IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AYURVEDIC LEAF CLASSIFICATION USING MACHINE LEARNING ALGORITHMS

Dr. Jai Ruby MCA., M.Phil. Ph.D and S.Ramila

Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.

***ABSTRACT*:** Ayurveda is an ancient system of medicine practiced in India. The main constituents of ayurvedic medicines are plant leaves and other parts of plants like root, bark etc. the Ayurvedic physicians themselves picked the medicinal plants and prepared the medicines for their patients. Today only a few practitioners follow this practice. Today the plants are collected by women and children from forest areas; those are not professionally trained in identifying correct medicinal plants. Manufacturing units often receive incorrect or substituted medicinal plants. Most of these units lack adequate quality control mechanisms to screen these plants. In addition to this, confusion due to variations in local name is also rampant. Some plants arrive in dried form and this make the manual identification task much more difficult. Incorrect use of medicinal plants makes the Ayurvedic medicine ineffective. It may produce unpredictable side effects also. In this situation, we have to know the ayurvedic leaf accurately. There are four steps involved. The first step is preprocessing that is by sharpening the RGB image. This is done by employing unsharp masking. This sharpening of the image improves the image appearance. Also, the edge or boundary points of an image are sharpened. The second step is Segmentation. Image segmentation is the process where image is divided into multiple parts. This is mainly employed to extract relevant information such as leaf portion from a digital image. Segmentation involves binarization. This binary image then passes through a morphological erosion and dilation process so that small imperfections like dots are removed. The third step is feature extraction like shape features, color features, texture analysis. Final step is classification. After feature extraction is completed, the values of these parameters will be stored in a temporary database as excel file. Based on these feature values, the image samples are classified into different classes and are ready to be tested for identification. The classification is done by using MLP and Decision Tree algorithm. Based on the correctly classified testing instances, accuracy is calculated as a measure of the efficiency of the algorithm.

***Index Terms* - Leaf Classification, Image processing techniques, Morphological operators, Machine Learning algorithms**

1. INTRODUCTION

Leaf recognition is a pattern recognition task performed specifically on leaves. It can be described as classifying a leaf either "known" or "unknown", after comparing it with stored known leaves. It is also desirable to have a system that has the ability of learning to recognize unknown leaves. Computational models of leaf recognition must address several difficult problems. This difficulty arises from the fact that leaves must be represented in a way that best utilizes the available leaf information to distinguish a particular leaf from all other leaves. Compared with other methods, such as cell and molecule biology methods, classification based on leaf image is the first choice for plant classification. Sampling leaves and photogene them are low-cost and convenient. One can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques. Some systems employ descriptions used by botanists. But it is not easy to extract and transfer those features to a computer automatically. It is difficult job to tell the just one algorithm alone is the best and successful at recognizing any and all variation of the same object. And it is more difficult to tell the same algorithm to be able to differentiate between different objects.

Currently there are a huge number of texture feature extraction methods available and most of the methods are associated with tunable parameters. It is difficult to find the most suitable feature extraction methods and their optimal parameters for a particular task. In addition, performance of classification methods also depends upon the problems, which makes selecting an optimal "feature extraction + classification" combination a difficult assignment.

2. LITERATURE REVIEW

The studies of plant disease refer to the studies of visually observable patterns of a particular plant. Nowadays crops face many traits/diseases. Damage of the insect is one of the major trait/disease. Insecticides are not always proved efficient because insecticides may be toxic to some kind of birds. It also damages natural animal food chains. A common practice for plant scientists is to estimate the damage of plant (leaf, stem) because of disease by an eye on a scale based on percentage of affected area. It results in subjectivity and low throughput.

A new automatic method for disease symptom segmentation in digital photographs of plant leaves,

In[1],they have proposed Traditionally, plant diseases were detected through visual inspection of plant tissue by trained experts this was costly and inappropriate paradigms as human intelligence is not perfect. This can be best solved using machine learning, where the image of infected plant's leaf is pre-processed and fit into neural network model for detection of disease. Many of the approaches used the same basic procedure, implementing basic supervised classification algorithms previously. In addition to this, artificial neural networks and support vector machine are also used for the image classification.

Leaf Disease Detection Using Image Processing

In[2],they have proposed India is fast developing country and agriculture is the back bone for the countries development in the early stages. Due to industrialization and globalization concepts the field is facing hurdles. On top of that the awareness and the necessity of the cultivation need to be instilled in the minds of the younger generation. Now a day's technology plays vital role in all the fields but till today we are using some old methodologies in agriculture. Identifying plant disease wrongly leads to huge loss of yield, time, money and quality of product. Identifying the condition of plant plays an important role for successful cultivation.

Texture Feature Extraction of RGB, HSV, YIQ and Dithered Images using GLCM, Wavelet Decomposition Techniques

In[3],they have proposed India is an agricultural country; where in about seventy percentage of the population depends on agriculture. Farmers have wide range of diversity to select suitable Fruit and Vegetable crops. However, the cultivation of these crops for optimum yield and quality product is highly technical. It can be improved with the aid of technological support. The management of perennial fruit crops requires close monitoring especially for the management of diseases that can affect production significantly and subsequently the post-harvest life. Cotton," The White Gold" or the "King of Fibers" enjoys a pre - eminent status among all cash crops in the country and is the principal raw material for flourishing textile industry. It provides livelihood to about sixty million people and is an important agricultural commodity providing remunerative income to millions of farmers both in developed and developing countries. This work exposes to automatic detection of disease on cotton leaves.

The Survey of Disease Identification of Cotton Leaf

In [4],they have proposed Cotton is an important cash crop in India. Disease on cotton is the main problem that decreases the productivity of the cotton. The main source for the disease is the leaf of the cotton plant. Without knowing about the diseases affected in the plant, the farmers are using excessive pesticides for the plant disease treatment. To overcome this, the detected spot disease in leaf are classified based on disease leaf types using artificial neural network,by this approach we can detect the leaf disease. The images required for this work are captured from the fields at Central Institute of Cotton Research Nagpur. We will take an input image of defected plant leaf and extract the features of leaf. With the help of this feature we will compare our defected plant leaf with data set. We will use Artificial Neural Network as our classifier for comparison of cotton leaf. An (ANN) artificial neural network also call (NN) neural network. We have created a database of cotton leaf disease considering two diseases they are Alternaria, foliar of cotton. We have extracted the separate H, Sand V feature and compared those feature with the feature that are extracted from the input test image.

An Identification of Variety of Leaf Diseases Using Various Data Mining Techniques

In [5], they have proposed India is an agricultural country where in most of the population depends on agriculture. Research in agriculture is aimed towards increase of productivity and food quality at reduced expenditure, with increased profit. Agricultural production system is an outcome of a complex interaction of soil, seed, and agro chemicals. Vegetables and fruits are the most important agricultural products. In order to obtain more valuable products, a product quality control is basically mandatory. 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 modifies or 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. Leaf presents several advantages over flowers and fruits at all seasons worldwide. This paper provides an introductory part includes importance of leaf disease detection; plant leaves analysis, various types of leaf diseases. By concludes this paper along with possible future directions.

Plant Disease Recognition Based on Plant Leaf Image

In [6], they have proposed India accounts approximately 25 percent of cotton land. Maharashtra is main cotton growing state in India. Cotton diseases are main problems for decreasing production of cotton. In many advance techniques for agriculture and medical field, image processing is used for multi-dimensional image analysis and applications. Hence not whole cotton plant but cotton leaves are mainly suffered from diseases like Alternaria leaf spot, fungus, foliar leaf spot etc. Disease management is a challenging task. Many diseases show visual symptoms of disease but it is quite difficult to precisely quantifying visual patterns. Hence there are many digital imaging techniques to identify such diseases precisely. This work represents agriculture atmosphere to sustain the farmers to identify the cotton diseases and get pest recommendation. Cotton foliar . diseases recognition and automatic classification using wavelet transform has been used for feature extraction while Support Vector Machine has been used for classification. The work shows that selforganizing feature map combined with a back-propagation neural network can be used for features extraction and those features are used to recognize the color of the image [1]. The modified selforganizing feature map is used to segmentation of cotton leaf disease with genetic algorithms for optimization and support vector machines for classification .

In [7], they have proposed As known, plants are very important for human beings. The photosynthesis of plants can maintain the balance of carbon dioxide and oxygen in the atmosphere. At the same time, plants are important resources of food and some products, and they also play a vital role in water conservation, inhibiting desertification and improving climate. However, the plant diseases many cause significant reduction in both quality and quantity of agricultural products (Ananthi and Varthini, 2012; Wang et al., 2008; Camargo and Smith, 2009; Arivazhaganet al., 2013). In 1943, in north eastern India, it is estimated that the outbreak of the rice helminthosporiose caused a heavy loss of food grains and death of a million people (Ananthi and Varthini, 2012). In 2007, in Georgia (USA), it is estimated that the plant disease losses was about $539.74 million, about $185 million was spent to control the diseases, and the rest was the value of damage caused by the diseases. So plant disease resistance and management are crucial to the reliable production of food. In fact, about 80% to 90% of disease on the plant is appeared on its leaves. So we interest in the plant leaf rather than whole plant. There are many leaf based plant disease recognition methods (Sabine Bauer et al., 2011; Al-Bashishet al., 2011; Al-Hiaryet al.,2011; Dheebet al., 2010; Arivazhaganet al., 2013). But, there is not an effective method because of the complexity of color and shape of the disease leaves. In this paper, a disease recognition method is proposed and the major steps of plant disease identification are introduced.

A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering

In[8],they have proposed The main objective of farming is feed ever the day by day growing population. Agricultural productivity is the main source of economy for developing country like India. Rice Crop is a staple food of almost over the entire population. So, disease found in leaves of Rice crop degrades both in quality and quantity of food product which directly affects the economy of the country. Therefore in the agriculture field, leaf disease identification and its management have turned into an important issue. Generally, Experts are observed through the naked eyes for the detection and identification of crop leaf diseases. But for this, the monitoring process is required Experts every time and it is too expensive in large fields. So in many under developed countries in the agricultural area, the farmer needs to take lots of efforts. This work will be described that how can we do the automatic detection of leaf diseases as this can give many benefits in monitoring large fields of crops and detect the symptoms of diseases If we are using the soft computing technique then the detection and the classification of the disease will be done faster at every stage. This is the best practical application of the image processing in the Agricultural industry. Different Image processing is used for automatic detection of Crop diseases and it will take less time, fewer efforts and become more accurate this paper explores, each of the stages involves and represents all the techniques.

Agricultural plant Leaf Disease Detection Using Image Processing

In[9],they have proposed Agriculture has played a key role in the development of human civilization. If there is decrease in agro products, total economy will get affected. Therefore judicious management of all input resources such as soil, seed, water, fertilizers etc. is essential for sustainability. As diseases are inevitable, detecting them plays major role. One can refer incident that occurred in 2007, Georgia (USA), it is estimated that approximately 539 USD was the loss incurred due to plant diseases as well as controlling them. The naked eye observation of farmers followed by chemical test is the main way of detection and classification of agricultural plant diseases. In developing countries, farming land can be much larger and farmers cannot observe each and every plant, every day. Farmers are unaware of non-native diseases. Consultation of experts for this might be time consuming & costly. Also unnecessary use of pesticides might be dangerous for natural resources such as water, soil, air, food chain etc. as well as it is expected that there need to be less contamination of food products with pesticides.

In [10], they have proposed India is a cultivated country and about 70% of the population depends on agriculture. Farmers have large range of diversity for selecting various suitable crops and finding the suitable pesticides for plant. Disease on plant leads to the significant reduction in both the quality and quantity of agricultural products. The studies of plant disease refer to the studies of visually observable patterns on the plants. Monitoring of health and disease on plant plays an important role in successful cultivation of crops in the farm. In early days, the monitoring and analysis of plant diseases were done manually by the expertise person in that field. This requires tremendous amount of work and also requires excessive processing time. The image processing techniques can be used in the plant disease detection. In most of the cases disease symptoms are seen on the leaves, stem and fruit. The plant leaf for the detection of disease is considered which shows the disease symptoms. This paper gives the introduction to image processing technique used for plant disease detection.

3. METHODOLOGY

The purpose of this chapter is to describe about modules in proposed system. A module is a part of a program. Programs are composed of one or more independently developed modules that are not combined until the program is linked. A single module can contain one or several routines. Each and every modules will explain about algorithm of work. The proposed system is developed by using following modules.

- PREPROCESSING
- SEGMENTATION
- FEATURE EXTRACTION
- CLASSIFICATION
  - DT
  - MLP

GRAY-LEVEL CO-OCUURRENCE MATRIX

This method was first proposed by Haralick in 1973 and still is one of the most popular means of texture analysis. The key concept of this method is generating features based on gray level co-occurrence matrices (GLCM). The matrices are designed to measure the spatial relationships between pixels. The method is based on the belief that texture information is contained in such relationships. Co-occurrence features are obtained from a gray-level cooccurrence matrix. We used 22 features that extracted from GLCM matrix in our paper

SHAPE FEATURE

Eccentricity

The eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. Eccentricity is the measure of aspect ratio. It can be calculated by principal axes method or minimum bounding rectangle method.

Eccentricity is the ratio of the length of the short (minor) axis to the length of the long (major) axis of an object:

Eccentricity = $\frac{\text{axislength}_{SHORT}}{\text{axislength}_{LONG}}$

- The result is a measure of object eccentricity, given as a value between 0 and 1.
- Sometimes known as ellipticity

Solidity

Scalar specifying the proportion of the pixels in the convex hull that are also in the region.

Solidity measures the density of an object.

• A measure of solidity can be obtained as the ratio of the area of an object to the area of a convex hull of the object:

$$\textbf{Solidity} = \frac{\text{Area}}{\text{Convex Area}}$$

Aspect Ratio

Eccentricity is the measure of aspect ratio, extent is the proportion of the pixels in the bounding box that are also in the region and circularity ratio represents how the shape is similar to circle. These three are translation, rotation and scale invariance, affine and occultation invariance and statistically independent.

$$\textbf{aspect ratio} = \frac{\text{Height}}{\text{Width}}$$

Elongation

In its simplest form elongation is the ratio between the length and width of the object bounding box:

$$\textbf{Elongation} = \frac{\text{width}_{\text{bounding-box}}}{\text{length}_{\text{bounding-box}}}$$

- The result is a measure of object elongation, given as a value between 0 and 1.
- If the ratio is equal to 1, the object is roughly square or circularly shaped. As the ratio decreases from 1, the object becomes more elongated.

Convexity

Convexity is the relative amount that an object differs from a convex object.

- A measure of convexity can be obtained by forming the ratio of the perimeter of an object's convex hull to the perimeter of the object itself:

$$\textbf{Convexity} = \frac{\text{convex perimeter}}{\text{perimeter}}$$

TEXTURE & COLOR FEATURE

Contrast

The contrast (Con) is defined in Equation, is a measure of intensity of a pixel and its neighbour over the image. In the visual perception.of the real world, contrast is determined by the difference in the colour and brightness of the object and other objects within the same field of view.

$$Con = \sum_{i=0}^{N-1} \sum_{j=o}^{N-1} (i-j)^2 M(i,j)$$

Mean and Standard Deviation

The **Mean** is the average value, so it tells us something about the general brightness of the image. A bright image will have a high mean, and a dark image will have a low mean.

Mean GL = sum (GLs .* pixel Counts) /number of Pixels

$$\text{M}=\left(\frac{1}{mxn}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x,y)$$

Standard Deviation

The **Standard Deviation**, which is also known as the square root of the variance, tells us something about the contrast. It describes the spread in the data, so a high contrast image will have a high variance, and a low contrast image will have a low variance.

sd = sqrt(variance GL)

$$\text{SD } (\sigma)=\sqrt{\left(\frac{1}{mxn}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (f(x,y) - M)^2}$$

Skewness

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right.

Skewness= sum ((GLs – mean GL). ^ 3.* pixel Counts) / ((number Of Pixels - 1) * sd^3)

$$\text{S}_{\text{k}}\,(\text{X})=\left(\frac{1}{mxn}\right) \sum \frac{(f(x,y)-M^3}{SD^3}$$

Kurtosis

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak.

Kurtosis = sum((GLs – mean GL) .^ 4 .* pixel Counts) / ((number Of Pixels - 1) * sd^4)

$K_{urt\ (X)}=(\frac{1}{mxn})\frac{\sum(f(x,y)-M^4}{SD^4}$

Contrast

Contrast measures the quantity of local changes in an image. It reflects the sensitivity of the textures in relation to changes in the intensity. It returns the measure of intensity contrast between a pixel and its neighborhood. Contrast is 0 for a constant image. It is the amount of local variation present in an image. If the amount of local variation is large, the contrast feature also has consistently higher values comparatively. If the gray scale difference occurs continually, the texture becomes coarse and the contrast becomes large. The texture becomes acute if the contrast has a small value.

$C_{on}=\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x-y)^2 f(x,y)$

Energy

Energy is defined based on a normalized histogram of the image. Energy shows how the gray levels are distributed. When the number of gray levels is low then energy is high.

$En=\sqrt{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f^2\ (x,y)}$

Homogeneity

$\mathbf{HOM}=\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\frac{f(x,y)}{1+||x-y||}$

Homogeneity is defined by name HOM. It passes the value that calculates the tightness of distribution of the elements in the GLCM to GLCM diagonal.

Root Mean Square

Rms average$=\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$

Root mean Square (rms) average is another way of calculating a mean for a set of numbers. The rms average of a series of numbers is the square root of the arithmetic mean of the squares of the numbers

Entropy (E)

$E=-\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f(x,y)\log_2 f(x,y)$

The entropy is a measure that tells us how many bits we need to code the image data. An image that is perfectly flat will have an entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images.

Variance (σ^2)

$\sigma^2=(\sqrt{\left(\frac{1}{mxn}\right)\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(f(x,y)-M)^2)^2}$

The dispersion of the value around the mean is represented by variance. It is used to descriptors of relative components

Inverse Difference Moment

$\mathbf{IDM}=\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\frac{1}{1+(x-y)^2}\ f(x,y)$

Inverse Difference Moment is a rate of the local homogeneity of an image.IDM can contain single or range of values so as to find out whether the image is textured or no textured.

Correlation (C_{orr})

$C_{orr}=\frac{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x,y)f(x,y)-M_x\ M_y}{\sigma_x\sigma_y}$

Correlation feature explains the spatial dependencies between the pixels.

4. RESULTS

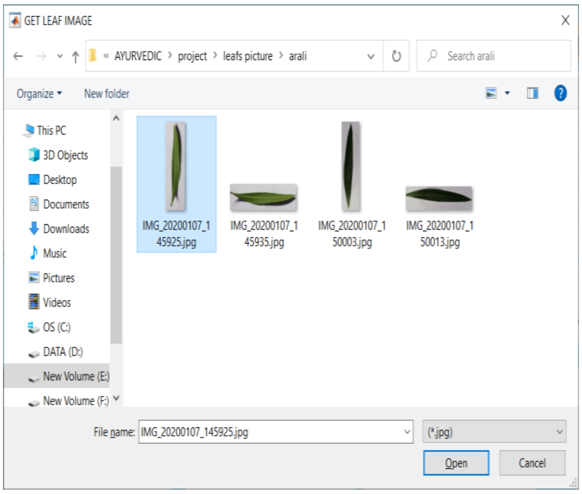

Figure 1 Input dataset

The above figure shows the dataset image selection by the user.

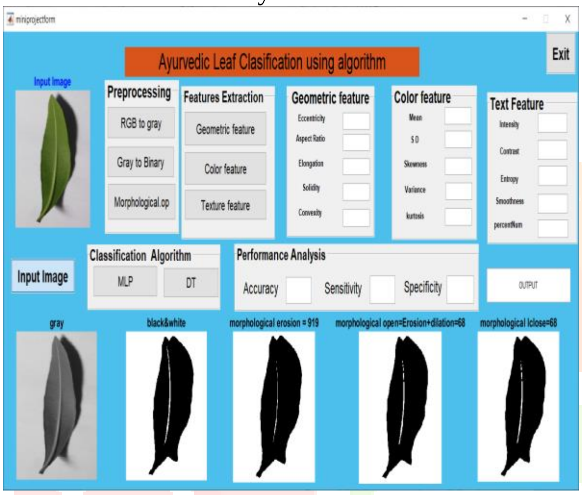

Figure 2 Preprocessing Output

The above figure shows the preprocessing such as RGB to gray image conversion, gray to binary conversion and morphological process output.

.5

Figure 3 Geometric Feature Output

The above figure shows the geometric feature extraction values.

Figure 4 Color and Texture feature ectraction and Performance of MLP

The above figure shows the color and texture feature extraction values. And also it displays MLP accuracy.

Figure 5 DT performance

The above figure shows the geometric decision tree accuracy, sensitivity and specificity.

5. CONCLUSION

This project has proposed and implemented a system for automatic identification of medicinal plants from their leaves. The proposed system makes use of computer vision and machine learning approaches to identify a pre-trained medicinal plant from its leaf. Main highlight of this work is the non-use of typical shape and color features of leaves which are computationally expensive to extract as they are spatial features. In this system, we used 25 samples for training and 25 for testing. This training and testing are based on the features namely entropy, contrast, correlation, solidity, eccentricity, extent, equivalent diameter, mean and standard deviation whose values are obtained from the images. This combination of features along with Decision Tree classifier provides an accuracy of 96.6677%. Highest identification rate is achieved using MLP Neural Network classifier for both dry and green leaves. Inclusion of both front and back features increase identification rate when green leaves are used to identify plants.

6. REFERENCE

1. Barbedo, Jayme Garcia Arnal, A new automatic method for disease symptom segmentation in digital photographs of plant leaves, European Journal of Plant Pathology, 2016, 1-16.
2. Bhong, Vijay S and Pawar B.V, Study and Analysis of Cotton Leaf Disease Detection Using Image Processing, International Journal of Advanced Research in Science, Engineering and Technology, 3 (2), 2016.
3. Lumb, Manisha, and Poonam Sethi, Texture Feature Extraction of RGB, HSV, YIQ and Dithered Images using GLCM, Wavelet Decomposition Techniques, International Journal of Computer Applications, 68 (11), 2013.
4. Reena Tijare, Pawan Khade, Rashmi Jain, The Survey of Disease Identification of Cotton Leaf, International Journal of Innovative Research in Computer and Communication Engineering, 2015.
5. Sasirekha N, Swetha N, An Identification of Variety of Leaf Diseases Using Various Data Mining Techniques, International Journal of Advanced Research in Computer and Communication Engineering, 4 (10), 2015.
6. Zhang S. W, Shang Y.J, and Wang L, Plant Disease Recognition Based on Plant Leaf Image, Journal of Animal and Plant Sciences, 25 (1), 2015, 42-45.
7. International Conference Animal and plant DSIAS International University, Xin Zheng Zheng Zhou, Henan 451150, P. R. China *Corresponding author:wjdw716@163.com
8. Joanna Sekulska-Nalewajko, Jaroslaw Goclawski, "A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering". Polyana-Svalyava (zakarpattya), Ukraine: pp.172-175, 2011 IEEE.
9. Prof. Sanjay B. Dhaygude, Mr.Nitin P.Kumbhar, "Agricultural plant Leaf Disease Detection Using Image Processing" IJAREEIE, vol. 2(1), pp. 599-602, January 2013.
10. M.E Student (Electronics & Telecommunication Engg.) Pimpri Chinchwad College of Engg. Savitribai Phule PuneUniversity 2015 International Conference on Computing Communication Control and Automation.