9119	7.5330	4.4652e+03	1.0000	3.9932	1.5873	255
11	0.8894	0.8263	0.8185	0.9651	16.4181	55.6534	1.3002	4.3228	2.8415e+03	1.0000	12.3204	3.3010	255

Figure 11: Feature Dataset

Figure 11 above shows that the characteristics were extracted and stored in a table used to form the SVM.

Various types of SVM were compared and found that an SVM Multi-Class would be the perfect fit for our system to achieve the desired speed and accuracy. This SVM formulation in which the multi-class issue is resolved using a single optimization issue is called an all-in-one SVM classifier. It's called an all-at-ounce approach. It constructs k binary decision functions where/h gi (X) decision function separates the drive data points of the Ci class from the other data points of the class. Decision limits for a three-class classification problem using the all-in-one SVM approach.

5.3 Dataset Description

Data Type: JPG File(.jpg)

Number of Images: 4000

The set of illnesses contains healthy leaves and sick leaves. Some of these illnesses include Alternaria Alternate, Anthracnose, Bacterial Blight, Cercospora Leaf Spot. The sample data are illustrated in Figure 12 (a-d) and the healthy leaves in Figure 12 (e) and (f).

Figure 12(a): Alternaria Alternate

Figure 12(b): Anthracnose

Figure 12(c): Bacterial Blight

Figure 12(d): Cercospora Leaf Spot

Figure 12(e): Healthy Leaf 1

Figure 12(f): Healthy leaf 2

In the following Performance Evaluation Graph, the performance of individual classifiers using SVM and ANN classifiers with combined features according to True Positive Rate (TPR), False Positive Rate (FPR), accuracy, recall, F-measure, and Average Classification Accuracy (ACA) is given. The graph shows that the SVM classifier provided greater precision in the identification and classification of plant diseases affecting agricultural/horticultural crops than the ANN model. This comparison was carried out for the same dataset. The results are presented in Table 1 and the performance appraisal graphic is presented in Figure 13.

Table 1 Comparison of Results

	TP Rate	FP Rate	Precision	Recall	F-Measure
SVM	0.931	0.917	0.976	0.921	0.945
ANN	0.847	0.953	0.823	0.847	0.834

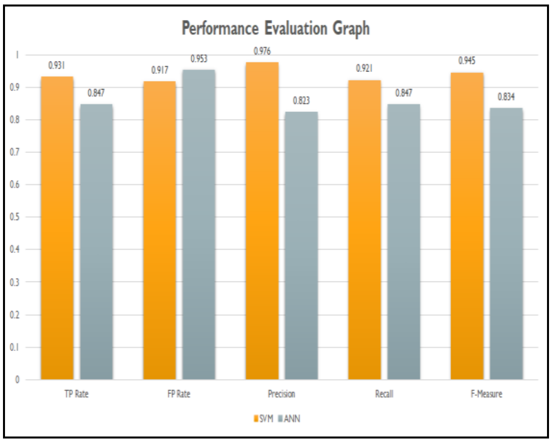


Figure 13: Performance Evaluation Graph

6. Conclusion

It detects and recognizes diseases in plants using the leaf image using image processing and machine learning concepts. We tested the system on various types of leaves and were reasonably satisfied with its operation and result. Also, performance and accuracy were higher than those of the existing NA method. This system is limited by the kind of diseases it can detect and recognize. Also, the request image should contain only the sheet; images containing additional data/information may result in inaccurate results. The part of the training is long and time-consuming. That being said, however, there is still scope for increments in the system. For example, the system could be formed with even more different types of diseases to extend its reach. Another might be to build the hardware part, such as a drone, and use it to permanently monitor the farm. The proposed system can lead to timely disease detection and potentially avoid a situation that could cause significant losses.

7. Conflict of Interest

Authors don't have conflict of interest.

References

1. Aroulanandam, V. V., Latchoumi, T. P., Bhavya, B., & Sultana, S. S. *Object Detection in Convolution Neural Networks Using Iterative Refinements. architecture, 15, 17.*
2. Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. (2019). *Solving current limitations of deep learning-based approaches for plant disease detection. Symmetry, 11(7), 939.*
3. Elangovan, K., & Nalini, S. (2017). *Plant disease classification using image segmentation and SVM techniques. International Journal of Computational Intelligence Research, 13(7), 1821-1828.*
4. Ferentinos, K. P. (2018). *Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311-318.*
5. Garikapati, P., Balamurugan, K., Latchoumi, T. P., & Malkapuram, R. (2020). *A Cluster-Profile Comparative Study on Machining AlSi 7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. Silicon, 1-12.*
6. Islam, M., Dinh, A., Wahid, K., & Bhowmik, P. (2017, April). *Detection of potato diseases using image segmentation and multiclass support vector machine. In 2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE) (pp. 1-4). IEEE.*
7. Kaur, R., & Kang, S. S. (2015, October). *An enhancement in classifier support vector machine to improve plant disease detection. In 2015 IEEE 3rd International Conference on MOOCs, Innovation, and Technology in Education (MITE) (pp. 135-140). IEEE.*
8. Latchoumi, T. P., Loganathan, J., Parthiban, L., & Janakiraman, S. (2016, August). *OFS method for selecting active features using clustering techniques. In Proceedings of the International Conference on Informatics and Analytics (pp. 1-4).*
9. Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., ... & Dandekar, A. M. (2015). *Advanced methods of plant disease detection. A review. Agronomy for Sustainable Development, 35(1), 1-25.*
10. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). *Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419.*
11. Padol, P. B., & Yadav, A. A. (2016, June). *SVM classifier-based grape leaf disease detection. In 2016 Conference on advances in signal processing (CASP) (pp. 175-179). IEEE.*
12. Patil, P., Yaligar, N., & Meena, S. M. (2017, December). *Comparison of performance of classifiers-SVM, rf, and ann in potato blight disease detection using leaf images. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) (pp. 1-5). IEEE.*
13. Pujari, D., Yakkundimath, R., & Byadgi, A. S. (2016). *SVM and ANN-based classification of plant diseases using feature reduction technique. IJIMAI, 3(7), 6-14.*
14. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). *Deep learning for image-based cassava disease detection. Frontiers in plant science, 8, 1852.*
15. Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N., & Vinod, P. V. (2018, April). *Plant disease detection using machine learning. In 2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C) (pp. 41-45). IEEE.*
16. Saleem, M. H., Potgieter, J., & Arif, K. M. (2019*). Plant disease detection and classification by deep learning. Plants, 8(11), 468.4*
17. Shruthi, U., Nagaveni, V., & Raghavendra, B. K. (2019, March). *A review on machine learning classification techniques for plant disease detection. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) (pp. 281-284). IEEE.*