Deep Learning Approach for the Detection of Plant Diseases

Padmavathi C.
Department of ECE, Sapthagiri College of Engineering
Bengaluru, V.T.U, Karnataka, India

ABSTRACT
Plant disease detection is one of the most active topics in the modern agriculture. The disease in plants are season-based which depends on the presence of the pathogen, crops, environmental conditions and varieties grown. The monitoring of leaf area is important in studying physiological capabilities associated with plant. This work makes use of image processing technique for the detection of disease and the use of Support Vector Machine for the classification of plant leaf disease. Plant Leaf disease detection and classification is performed, depending on various extracted features from plant leaves utilizing different image processing and deep learning techniques. Detection of plant leaf disease involves steps like data collection, image processing techniques like contrast enhancement, RGB to HSI, K-means clustering, feature extraction, segmentation and SVM based classification. This approach is useful when image dimensions are large and a reduced feature representation is required to efficiently complete tasks such as image matching and retrieval. The proposed work mainly concentrates on four major diseases that affect the plant leaf namely Alternaria alternata, Anthracnose, Bacterial blight and Cercospora leaf spot. The dataset considered for each disease is 22, 23, 20 and 20 respectively. The results of a test case for each of the four diseases are quantified and the percentage of disease affected area was observed to be 15.0013% in Alternaria alternata, 15.0015% in Anthracnose, 15.0142% in Bacterial blight and 23.0225% in Cercospora leaf spot.

Keywords
Plant leaf disease, K-means clustering, image processing, deep learning

1. INTRODUCTION
India is an agricultural country wherein most of the population depends on agriculture. Increase in productivity and food quality at reduced expenditure and increased profit are the main aims of agricultural research. Complex interaction of soil, seed, and agro chemicals are major components of agricultural production system. Product quality control is mandatory to obtain more valuable products. Many studies show that quality of agricultural products may be reduced due to plant diseases. Diseases are impairment to the normal state of the plant that interrupts its vital functions such as photosynthesis, transpiration, pollination, fertilization, germination etc. These diseases are caused by pathogens viz., fungi, bacteria and viruses, and due to adverse environmental conditions. Therefore, the early stage diagnosis of plant disease is an important task. Farmers require continuous monitoring of experts which might be prohibitively expensive and time consuming. Therefore, looking for fast, less expensive and accurate method to automatically detect the diseases from the symptoms that appear on the plant leaf is of great realistic significance.

Agriculture is the backbone of the Indian Economy. Almost the entire economy is being sustained by agriculture. It contributes 16% of the overall GDP and accounts for employment of approximately 52% of the Indian population. Proper monitoring of plants, detection of the plant disease, identification and amount of pesticide/insecticide used to increase the crop yield, contribute to the country’s economy. Manually the process is time consuming and involves manual labor. The work aims to learn about the challenges involved in image processing and machine learning frameworks involved in plant disease detection and understanding the technical feasibility of the system to diagnose diseases based on automated image recognition and subsequent implementation on a large scale thereby improving the performance and yield of agricultural sector.

The proposed work is to develop a plant leaf disease detection system using MATLAB and deep learning techniques which enables machine vision, that is to provide image based automatic inspection to detect the disease. This is very simple because the user clicks the image of the crop and sends it. The image is processed using the image processing techniques and the disease is detected. The details of the disease and amount of area affected are displayed and the user can see the details in the application. This may prove beneficial in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves.

2. BACKGROUND
Research in agriculture domain is aimed towards increasing the quality and quantity of the product at less expenditure with more profit. The quality of the agricultural product may be degraded due to plant diseases. Study and analysis of plant leaf disease detection using image processing work is carried out. Comparison of different detection technique of leaf disease detection is mentioned. SVM and k-means clustering is used in this system. The experiment results show that the algorithm is an efficient algorithm with high clustering accuracy. In the same time results obtained in standard and genetic versions of k-means algorithms relative to validity indices are also comparable.

The technology is needed for an automatic plant disease detection using image processing technology. This technique will give an alert to the farmers much before the disease starts to spread [1]. For data classification it is required to split the dataset into different classes. Here the WEKA Tool is used to classify the dataset using some algorithms like Nearest Neighbors Random Forest and J48 which extracts the models to perform the classification on different datasets [2]. Leaf
images are acquired in a digital camera and then images are pre-processed for enhancing the image quality. By using the classification, segmentation, clustering techniques, identifying the pixel, and masking the infected portion, the reports are statistically analyzed [3]. The image processing technique for groundnut plant disease detection namely early leaf spot (cercospora) and late leaf spot (cercoperidium personatum) is proposed. In [4] the leaf images in RGB are converted to HSV color images. In color and texture feature extraction analysis, co-occurrence matrices technique is used. In texture feature extraction there are two ways to analyze the texture images. First method is structured approach and second method is statistical approach. A statistical approach extracts quantitative features such as the number of horizontal, vertical, and diagonal segments which are then passed to a decision-theoretic classifier. A structural approach extracts morphological features and their interrelationships within each figure. Back propagation algorithm is applied for classification and recognition of groundnut diseases.

In [5] there are three different phases involved in detection of plant disease, first phase is to create color transformation structure for the RGB leaf image and convert Color values from RGB to the space specified in that structure. Then apply Color space transformation and image is segmented using the K-means technique. In the second phase called as Masking of green pixels, the unnecessary part such as green area within leaf area is removed. In third phase authors calculate the texture features for the segmented infected object also remove masked cells inside the boundaries of the infected cluster. Infected cluster is converted from RGB to HSI and SODM matrix is generated for H and S. In the fourth phase GLCM function is used to calculate the features and compute of texture statistics.

In [6] the image processing techniques is used to identify and classify fungal disease symptoms affected on different agriculture crops. They covered the fruit, vegetable, commercial, and cereal crops. In fruit crops, they applied segmentation using K-means clustering, then feature selection and texture selection, after that they used ANN and NN for classification. For the vegetable crops, they used the channel segmentation, then feature selection using local binary patterns and they have used SVM, k-NN for classification. For commercial crops, they have used a grab-cut for segmentation, a wavelet based on feature selection and Mahalanobies distance and PNN for the classifiers. Finally, for the cereal crops, they have used the k-means clustering, canny edge for segmentation, and for feature selection: Color shapes, texture, Color texture and radon transform and they have used SVM, k-NN for classification.

In [7], some segmentation and feature extraction algorithm is proposed to be used for the detection of plant diseases by using the image of leaves. The author has divided the entire process of plant leaf diseases detection into five steps: Image acquisition, Pre-processing, Segmentation, Feature extraction and Final classification of diseases. Image acquisition used the transformation structure for RGB leaf image. Then image is pre-processed to remove the noise and enhance the image contrast. Segmentation is done for the partitioning of image into various feature parts using k-means clustering, Otsu filters etc. This segmented image is further used for feature extraction and then final classification is performed using various classification techniques. In this way, plant diseases can be efficiently identified.

From the survey the plant diseases have become a major concern as it can cause significant reduction in both quality and quantity of agricultural products. The proposed work consists of the development of a classifier and the extracted features are passed through the classifier and the classifier outputs the disease the plant has been affected with.

### 3. METHODOLOGY

Initially, the infected plants are identified and captured. Since the captured images are of different size, shapes and are varying in many aspects, all the images are mapped to a normalized size. Normalized images are obtained after applying pre-processing techniques like enhancing the contrast and resizing the images. Segmentation is done on the region of interest by K-means clustering. Segmentation is followed by feature extraction. Classification is done by SVM method, first it goes through learning phase with existing data, and training and testing is done with new input data. Figure 1 gives the generalized flow of work for the detection of plant leaf disease.

![Flow of work for the detection of leaf disease](image)

**Fig 1: Flow of work for the detection of leaf disease**

Image acquisition is the process of selecting the plant which is affected by the disease and then the collection of the leaf and take a snapshot of leaf and load the leaf image into the system. In segmentation a digital image is partitioned into multiple segments which are defined as super-pixels. Generally, the pixel range of RGB is [0,255] and in this the pixel range is [0, 1]. Conversion of pixel range can be done by calculation of the components: Hue, Saturation and Intensity. K-means clustering algorithm is used to cluster or divide the object based on the feature of the leaf into k number of groups and is done by using the Euclidean distance metric. Support vector machine is used to map the various input images to the trained images depending on the symptoms.
**Fig 2: Algorithm for the detection of plant leaf disease**

### 3.1 Dataset
The dataset considered involves leaf samples of different plants infected with the following diseases Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora Leaf Spot.

**Fig 3.1: Alternaria Alternata**

**Fig 3.2: Anthracnose**

**Fig 3.3: Bacterial Blight**

**Fig 3.4: Cercospora Leaf Spot**

### 3.2 Preprocessing
Converting Red Green Blue (RGB) Image to HIS (H stands for Hue, S for Saturation and I for Intensity). Steps to be followed:
1. Read a RGB image
2. Represent the RGB Image in the range \([0 1]\)
3. Find HIS component

\[
\theta = \cos^{-1}\left(\frac{(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right)
\]  

(1)
\[ H = \begin{cases} \theta & \text{if } B < G \\ 360 - \theta & \text{if } B > G \end{cases} \]

\[ I = \frac{R+G+B}{3} \]

\[ S = 1 - 3 \times \min(R, G, B)/I \]

Equation 1, 2, 3, 4 give mathematical formulas for the calculation of theta, intensity, hue and saturation respectively which are the HSI components from RGB components.

### 3.3 Segmentation using K-means

Clustering

K-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups where each data point is a unique entity belonging to only one group. It tries to make the inter-cluster data points as similar as possible and also ensures that the clusters are as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster’s centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. Lesser variation within clusters, the more homogeneous and uniform (similar) the data points are within the same cluster. The way k-means clustering algorithm works as follows:

Step 1: Read the image.
Step 2: Get the number of clusters to be formed.
Step 3: Convert the Colour image into its corresponding gray image.
Step 4: Resize the two dimensional image into one dimensional array of length “r×c”.
Step 5: Find the intensity range of the image.
  \[ \text{Range} = [(\text{Maximum intensity value}) - (\text{Minimum intensity value})] \]
Step 6: Find the centroid value
  \[ \text{Centroid 1} = \text{Range/Number of clusters} \]
  \[ \text{Centroid 2} = (2 \times \text{Centroid 1}) \]
Step 7: Find the difference between the first intensity value and the various centroid values.
Step 8: Based on the minimum difference obtained, group the intensity values into the corresponding clusters.
Step 9: Repeat step 1 & 2 for all the other intensity values of the image.
  \[ \text{Centroid 3} = (3 \times \text{Centroid 1}), \text{Centroid n} = (n \times \text{Centroid 1}) \]

The approach k-means clustering follows to solve the problem is referred to as Expectation-Maximization. The E-step is the assignment of the data points to the closest cluster.

### 3.4 Evaluation of disease affected area

Generating the Gray Level Co-occurrence Matrices (GLCMs, \text{graycomatrix}(I)) create a gray-level co-occurrence matrix (GLCM) from image I. A gray-level co-occurrence matrix is also known as a gray-level spatial dependence matrix. \text{Graycomatrix} creates the GLCM by calculating the frequency of occurrence of a pixel with gray-level (grayscale intensity) value i occurs horizontally adjacent to a pixel with the value j. (it is possible specify other pixel spatial relationships using the 'Offsets' parameter.) Each element (i,j) in GLCM specifies the number of times that the pixel with value i occurred horizontally adjacent to a pixel with value j. Calculate the gray-level co-occurrence matrix (GLCM) for the grayscale image. By default, \text{graycomatrix} calculates the GLCM based on horizontal proximity of the pixels: [0 1] which is the pixel next to the pixel of interest on the same row.

### 3.5 Properties for enhancing and evaluation of disease affected regions of plant leaf

1. Contrast- Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image. Range \[= [0 \text{ size(GLCM,1)-1} \times 2]\] Contrast is 0 for a constant image. The property Contrast is also known as variance and inertia.
2. Correlation- Returns a measure of how correlated a pixel is to its neighbor over the whole image. Range \[= [-1 1]\] Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant image.
3. Homogeneity- Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Range \[= [0 1]\] Homogeneity is 1 for a diagonal GLCM. Statistics derived from the GLCM is returned as a structure with fields that are specific by properties. Each field contains a 1-by-p array, where p represents the number of gray-level co-occurrence matrices in GLCM. For example, if GLCM is an 8-by-8-by-3 array and properties is 'Energy', stats is a structure containing the field Energy, which contains a 1-by-3 array. Kurtosis, \text{kurtosis(X,flag)} specifies whether to correct for bias (flag is 0) or not (flag is 1, the default). When X represents a sample from a population, the kurtosis of X is biased, meaning it tends to differ from the population kurtosis by a systematic amount based on the sample size. The flag can be set to 0 to correct the systematic bias.

### 3.6 Support Vector Machine

The objective of the support vector machine algorithm is to identify a hyperplane in an N-dimensional space (N — the
number of features) that distinctly classifies the data points. To separate the data points of the two classes, there are many possible hyperplanes that could be chosen. The objective is to find a plane that has the maximum margin, i.e. the maximum distance of separation between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more accurately.

The use of classification algorithms is to make sense of and extract value from large sets of structured and unstructured data. Classification is carried out by the use of Support Vector Machine. Classification is used in the interpretation of the extracted diseased region in an image which helps in the identification of the type of disease in leaves. Analysis using Support vector machine is used which help in building association between known pattern of input and specific output.

4. RESULTS AND DISCUSSIONS

The proposed method is implemented using images of size 256X256. Data of Alternaria alternata (22 samples), Anthracnose (23 samples), Bacterial blight (20 samples), cercospora leaf spot (20 samples) is considered from different plant species. The white colors in the healthy and diseased images represent the portions of the leaf that are healthy and diseased respectively. The histogram of the hue image added additional color to the hue image for proper image visualization. Gaussian filter was included in the system, and simulation is run for the same image. This clearly shows that the original image contains some levels of unwanted signals in the form of higher frequencies and noise.

In order to validate the experimental results, considering that the goal is to automatically identify whether a leaf is infected or not each of the resulting images is used to extract a set of features that will be used later to classify the image, then validated against the human judgment of the infection in the image. Image features can be summarized as follows: Entropy, mean, standard deviation of the image after the first step of filtering. After the feature extraction and applying segmentation on the image, the image which seems to have the most disease affected part is selected for the further classification.

4.1 Alternaria Alternata

From Figure 5.2, the image which seems to be more affected by disease is selected.

Figure 5.1 is the image of the diseased plant leaf. After segmentation, in Figure 5.2, three different images are obtained. From that the image which seems to be more affected is chosen. In Figure 5.3 the disease Alternaria alternata is detected. Figure 5.4 quantifies the obtained result.

![Image of Diseased plant](image1.png)

**Fig 5.1: Image of Diseased plant**

**Fig 5.2: Based on segmentation 3 different images are given**

**Fig 5.3: Disease detected by the classifier**

**Fig 5.4: The calculated percentage of area that is affected by the disease**

**Alternaria Alternata**

**Affected Area is: 15.0113%**

![Image of Disease Area](image2.png)
by displaying the percentage of affected region in the input leaf image.

### 4.2 Anthracnose

![Contrast Enhanced Image](image)

**Fig 5.5: Image of Diseased plant**

![Cluster 1](image)

![Cluster 2](image)

![Cluster 3](image)

**Fig 5.6: Based on segmentation 3 different images are given**

From Figure 5.6, the image which seems to be more affected by disease is selected.

![Anthracnose](image)

**Fig 5.7: Disease detected by the classifier**

**Affected Area is: 15.0015%**

**Anthracnose**

**Fig 5.8: The calculated percentage of area that is affected by the disease**

Figure 5.5 is the image of the diseased plant leaf. After segmentation, in Figure 5.6, 3 different images are obtained. From that, the image which seems to be more affected is chosen. In Figure 5.7, the disease Anthracnose is detected. Figure 5.8 quantifies the obtained result by displaying the percentage of affected region in the input leaf image.

### 4.3 Bacterial Blight

![Contrast Enhanced Image](image)

**Fig 5.9: Image of Diseased plant**

From Figure 5.9, the image which seems to be more affected by disease is selected.
4.4 Cercospora leaf spot

Figure 5.9 is the image of the diseased plant leaf. After segmentation, in Figure 5.10, 3 different images are obtained. From that the image which seems to be more affected is choosen. In Figure 5.11 the disease Anthracnose is detected. Figure 5.12 quantifies the obtained result by displaying the percentage of affected region in the input leaf image.

From Figure 5.14, the image which seems to be more affected by disease is selected.

**Affected Area is: 15.0142%**

**Bacterial Blight**

Figure 5.13: Image of Diseased plant

**Cluster 1**

**Cluster 2**

**Cluster 3**

Fig 5.10: Based on segmentation 3 different images are given

From Figure 5.10, the image which seems to be more affected by disease is selected

Fig 5.11: Disease detected by the classifier

Fig 5.12: The calculated percentage of area that is affected by the disease

Fig 5.13: Image of Diseased plant

Fig 5.14: Based on segmentation 3 different images are given
digital images of diseased plant leaves to improve the database. As a part of enhancement, the complete process described can be automated so that the result can be delivered in a very short time and the size of the dataset is increased.

The K means clustering algorithm and Artificial Neural Network algorithms can be used to design an expert system for the farmers for the early detection of plant diseases. Presently four Diseases as mentioned earlier can be detected by this process. The K means clustering algorithm and Artificial Neural Network algorithms can be expanded for detection of multiple diseases on a significantly large scale. If this technique is developed into a sophisticated interface in the form of a website or an android application, it may prove to be great asset to the agricultural sector. In the future this methodology can be integrated with other yet to be developed methods for disease identification and classification.

6. REFERENCES


5. CONCLUSION

In order to detect the most common plant diseases like Alternaria alternata, Bacterial blight, Anthracnose and Cercospora a methodology is proposed. For efficient disease identification at various stages, the training samples can be increased with the optimal features given as input condition for disease identification. Major image processing techniques used for identification of leaf diseases are k-means clustering and SVM. A histogram segmentation method is proposed, this method can find appropriate threshold automatically rather than manually, and is more scientific, reliable, and efficient. By the use of k-mean clustering algorithm, the infected region of the leaf is segmented and analyzed. SVM is used for the classification of the disease that affects the plant leaf. This approach can significantly support an accurate detection of leaf disease. By using this concept, the disease identification is done for all kinds of leaves and also the affected area of leaf in terms of percentage is identified at low cost. ANN methods for classification of disease in plants such as self-organizing feature map, back propagation algorithm, SVMs etc. can be efficiently used for the identification of disease in a leaf. This method provides an efficient method to identify and classify various plant diseases using image processing techniques. The results are quantified by evaluating the percentage of affected area in the diseased leaf.

The future scope of the work is that the users in scientific agriculture fields will be given the option for adding the


