

**department of Computer Engineering & Applications**

**INSTITUTE OF ENGINEERING & TECHNOLOGY**

**B.Tech. IV Year CSE**

**Research Paper**

**On**

**“Plant Disease Detection”**

**Supervisor Name:**

Mr. Shashi Shekhar

**Submitted by:**

1.Devesh Thakur (161500195)

2.Rambandhu Sharma (161500438)

3.Rajeev Kumar (171599007)

4. Paritosh Panday (161500374)

**Group No: G-76**

**2019-2020**

## Abstract

*When plants and crops are affected by pests it affects the agricultural production of the country. Usually farmers or experts observe the plants with the naked eye for detection and identification of disease. But this method can be time processing, expensive and inaccurate. Automatic detection using image processing techniques provide fast and accurate results. This paper is concerned with a new approach to the development of plant disease recognition model, based on leaf image classification, by the use of deep convolutional networks. Advances in computer vision present an opportunity to expand and enhance the practice of precise plant protection and extend the market of computer vision applications in the field of precision agriculture. Novel way of training and the methodology used to facilitate a quick and easy system implementation in practice. All essential steps required for implementing this disease recognition model are fully described throughout the paper, starting from gathering images to create a database, assessed by agricultural experts, a deep learning framework to perform the deep CNN training. This method paper is a new approach in detecting plant diseases using the deep convolutional neural network trained and fine-tuned to fit accurately to the database of a plant's leaves that was gathered independently for diverse plant diseases. The advance and novelty of the developed model lie in its simplicity; healthy leaves and background images are in line with other classes, enabling the model to distinguish between diseased leaves and healthy ones or from the environment by using KNN. This paper also compares the benefits and limitations of these potential methods. It includes several steps viz. image acquisition, image pre-processing, feature extraction and neural network-based classification.*

*Keywords: Disease detection, Image acquisition, pre-processing, features extraction, classification, symptoms, and neural network.*

**Introduction**.

India is an agricultural country and ranks second worldwide in farm outputs. Agriculture employed more than 50℅ of the Indian workforce and contributed 17–18% to the country's GDP. Research in this sector is aimed to increase the productivity, food quality and quantity at reduced expenditures. When plants and crops in the field are suffering from pests it affects the agricultural production of the country. Studies show that the quality of products is reduced due to plant diseases. Diseases are impairment to the normal state of the plant that interrupts its vital functions such as photosynthesis, pollination, fertilization, etc. Pathogens (like fungi, bacteria and viruses), and adverse environmental conditions are a major cause of diseases. Therefore, the first stage diagnosis of disease is a crucial task. Usually farmers observe the plants with the naked eye for detection and identification of disease which calls for a tough decision on selecting the right fertilizer to use and this method can be time-consuming, expensive, and inaccurate. Automatic detection using image processing techniques provide fast and accurate results. This paper comes with a reviewed approach for plant disease recognition by leaf image classification. There are several ways to detect plant pathologies but some diseases don't have any visible symptoms, or the effect becomes noticeable too late to act, and in those situations, a sophisticated analysis is mandatory. This technique will improve the productivity of crops. The objective of this paper is to concentrate on plant diseases because timely and accurate diagnosis of plant diseases is one of the pillars of precision agriculture. It is crucial to prevent unnecessary waste of financial and other resources, thus achieving healthier production.

We can use Machine Learning libraries like Tensorflow, Keras, OpenCV, which provides many inbuilt functions for image processing to detect plant disease. This paper is organized into the following sections.

Section 1-Introduction includes the importance of plant disease detection, leaf analysis, various sorts of leaf diseases, and its symptoms.

Section 2- A brief on recent work carried out in this area.

Section 3- Basic methodologies and various image processing techniques.

Section 4- Conclusion along with possible future directions.

**1.1. Plant diseases analysis and its symptoms**

The RGB image feature pixel counting techniques are extensively applied to agricultural science. Image analysis can be applied for the following purposes:

1. To detect plant leaf, stem, and fruit diseases.

2. To quantify affected area by disease.

3. To find the boundaries of the affected area.

4. To determine the color of the affected area

5. To determine the size & shape of fruits.

Following are some common symptoms of plant leaf diseases.

**1.1.1. Bacterial disease symptoms**

* Bacterial ooze
* Water-soaked lesions
* Bacterial streaming in water from a cut stem

**1.1.2. Viral disease symptoms**

* Mosaic leaf pattern
* Crinkled leaves
* Yellowed leaves
* Plant stunting



(a)Bacterial leaf spot (b) mosaic virus

**Figure 1.** Bacterial and Viral disease on leaves

**1.1.3. Fungal disease symptoms**

* Birds-eye spot on berries (anthracnose)
* Damping off of seedlings (phytophthora)
* Leaf spot (septoria brown spot)
* Chlorosis (yellowing of leaves)

(a) late blight (b) early blight (c) downy mildew

**Figure 2.** Fungal disease on leaves

**II. Literature review**

A Proliferation of literature is out there in plant disease detection. We have highlighted some of the key contributions. A method for detecting plant diseases early and accurately using diverse image processing techniques has been proposed by Anand H.Kulkarni. [1], where Gabor filter has been applied for feature extraction and ANN-based classifier has been used for classification with recognition rate up to 91%. F. Argenti, [2] proposed a quick algorithm for calculating parameters of the co-occurrence matrix by supervised learning and maximum likelihood method for fast classification. Homogenize techniques like Sobel and canny filter has been used to identify the sides by P. Revathi [3]. These extracted edge features have been used in classification to identify the disease spots. The proposed homogeneous pixel counting technique for the cotton disease detection (HPCCDD) algorithm has been used for categorizing the diseases. They claim an accuracy of about 98.1%. Tushar H Jaware [4] proposed a unique and improved k-means clustering technique to unravel low-level image segmentation. Spatial gray-level dependence matrices (SGDM) for extracting statistical texture features by Sanjay B. Dhaygude et al. [5]. RGB images are transformed into Hue Saturation Value (HSV) color space representation and showed the H, S and V components. Mokhled S. Al-Tarawneh [6] presented an empirical investigation of olive leaf spot disease using auto-cropping

segmentation and fuzzy c-means classification. RGB to Lab color space and the median filter is used for image enhancement. At the end present comparative assessment of fuzzy c-means and k-mean clustering. The fuzzy feature selection approach namely fuzzy curves (FC) and fuzzy surfaces (FS) are proposed to pick features of cotton plant disease by Yan-Cheng Zhang, [7]. Which resulted in reduced dimensional feature space. Back-propagation (BP) networks are used to classify the grape and wheat diseases by Haiguang Wang [8]. Also, by using the principal component analysis (PCA) dimensions of the feature data has been reduced. Texture features supported the local power spectrum of Gabor filters have been proposed by Simona E. Grigorescu [9] where complex moments, Gabor energy, and grating cell operator features are discussed. They concluded that the grating cell operator responded only to texture features. Detection of unhealthy region and classification using texture features has been proposed by S. Arivazhagan,[10]. Their algorithm has been tested on ten species of plants namely banana, beans, jackfruit, lemon, mango, potato, tomato, and sapota. 94.74% accuracy has been achieved by the Support Vector Machine (SVM) classifier. Dheeb Al Bashish,[11] developed a neural network classifier supported statistical classification and will successfully detect and classify the diseases with a precision of around 93%. A Research of maize disease image recognition of corn supported BP networks effectively identified by Song Kai,[12] where YCbCr color space technology is employed to segment disease spot, Co-occurrence matrix (CCM) spatial gray level layer is deployed to extract disease spot texture feature, and Back Propagation neural network has been used to classify the maize disease. The applications of K-means clustering, BP neural networks had been formulated for clustering and classification of diseases that effect on plant leaves by H. Al-Hiary,[13]. This provided adequate support for the accurate detection of leaf diseases. The proposed algorithm has been tested on five diseases viz. Early and late scorch, cottony and ashen mold, tiny whiteness. Menukaewjinda et al. [14] tried another ANN, i.e. back propagation neural network (BPNN) for efficient grape leaf color extraction with complex background. They also explore a modified self-organizing feature map (MSOFM) and genetic algorithm (GA) and located that these techniques provide automatic adjustment in parameters for grape plant disease color extraction. SVM is also found to be very promising to realize efficient classification of leaf diseases. 21 color, 4 shape and 25 texture features have been extracted by Haiguang Wang[15] and principal component analysis (PCA) has been performed for reducing the dimensions in feature processing, then back-propagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) all are used to spot diseases.

**III. Methodology**

In all there are five main steps used to detect plant leaf diseases. The processing scheme consists of image acquisition using a camera or web image. Pre-processing includes image enhancement and image segmentation, feature extraction, and classification. Finally, the presence of diseases on the plant leaf will be identified. In the initial step, RGB images of leaf samples were picked up. The step-by-step procedure is as shown below:

1) RGB image acquisition;

2) convert the input image into color space;

3) Segment the components;

4) obtain the useful segments;

5) Computing the texture features;

6) Configuring the neural networks for recognition.

**3.1. Image acquisition**

Firstly, the images of various leaves are acquired using a digital camera with the required resolution. The construction of an image database is dependent on the application. The image database is responsible for the better efficiency of the classifier which decides the robustness of the algorithm.

**3.2. Image pre-processing**

In the second step, this image is pre-processed for improving the image data that suppress undesired distortions, enhancement of some image features is important for further processing and analysis task. It includes color space conversion, image enhancement, and image segmentation. The RGB images of leaves are converted into color space representation. The purpose of the color space is to provide the specification of colors in some standard accepted way. RGB images are converted into the Hue Saturation Value (HSV). HSV model is an ideal tool for color perception. Hue is a color attribute that describes the pure color as perceived by an observer. The Basic Methodology YCbCr color system is a common color space, which is applied by the jpeg image. Y, Cb, and cr, indicates a luminance component and two-color component signals. Different from other color space, YCbCr color space is orthogonal, which fully takes important factors of composition of RGB from other colors into account. YCbCr color space model is often used in image compression [12]. A, U, and Cr components from LAB [6], UVL, and YCbCr color space used to extract affected leaf color with the purpose of less illumination effects [14]. Image segmentation helps in the representation of an image into something more meaningful and easier to analyze. As the premise of feature extraction and pattern recognition, image segmentation is also the fundamental approach of digital image processing. There are various techniques for image segmentation discuss below.

**3.2.1. Region based**

In this technique related pixels to an object are grouped. The area that is detected for segmentation should be closed. Due to missing edge pixels in region-based segmentation there will not be any gap at all. The boundaries are identified for segmentation. In every step at least one pixel is related to the region and is taken into account. After identifying the change in the color and texture, the edge flow is converted into a vector. Then these edges are detected for further segmentation.

**3.2.2. Edge based**

Segmentation also uses edge detection techniques. There are various techniques viz. gradient, log, canny, Sobel, Laplacian, Robert. In this technique the boundary is identified to be segmented. Edges are detected to identify the discontinuities in the image. For classification both fixed and adaptive features of support vector machine are used.

**3.2.3. Threshold based**

In this segmentation is done through the threshold values obtained from the histogram of those edges of the original image. It is the easiest method. So, if the edge detections are accurate then the threshold will be too. Segmentation through thresholding has fewer computations compared to other techniques. The only disadvantage of this technique is that it is not suitable for complex images.

**3.2.4. Feature based clustering**

Segmentation is also performed via. Clustering [8]. The image is converted into a histogram and then clustering is done. Pixels of the color image are clustered using an unsupervised technique Fuzzy C. This is applied to ordinary images which results in fragmentation if it is a noisy image or not. Then a basic clustering k-means algorithm is used for segmentation in textured images [6, 13]. It clusters the related pixels to segment the image. Segmentation is done through feature clustering and will be changed according to the color components. Segmentation is purely dependent on the characteristics of the image. Features are recorded for segmentation. The difference between the intensity and color values are used for segmentation. Improved k-mean used to solve low-level image segmentation [4]. For segmentation of color image use of fuzzy clustering technique is to iteratively generate color clusters using fuzzy membership function in color space regarding image space. Real-time clustering-based segmentation. A Virtual region is captured accurately for segmentation. The image is segmented coarsely by multi-thresholding. It is then refined by fuzzy C means clustering [6]. It can be applied to any multispectral images.

I. Image Acquisition

1. Pick up a leaf from plants

2. Capture leaf image

II. Image Preprocessing

1. Image Enhancement

2. Color space conversion

3. Image segmentation

III. Feature Extraction

IV. Classification

V. Diagnosis Of diseases

**3.2.5. Model based**

Markov Random Field (MRF) based segmentation also known as model-based segmentation. An inbuilt region smoothness constraint which is used for color segmentation. MRF is combined with edge detection for identifying the edges with accuracy [4].

**3.3. Feature extraction**

After segmentation, significant features are extracted and those features can be used to determine the meaning of a given sample. Image features usually include color, shape, and texture features. Currently most of the researchers are targeting plant leaf texture as the most important feature in classifying plants. With the help of texture features, plant diseases are classified into different types. There are various methods for feature extraction as discussed below.

**3.3.1. Texture analysis methods**

Textures are a pattern of non-uniform spatial distribution of differing image intensities, which focus mainly on the individual pixels of an image. Texture is termed as quantifying the spatial relationship between materials in an image. Some important properties in recitation of texture viz. uniformity, regularity, density, linearity, roughness, and frequency. The method to characterize texture force into four major categories such as structural, statistical, fractals, and signal processing.*Structural*: The structural models presume that textures are combinations of texture primitives. Conceptually, structural texture analysis is carried out into two major steps as the extraction of the texture elements and inference of the placement rule. Two different structural methods are considered. *Statistical*: Statistical type includes grey-level histogram, co-occurrence matrix, and run-length matrices for pattern extraction. *Fractals*: Many surfaces possess a quality of roughness and self-similarity at different scales. Fractals are very useful and popular in modeling these properties. *Signal processing*: Texture is especially suited for this type of analysis because of its properties which include spatial domain filters, Fourier domain, two-dimensional Gabor function.

**3.3.2. Texture feature extraction methods**

The methods which are used for extracting interesting and relevant features from the input

Image which is used for the extraction of texture features from images is called a texture feature extraction method.

*Color co-occurrence Method:* In texture analysis, the features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image [2], [10]. Grey Level Co-occurrence Matrices (GLCM) is a statistical method. It is an old feature extraction method used for texture classification. It calculates the relationship between pixel pairs in the image. The textural features can be calculated from GLCMs, e.g. contrast, correlation, energy, entropy, and homogeneity. However, in recent times, instead of using the GLCM individually, it is combined with other methods. Here are a few other implementations of the GLCM, other than the traditional implementation e.g. 1-D GLCM, second-order statistical GLCM. It can be also applied to the color co-occurrence matrix [12].

*Gabor Filters*: Gabor filters also known as the Gabor wavelets is a widely used signal processing method. The Gabor filters consist of parameters such as the radial center frequency, standard deviation, and orientation. As the signal processing method produces large feature size, Gabor filters require to be downsized to prevent dimensionality issues [9]. Principal Component Analysis (PCA) can be a good choice to downsize the feature space as it can be combined with other methods too [1]. *Wavelets Transform*: Another popularly used signal processing method in digital image processing and pattern recognition. Currently, it became an important feature to be used in texture classification. Several wavelet transforms are used nowadays such as Discrete Wavelet Transforms (DWT), Hear wavelet, and Daubechies wavelets. Among these DWT is the most widely used wavelet transform. Therefore, despite being more complex and slower, wavelet transforms usually produces better features with a higher accuracy [3].

*Principal component analysis*: PCA is a statistical method that uses an orthogonal transformation to convert correlated variables into a set of linearly uncorrelated variables called principal components. It is sensitive to the relative scaling of the original variables [15]. PCA is the simplest of the true eigen vector-based multivariate analyses; PCA is closely related to factor analysis [8].

**3.4. Classifier**

A software cycle was written in which training and testing is performed via several neural network classifier. Texture Feature Classification Methods are as follows.

**3.4.1. Convolutional Neural Network**

It is a neural network that has convolution input layers acts as a self-learning feature extractor directly from input images. Hence, it can perform both feature extraction and classification under the same architecture.CNN takes advantage of local spatial coherence in the input (often images), which allows them to have fewer weights as some parameters are shared. This process, taking the form of convolutions, makes them especially well suited to extract relevant information at a low computational cost.The procedure is as follows:

1. Convolutional Operation
2. Pooling
3. Flattening
4. Full Connection

**3.4.2. Radial basis function**

A radial basis function (RBF) is a function whose value depends only on the distance from the origin. Usually used measuring norm is Euclidean distance. RBF's are the networks where the activation of hidden units is based on the distance between the input vector and a prototype vector [15].

**3.4.3. Artificial neural networks**

Multilayer Perception (MLP) is the basic form of ANN that updates the weights through backpropagation during the training time. There are variations in neural networks, which have recently, became popular in texture classification. *Probabilistic Neural Network (PNN)*: Derived from the Radial Basis Function (RBF) network, it has a parallel distributed processor that has a natural tendency for storing experiential knowledge. PNN is an implementation of a statistical algorithm called kernel discriminate analysis in which the operations are organized into a Multi-layered feed-forward network having four layers viz. input layer, pattern layer, summation layer, and output layer. *Backpropagation network (BPN)*: It consists of three parts: an input layer, hidden layer, and output layer which connect through the weighted values between nodes. The biggest characteristic of the BP network is that network weight value can reach expectations through the sum of error squared between the network output and the sample output, and then iteratively adjusted network structure's weight value [8,12,13]. It is widely popular and extensively used for training feed-forward networks.

**3.4.4. Support vector machine**

SVM is a non-linear classifier, and is new in the market. SVM is popularly used in many pattern recognition problems including texture classification [14]. SVM is designed to work with only two classes. This is done via. maximizing the margin from the hyperplane. The closest results to the margin that were chosen to determine the hyperplane are known as support vectors. Multiclass classification is applicable and is built up by two classes to solve the problem, either by using one-versus-all or one [10].

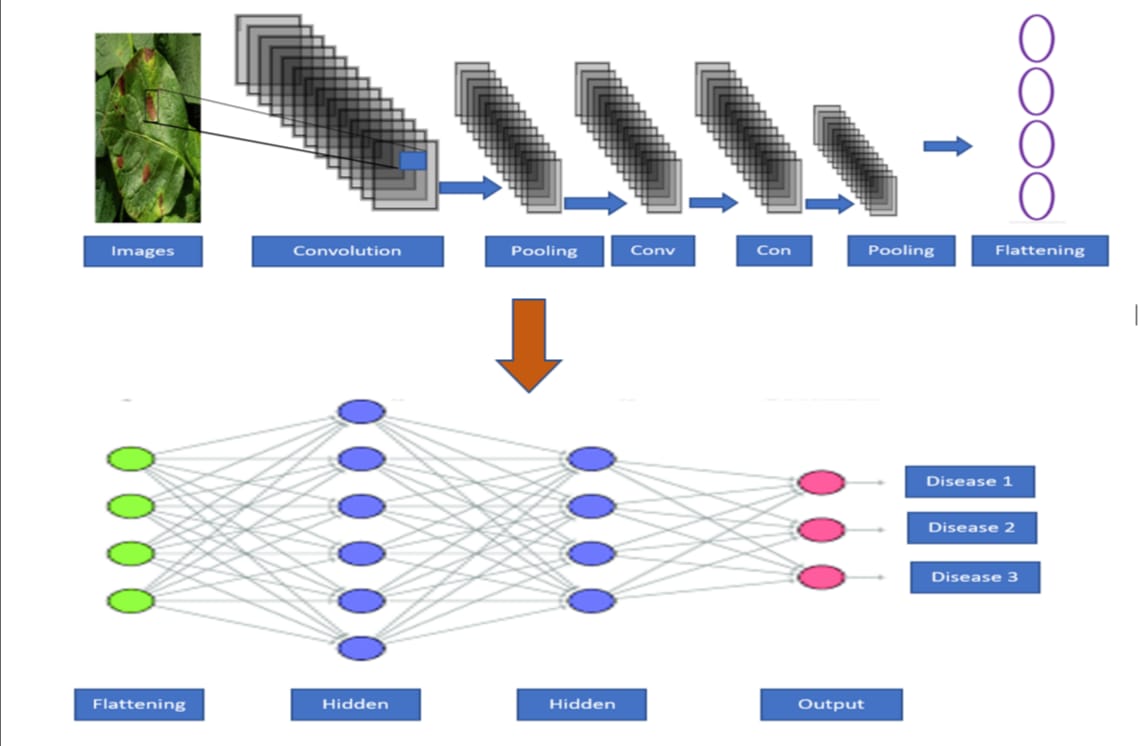
Table.1. Texture classification techniques comparison

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Technique** | **Advantages** | **Disadvantages** |
| 1 | Convolutional Neural Network  (CNN) | * local spatial coherence * low computational cost. | * Can not identify spatial patterns. |
| 2 | Radial Basis  Function  (RBF) | * Faster Training. * Hidden layer is easier to interpret. | * Slower in execution. |
| 3 | Probabilistic  Neural  Networks  (PNN) | * Tolerant of noisy inputs. * Instances are classified by more than one output. * Transformable to changing data. | * Long training time required. * Large complexity of network structure. * Need lot of memory for training the data. |
| 4 | Back  propagation  Network  (BPN) | * Easy to implement. * Applicable to wide range of problems. * Able to form arbitrarily complex nonlinear mappings | * Slow Learning rate. * Difficult to acknowledge how many neurons as well as layers are required. |
| 5 | Support  Vector  Machine  (SVM) | * Simple geometric interpretation. * Robust, even when training sample has become biased. | * Slow training rate. * Large no. support vectors are needed from training set to perform classification. |

**Image Preprocessing and Labelling [16]**:

Preprocessing images commonly involves removing low-frequency background noise, normalizing the intensity of the images, removing reflections, and masking portions of images. Image preprocessing is the technique of enhancing the data. Furthermore, the procedure includes cropping of all the images manually, making the square around the leaves, to highlight the region of interest (leaves). During the collecting phase of the images for the dataset, images with smaller resolution and dimension (< 500 pxls) were not considered as valid images only the images with higher resolution were marked as eligible candidates for the dataset. Therefore, it was ensured that images contain all the needed information for feature learning. As it is important to use accurately classified images for the training and validation dataset. Only in this way an appropriate and reliable detecting model is developed. In this stage, duplicated images that were left after the initial iteration of gathering and grouping images into classes were removed from the dataset.

**Working Stages of Algorithm(CNN):**

****

**IV. Conclusion**

The present paper reviews and summarizes the image processing techniques for several plant species which have been used for recognizing plant diseases. The major techniques for the detection of plant diseases are: BPNN, SVM, CNN, and SGDM. These techniques are used to analyze the plants leaves as healthy or diseased. Some of the challenges in these techniques viz. effect of background data in the resulting image, optimization of the technique for a specific plant leaf disease, and automation of the technique for continuous automated monitoring of plant leaf diseases under real-world field conditions. The review suggests that this disease detection technique shows good potential with an ability to detect plant leaf diseases. But, there is still scope of improvement in the existing research with advanced technologies.

**REFERENCES**

[1] Anand H. Kulkarni, Ashwin Patil R. K., Applying image processing technique to detect plant diseases, International Journal of

Modern Engineering Research, vol.2, Issue.5, pp: 3661-3664, 2012.

[2] F. Argenti, L. Alparone, G. Benelli," Fast algorithms for texture analysis using co-occurrence matrices" Radar and Signal Processing,

IEEE Proceedings, vol. 137, Issue 6, pp:443-448, No. 6, December 1990.

[3] P. Revathi, M. Hemalatha, Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques, IEEE

International Conference on Emerging Trends in Science, Engineering and Technology, pp-169-173, Tiruchirappalli, Tamilnadu,

India, 2012.

[4] Tushar H. Jaware, Ravindra D. Badgujar and Prashant G. Patil, Crop disease detection using image segmentation, National

Conference on Advances in Communication and Computing, World Journal of Science and Technology, pp:190-194, Dhule,

Maharashtra, India, 2012.

[5] Prof.Sanjay B. Dhaygude, Mr. Nitin P. Kumbhar, Agricultural plant Leaf Disease Detection Using Image Processing, International

Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, S & S Publication vol. 2, Issue 1, pp:

599-602, 2013.

[6] Mokhled S. Al-Tarawneh An Empirical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy

C-Means Classification, World Applied Sciences Journal, vol.23, no.9, pp:1207-1211,2013.

[7] Yan-Cheng Zhang, Han-Ping Mao, Bo Hu, Ming -Xi Li, Feature Selection of Cotton Disease leaves Image Based on Fuzzy feature

Selection Techniques, Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, pp:124-129,

Beijing, China, Nov. 2007.

[8] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li, Image Recognition of Plant Diseases Based on Backpropagation

Networks, 5th International Congress on Image and Signal Processing, pp-894-900, Chongqing, China, 2012

[9] Simona E. Grigorescu, Nicolai Petkov, and Peter Kruizinga, Comparison of Texture Features Based on Gabor Filters, IEEE

transaction on Image Processing, vol. 11, No. 10, pp-1160-1167, Oct 2002.

[10] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini, Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, Commission Internationale du Genie Rural (CIGR) journal, vol. 15, no.1,

pp:211-217, March 2013.

[11] Dheeb Al Bashish, M. Braik, and S. Bani-Ahmad, A Framework for Detection and Classification of Plant Leaf and Stem Diseases,

2010 International Conference on Signal and Image Processing, pp: 113-118, Chennai, India, 2010.

[12] Song Kai, Liu zhikun, Su hang, Guo chunhong, A Research of maize disease image recognition of Corn Based on BP Networks,

Third International Conference on Measuring Technology and Mechatronics Automation, pp:246-249, Shenyang, China, 2011.

[13] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, Fast and Accurate Detection and Classification of Plant

Diseases, International Journal of Computer Applications, Wageningen Academic Publishers, vol. 17, no.1, pp: 31-38, March 2011.

[14] A. Menukaewjinda, P. Kumsawat, K. Attakitmongcol, A. Srikaew, Grape leaf disease detection from color imagery using hybrid

intelligent system, Proceedings of electrical Engineering/electronics, Computer, Telecommunications and Information technology

(ECTI-CON), vol 1. pp: 513-516, Krabi, Thailand, 2008.

[15] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li, Image Recognition of Plant Diseases Based on Principal Component

Analysis and Neural Networks, 8th International Conference on Natural Computation, pp-246-251, Chongqing, China, 2012.

[16] https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761