*Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 3340-3347*

*Research Article*

# A Modern Approach for Detection of Leaf Diseases Using Image Processing and ML Based SVM Classifier

**Mohammed Khalid Kaleem[a], Nitin Purohit[b], Kassahun Azezew[c], Smegnew Asemie[d]**

[a]Assistant Professor, School of Computing and Informatics, Mizan Tepi University - Ethiopia
[b,c,d] Lecturer, School of Computing and Informatics, Mizan Tepi University – Ethiopia
[a]khalid_ccbc@yahoo.ca, [b]nitinpurohit111@gmail.com, [c]azeze2912@gmail.com, [d]simegnewasemie@gmail.com

**Article History**: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

**Abstract:** Agriculture maintains a key role in India, because it provides food to the human beings even there is a drastically increasing number of populations. So in the field of agriculture, plant disease detection techniques are used. But human eyes are not able to detect these leaf diseases exactly. Therefore machine learning and image processing techniques are used for disease detection with specialized algorithm accurately. Image classification and analysis techniques are implemented for detecting leaf diseases. In this paper a modern approach for detection of leaf diseases using Image processing and Machine learning (ML) based Support Vector Machine (SVM) classifier is analyzed. By using digital cameras leaf images are collected from the agricultural fields and background removal, filtering, enhancement are the different techniques which are preprocessed on images. K-means clustering process is used for segmentation on colour based, which can detect the disease effecting leaf. Statistical Gray-Level Co-Occurrence Matrix (GLCM) features are used in feature extraction with the help of image segmentation. SVM classifier is used in appropriate feature section as image texture and colour. Experimental results can be shown that the accuracy of leaf disease detection using this modern approach is better than other state-of –art-techniques.

Keywords: Leaf disease detection, Image processing, ML approach, K-means clustering, SVM classifier.

## 1. Introduction

One of economic sector supportive field all over the world is agriculture. Disease infected plants are must be recognized time to time in this agriculture. The overall crop production is reduced when plant get attacked by the disease. So these diseases in the plant must be detected in early [1]. This early detection can be processed through image processing in which leaf plant diseases are detected on the leaves. Various diseases are causing great social, ecological and financial losses. An accurate manner these diseases must be detected by using digital image processing. Based on detection, disease type can be analyzed and early prevention techniques are made [2]. In the farming field, an innovating computerized imaging technique is required. Therefore diseases in leaf plant can be detected in early stages and disease classification should be known by the formers. Some of infections are attacked in the plant leaf. The quality of the produced good is influenced by these infections. If these diseases are detected in early stage then its effect on the economy and quality is quite less. There are many advanced image processes are possible and utilized in the recognition and classification of leaf diseases. This examination process can be done on the leaves which decide the disease type. Image processing gives the best results and consistent technique for leaf plant disease detection [3].

Disease detection and recognition method is proposed in this paper in order to make good decisions in the detection. In some cases the leaf may affected by one or more diseases at a time which are having different characteristics then image processing is proceeds action with SVM classifier [4]. Thus the diseases or infections in the plant leaf can be detected. If inaccurate detection is made then massive amount of pesticides are used by the formers which effects the plant life cycle and its production. In the large fields, there are many plants existed. In these fields the sickness leaf location can be identified by the SVM classifier with the predefined automatic programming techniques are presented in this paper [5]. This process can be activated at the early stage of leaf spotted diseases. After recognizing disease effected leaf among all other leaves then the classification is required for feature extraction. Four main stages are included in the proposed system and these are

- Preprocessing
- segmentation
- feature extraction
- classification

In first step, different types of plant leaf diseases and their key challenges along the problems are discussed. By using K-Means algorithm the segmentation is processed in the second step [6]. Extraction of features are done by using Gray-Level Co-Occurrence Matrix (GLCM) in the last step, Support Vector Machine (SVM) can help in disease classification.

**2. Single-Phase AC Microgrid**

**2.1 Types of Plant diseases**

Researches can be done by the researchers in different plants on different diseases and as a result some techniques are introduced for recognizing the diseases. This researches can be better understand, when former get an idea about different types of plants and their diseases. Image processing operation type and feature type are identified in this section by observing a variety of diseases which can be understand by the researcher. Normal growth of the plant is effected by disease in the plant. If the plant is infected with bacteria or virus then that plant is called disease affected plant. The leaves of a disease affected plant may change its colour and lead to death. Because of viruses, microbes, nematodes and fungi, diseases are raised. Some common diseases in areca nut, Cotton, Tomato, coconut trees, Brinjal, Papaya, Chilli, Maize are discussed.

Different diseases are described in below sub sections.

Rust: It is usually found on leaves lower surfaces of mature plants. As time passes these spots become reddish orange spore masses. Later, leaf pustules turn to yellow green and eventually black. Severe infestations will bend yellow leaves and cause leaf drop [7].

- Kole Roga: In areca nut one of the major diseases is Kole Roga.
- Yellow leaf disease: By pathogen Phytoplasma in areca nut this disease caused. Where green leaves tuning into yellow that gradually decline in yield.
- Leaf rot: In coconut tree, this disease is appeared and by the fungi or bacteria it is caused. Main features are, leaf changes its shape, colour and size.
- Leaf curl: by genus Taphrina or virus, fungus this disease is caused and it is characterized by leaf curl [8].
- Angular leaf spot: in the cotton crops this disease is appeared and these plants die due to this disease because it appears on leaves. Finally turn black and form holes in leaves.
- Leaf spot: In chili crops this disease is spotted by bacteria and spread. The symptoms like small yellow green legions and patches on leaves.
- Late Blight: Its spreading is too fast. The development of the fungus due to Cool and wet weather. It forms irregularly shaped ashen spots signs on leaves. White mold is appeared around the spots as a ring [9].
- Bacterial wilt: In the Brinjal crops bacterial wilt is observed. Whole plant is going to die because of this bacterial wilt [10].

**2.1 Challenges in Disease Analysis**

Researches can be done by the researchers in different plants along with different diseases and as a result some techniques are introduced for recognizing the diseases [11]. From different sources the input data is collected for identifying disease. According to different research papers, the key issues are identified and different techniques are discussed in given below.

1. Plant leaves image must be high quality.

Large amount of data set is needed for consideration.

2. Background data and noises are affecting the images.
3. Disease of leaf can be identified by the Segmention. Input image can be trained and tested [12].
4. In the process of detecting disease in a leaf classification plays a vital role [13] [14].
5. According to weather conditions some plant leaves colour is changed.
6. Daily observation is required for some specific plants.
7. Different plants are having different diseases, so this detection process is little bit hard.
8. Based on reviews, image processing gives the best results with machine learning techniques [15].

All the above mentioned challenges in mind then the proposed approach is processed for getting high accuracy.

**3. Reactive Power Flow Optimization method for Single-Phase MMG Converters**

Plant leaf disease detection process is presented in this paper and this approach is processed in four stages as pre-processing of images which are collected from the different plant leaves, segmentation of images, image feature extraction and last one is disease classification according to different image collections. Work flow sequence by using ML is depicted in below Fig. 1.

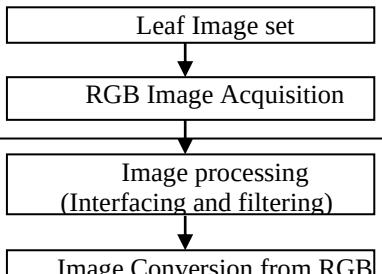

**Figure.1** Flow Work Sequence of Modern Leaf Diseases Detection Using ML

**3.1 Leaf Image set**

With the help of mobile camera, images of plant leaves are captured. 120 images are collected for the data set in that both healthy and damaged leaves are combined. So two types of image sets are inside the data set, in that first one is having collection of infected leaf images and second one is healthy plant leaf images.

**3. 2 Acquisition of RGB Image**

By using digital camera the data set images are collected. Agriculture field survey is required these type of images. The digital camera is used for collecting images which is having the resolution as 2 megapixels and above. This acquisition of images is the primary step in the proposed model. Infected leaves and healthy leaves images are contained by the dataset.

**3.3 Preprocessing of acquired Image**

The leaves images are pre-processed for getting better results by applying the pre-processing techniques to input leaves images. Image of leaf is stable condition and this can be achieved with image resizing. The troubles in the data set can be eliminated by filtering process and different pre-processing techniques like contrast enhancement, colour transformation and image cropping are used in the data set. Image is resized to 300 × 300 and quality image is obtained by contrast stretching.

**3.4 Image Segmentation**

Digital image can be divided into number of sections called segmentation and it is used to identifying and removing the effected portion in the leaf. K-means clustering technique is used in the segmentation process with $L * a * b$ color-space analysis of leaf by adding colour to the segmented leaf part. CIE XYZ tri-stimulus values are giving the $L^*a^*b^*$ color space. The $L^*a^*b^*$ space containing the radiance layer '$L^{*\prime}$, red-green pivot colour drops are discussed in chromaticity-layer '$a^*$, and blue-yellow pivot colour drops are discussed in chromaticity-layer '$b^*$ and equations are given below as.

$L^* = 116YY_0 3 - 16$ --- (1)

$a^* = 500XX_0 3 - YY_0 3$ --- (2)

$b^* = 200YY_0 3 - ZZY_0 3$ --- (3)

Where, for standard white $X, Y, Z$ values fraction $YY_0 > 0.01$, $X/X_0 > 0.01$, $Z/Z_0 > 0.01$, ($X_0$, $Y_0$, $Z_0$).

A colour space of each point is appeared as ($L^*a^*b^*$) three-dimensional colour space point and colour differences are calculated by the Euclidean separation as a goal. Normalized colour values between two colours to the white and this equation is as follows.

$\Delta E_{a,b} = (\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2$ ----- (4)

($l * a * b$) color space is obtained from the preprocessed RGB image because image visual differences and similar objects clustering process are measured in the colour space.

**K-means clustering algorithm:** Based on grouping of colours, diseased part, foreground and background. In a∗ and b∗ color space, objects (pixels) on leaf are classified by the K-means clustering algorithm. Segmented image part of every pixel labeled by K-means cluster. K means clustering algorithm involves the following steps as:

1. For a set of data points $X_1 \ldots\ldots X_n$, centroids and clusters j are initialized.
2. For the nearest centers clusters are assigned and j number of random clusters for the centroids $C_1 \ldots\ldots C_j$.
3. Euclidean distance d is calculated between the every pixel of image and center given as

$X_j(a) = \frac{1}{n_j}\sum_{x\ i} x(a), for\ a = 1,2, \ldots\ldots d$ ------- (5)

Where, jth cluster data points are denoted by $n_j$.

4. Mean can be calculated for finding the center of each cluster.
5. This process is continued till no changes happen in cluster.
6. Cluster pixels are again shaped into image.

### 3.5 Feature Extraction of Segmented Image

Extraction of image GLCM features from the image is processed after the segmentation of image. Spatial relationship of pixels is calculated because it is used in the texture investigating with the statistical method of Gray-Level Co-Occurrence Matrix (GLCM). From the matrix statistical measures are extracted. Changing direction pixel relationships are described in the array of offsets while creating the GLCM and distance is also calculated. Four features are commonly evolved in the feature extraction as energy, contrast, correlation and homogeneity. In the normalized Gray-Level Co Occurrence Matrix the $(i,j)^{th}$ entry is denoted by $P_{ij}$. The number of distinct gray levels is denoted by N in the quantized image. The extracted features are described in the below sub sections as:

a. **Contrast**: Intensity of the pixel and its neighbor pixel is measured by the contrast in full image. The contrast value of image is zero if constant. The contrast equation of a image is as follows,

$\text{Contrast} = \sum_{\mathrm{i,j=0}}^{\mathrm{N-1}}(\mathrm{P_{j,j}})\ (\mathrm{i}-\mathrm{j})^2$ ----- (6)

b. **Energy:** consistency is measured by the Energy by adding squared elements in the GLCM. Energy range is in between 0 and 1. For constant image Energy is 1. General equation Energy is as:

c. $Energy = \sum_{\mathrm{i,j=0}}^{\mathrm{N-1}}(\mathrm{P_{j,j}})^2$ ------ (7)

d. **Homogeneity:** pixels similarity is calculated by the Homogeneity. Range of Homogeneity is in between 0 and 1. Diagonal GLCM is having the Homogeneity value as 1. The Homogeneity equation as follows.

$Homogeneity\ = \sum_{\mathrm{i,j=0}}^{\mathrm{N-1}} \frac{(\mathrm{P_{j,j}})^2}{[1+(\mathrm{i}-\mathrm{j})^2]}$ --- (8)

e. **Correlation:** the process of intermixing pixels with other pixels is called as Correlation and its value range is in between -1 and 1.

$Correlation = \sum_{\mathrm{i,j=0}}^{\mathrm{N-1}} \mathrm{P_{jj}} \left(\frac{(\mathrm{i}-\mu)(\mathrm{j}-\mu)}{\sigma^2}\right)$ -(9)

In GLCM pixels mean value is denoted by μ and variance is represented by σ.

### 3.6 Machine learning Classifier

Machine learning classifier is having some steps to explore or access data, data pre-processing and predictive model development systems and analytics integration with iteration. Some problems are raised in the above steps as lack of domain tools and data diversity, platform diversity requirement and high time consumption. MatLab can eliminates the above mentioned difficulties by giving the extensive data support, workflow features of interactive app-driven, industry standard libraries with high quality and execution can be processed in anywhere by platform diversity.

**SVM:** set of images to be trained the system with different diseases and their features and different patterns in machine learning technique. For knowledge, standard data set is used by the system. Let's take leaf image of Custard apple data set with healthy leaf images and disease affected leaf images for the purpose of confirmation and training. The disease of plant leaves and their respective names are identified by the algorithm based on parameter of feature extraction. K-means clustering and support vector machine (SVM) techniques are used by the machine learning algorithm in order to detect disease classification and condition of affected leaf. Leaf disease detection process is uses the supervised learning of Support vector machine (SVM). High dimensional spaces are efficient, memory efficient and adaptability are acquired by the SVM and it is used for labeling. Therefore Support vector machine gives the more accurate results. Whereas K-means clustering is an unsupervised learning algorithm and it aims to divide the N observation into K-clusters. The similar feature data can be grouped to form the cluster.

Training data is assumed in SVM classification as $(x_i, y_i)$for i=1,2,……N and $y \in \{-1, +1\}$, where N is the sample number $y = +1$ for class C1 and $y = +1$ for class C2.

Minimum hyperplane calculation is given below where b is bias and v is vector.

$yi(v, xi + b) > 0, i = 1,2, \ldots . N$ ------ (10)

v and b values are again calculated then,

$$\min yi(v.xi + b) \geq 1, \quad i = 1,2, \ldots N$$

i.e. the distance between the point class and hyperplane to $1/|v|$ equation (1) can be,

$yi(v.xi + b) \geq 1$---- (11)

In high dimensional space hyperplanes are created by this technique which is used to divide the data points into several classes. The perfect hyperplane is detected by the Support vector machine with classification techniques. It is different from the data of different categories. The best hyperplane for SVM is having largest gap between two classes.

## 4. Results

Classification requires Training and Testing sections. Specified esteem and some highlighted esteems are prepared by the classifier in preparing stage. Test images are set to be in sequence after the classifier preparation. 130 images are contained by the classifier data set of proposed work. In this 30 images are representing healthy leaves, 50 images are leaves with Magnesium deficiency and remaining images are bacterial infected leaves. These images are used in classifier training but in testing classifier one image is randomly selected. 60 citrus leaves images, 25 normal leaf images and 35 disease affected leaves are used in the dataset at starting. Based on K-Means algorithm L*a*b* color space formats are designed from the RGB color images. Three clusters are selected. By using GLCM, texture features of cluster considering region are extracted. In the data base all features of GLCM are extracted and these images are stored. GLCM features as energy, correlation, homogeneity and contrast.

The inputs of classifier are all data base feature images. The disease affected leaves are recognized by the segmentation of the leaves. Classification process is activated after the classifier training with training vectors. By using this proposed method, an accuracy level of classification is in between 0.9 to 1.0. Support vector machine is used in the classification of leaves weather the leaf is affected by the disease or not. SVM classifies the diseases of leaves with GUI model. Preprocessed sample image is shown in Fig. 2. Segmentation of image with 3 clusters is displayed in Fig. 3.

(a) Diseased leaf image (b) Contrast enhanced

**Figure.2** Pre-processing Results

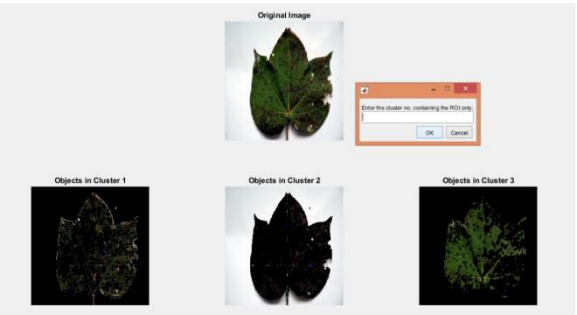

**Figure.3** GUI window for K-means clustering

The detection image results of some types of diseases by using the proposed modern approach are given as follows:

Bacterial Blight is shown in Fig. 4. Which is characterized by water-soaked pale and tiny inexpert spots appeared in the leaf. These are also called as dry dead spots and after that it expands to whole leaf length.

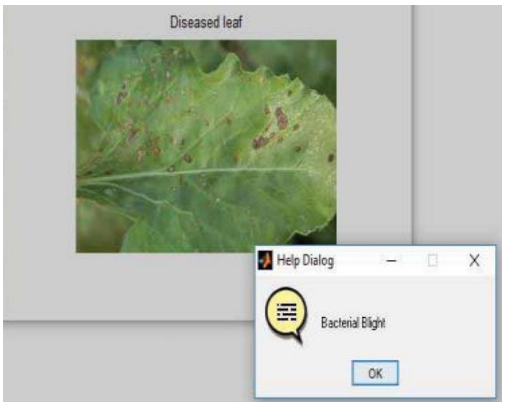


**Figure.4** Bacterial Blight Diseased leaf and Output

Spot diseased leaf of Cercospora is shows the brown spots in the leaf with bended margins of leaf. After 2 or 3 days the total leaf is turned into brown colour and shrink because progress in disease. Then flower components, stems and leaf petioles are also turned into brown colour. Fig.5. Shows the spot disease affected Cercospora Leaf.

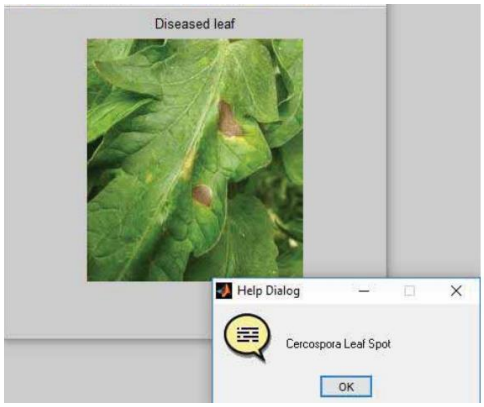


**Figure.5** Cercospora Leaf Spot Diseased Leaf and Output

The k-means clustering and SVM classifier used in the modern approach for classifying and detecting the various leaf diseases giving detection accuracy of 95.46%. In Table 1 the comparative results of proposed approach with different methods and their efficient accuracy are depicted. Graphical representation of accuracy with various approaches and proposed approach is depicted in Fig. 6.

**Table 1:** Comparative Analysis On Detection Accuracy

| **Label for method** | **Methods** | **Accuracy** |
|---|---|---|
| 1 | color co-occurrence method, Multi SVM Classifier | 85.5% |
| 2 | Artificial Neural Networks, GLCM | 79.96% |

| | | |
|---|---|---|
| 3 | K-means clustering, Naïve Bayesian classifier, color co-occurrence method. | 86% |
| 4 | K-means clustering algorithm with SVM, color co-occurrence method. | 89% |
| Proposed | GLCM+SVM+K-means (Proposed) | 95.46% |

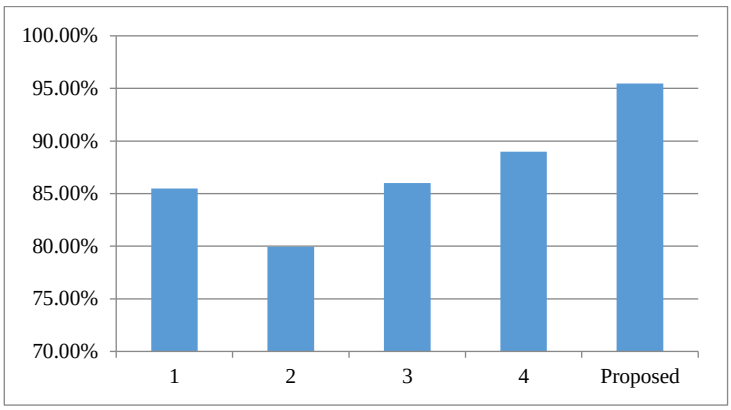


**Figure.6** Graph Representation of accuracy values

## 5. Conclusion

Leaf diseases detection method is analyzed in this paper by using the modern approach with the Image processing and ML based SVM classifier. Pre-processing, segmentation of image, feature extraction and classification of disease and its characteristics are the important steps in the proposed approach. The leaves images are pre-processed for getting better results by applying the pre-processing techniques to input leaves images. In some cases the leaf may affected one or more diseases at a time which are having different characteristics then image processing is proceeds action with SVM classifier. K-means clustering technique is used in the segmentation process with $L * a * b$ color-space analysis of leaf by adding colour to the segmented leaf part. By using GLCM, texture features of cluster considering Region are extracted. Extraction of image GLCM features from the image is processed after the segmentation of image. Four features are commonly evolved in the feature extraction as energy, contrast, correlation and homogeneity. So the better performance is given by the machine learning algorithm and it gives the knowledge and brief information about the disease which helps more to the formers with minimum efforts.

**References**

1. Yi-Bing Lin, Yun-Wei Lin, Wen-Liang Chen, Chun-You Liu, Fung-Ling Ng, "RiceTalk: Rice Blast Detection Using Internet of Things and Artificial Intelligence Technologies", IEEE IOT Journal, Volume: 7, Issue: 2, 2020
2. Aditya Sinha, Rajveer Singh Shekhawat, "Detection, Quantification and Analysis of Neofabraea Leaf Spot in Olive Plant using Image Processing Techniques", I nternational Conf. on Signal Processing, Computing and Control (ISPCC), 2019
3. Fenina Adline Twince Tobing, Yampi R. Kaesmetan, Marlinda Vasty Overbeek, "Identification of Maize Leaf Diseases Cause by Fungus with Digital Image Processing (Case Study: Bismarak Village Kupang District - East Nusa Tenggara)", 5th International Conf. on New Media Studies (CONMEDIA), 2019
4. N.M. Ghangaonkar, P. Raja, P. Maheswari, "Intelligent Disease Detection System for Early Blight of Tomato Using Foldscope: A Pilot Study", IEEE 4th International Symposium in Robotics and Manuf. Automation (ROMA), 2018
5. Rokeya Mumtahana Mou, M. Abdur Razzak, Selim Hossain, Sajib Chakraborty, Mohammed Mahedi Hasan, "Recognition and detection of tea leaf's diseases using support vector machine", IEEE 14th International Colloquium on Signal Processing & Its Apps. (CSPA), 2018

6. Jia-hui Sun, Shi-yu Zhou, Yuan-yuan Liu, "Detection of Ginseng leaf cicatrices base on K-means clustering algorithm", 10th International Congress on Image and Signal Processing, BioMedical Engg. and Informatics (CISP-BMEI), 2017
7. Medha V. Wyawahare, Shriroop C. Madiwalar, "Plant disease identification: A comparative study", International Conf. on Data Manag., Analytics and Innovation (ICDMAI), 2017
8. Muhammad Ilham Jamaluddin, Yeni Herdiyeni, Vektor Dewanto, Tri Setio, Bayo Alhusaeri Siregar, Budi Tjahjono, "An integrated smart surveillance system for diseases monitoring in tropical plantation forests", IEEE 17th International Conference on Comm. Tech. (ICCT), 2017
9. S.M. Meena, Nagaratna Yaligar, Priyadarshini Patil, "C omparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images", IEEE International Conf. on Computational Intelli. and Computing Reser. (ICCIC), 2017
10. J Aravinth, S Veni, R Anand, " An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method", International Conf. on Recent Trends in IT (ICRTIT), 2016
11. Harshadkumar B. Prajapati, Vipul K. Dabhi, Bhumika S. Prajapati, "A survey on detection and classification of cotton leaf diseases", International Conf. on Electrical, Electronics, and Optimization Techni. (ICEEOT), 2016
12. Antoine Vacavant, Manuel Grand-Brochier, Camille Kurtz, Guillaume Cerutti, Laure Tougne, Jonathan Weber, "T ree Leaves Extraction in Natural Images: Comparative Study of Preprocessing Tools and Segmentation Methods", IEEE Trans. on Image Processing, Volume: 24, Issue: 5, 2015
13. Sahaya Anselin Nisha, M. Ramakrishnan, "Groundnut leaf disease detection and classification by using back probagation algorithm", International Conf. on Comm. and Signal Processing (ICCSP), 2015
14. Lisa Hallau, Christian Bauckhage, Marion Neumann, Kristian Kersting, Benjamin Klatt, "Erosion Band Features for Cell Phone Image Based Plant Disease Classification", 22nd International Conf. on Pattern Recog., 2014
15. Shun'ichi Kaneko, Motoshige Shimizu, Rong Zhou, Miyuki Kayamori, Fumio Tanaka, "Early Detection and Continuous Quantization of Plant Disease Using Template Matching and Support Vector Machine Algorithms", First International Symposium on Compu. and Networking, 2013