International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 5, ISSN No. 2455-2143, Pages 120-125
Published Online March-April 2017 in IJEAST (http://www.ijeast.com)

PCA AND RF: AN AUTOMATIC PLANT LEAF DISEASE DETECTION USING TEXTURE, SHAPE AND COLOR FEATURES

Dr. N. Palanivel,
Associate Professor,
Department of CSE,
Manakula Vinayagar Institute of Technology

Lavanya.S, Devapriya.E, Vinitha.M
Department of Computer Science & Engineering,
Manakula Vinayagar Institute of Technology

***Abstract*— The main objective of the project is classification of leaf, based on the characterization of texture, shape, and color properties. An original plant leaf is preprocessed initially using the cellular automata (CA) filter to minimize the noise. For enhancing the contrast and quality of the image, the histogram equalization are applied respectively. The GLCM and the LBP systems are introduced for the extraction of features. It has the issues of lower accuracy and recognition rate. The proposed feature extraction techniques overcome the difficulties faced by the existing method. The feature comprises of Haralick texture based features, Gabor features, shape features, and color features. The features are extracted from each leaf image, which increases the time complexity. Subsequently, the PCA(Principal Component Analysis) is presented to overwhelm the above issue of selecting the optimum features. Finally, the Random Forest is employed to characterize the type of leaves. The main objective of the proposed RF classification is to accurately predict the type of leaf from the given input leaf images. The experimental analysis showed better results such as accuracy, sensitivity and specificity of 99.87%, 99.5%, and 99.9% respectively, which are the improved values over the literature.**

***Keywords*— rf, GLCM, CA**

I. INTRODUCTION

In preservation of our earth ecology and environment, Plants plays important role. In the identification of the plants leaves are very essential. Visualization of plants is based on the significant shape of the leaves. Categorization process of leaves are very crucial work, it takes much amount of time in some agricultural based industries. In botanical research area, classification of leaves plays a paramount role, in order to identify their particular classes and their belonging families of the plants. Other than a shape, plant leaves also exhibits various features in order to identify their belongings. In general, the identification of the plants are done by its Morphological Characteristics(structure of stems, roots and leaves).Leaf consists much amount of information about the taxonomic plant characteristics. Leaves can survive in the plants for various months when compared with the flowers and fruits, because it survive for the little amount of days. Classification of plants is the technique, where the leaves are classified based on some features. Normally, Leaves are classified based on the texture, shape and color characteristics. There are various classification techniques, they are K-nearest Neighbor Classifier, Probabilistic neural network(PNN), Support Vector Machine(SVM)

Classification process of the leaf is a difficult task because, the out coming result varies for the different variety of input data. There are certain drawbacks for the above mentioned classifiers. K-nearest Neighbor classifier needs much cost to check for each and every instances and also it is very sensitive to noise. Due to this, it produces irrelevant inputs.PNN needs large network infrastructure and too many result attributes in over fitting of the network. SVM classifier has some speed and size limitations in both testing and training the data and it possess unclear understanding of the algorithm. In order to overcome the above mentioned classifiers drawbacks, the new technique is proposed.

In this proposed work, large amount of leaves are taken under classification process. Initially, the leaf is preprocessed using Cellular Automata (CA)filter. In order to enhance the contrast and brightness of the image, Histogram equalizer is used and also it removes the noise in the images as a preprocessing step. After segmentation, here comes the feature extraction process. This process extracts features such as texture, color and shape for providing more accuracy when compared with existing (GLCM and LBP). This process increases the time complexity. To overcome this issue, the feature selection technique called principal component analysis (PCA) is used. PCA is classification technique, which is used to classify the leaves in accuracy manner and also delivers it in high quality. The advantage of proposed system is Robustness, Efficiency, Localize, High Accuracy.

The main objective of the proposed system is to present the high accuracy using the incomplete training data, which can be easily achieved using PCA. Even the botanical research scientist need the much knowledge in order to achieve the mastery of this subject. The synopsis of this paper is, section 2

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 5, ISSN No. 2455-2143, Pages 120-125
Published Online March-April 2017 in IJEAST (http://www.ijeast.com)

explains the material and methods in which an automatic plant leaf detection is carried out are explained. section 3 explains the analysis of proposed system over existing system. At last the paper concludes with section 4 with the upcoming results of the paper.

II. PROPOSED ALGORITHM

The materials used for the automatic identification and classification of the plant leaf is taken from the ICL leaf dataset (Wanet al., 2015). The ICL leaf dataset consists of 220 species of leaf images from them 60 species of leaf types are taken for our experiment. It is the publicly available dataset, which is freely available and is the largest dataset containing of 220 species of leaf categories. The classification of leaf and the retrieval performances are evaluated on this data set. The camera used is with the normalized background and the actual size of the image is 200–500 both in horizontal and vertical directions. The minimum and maximum number of leaves used for our processing is 50 and 150 respectively. If the background noise is present in the images, then it is removed using the preprocessing stage.

2.1. An Automatic leaf detection using texture, shape, and color features

This section describes the overall flow explanation of the proposed methodology for an automatic leaf detection using the features such as the texture, shape, and color.

Fig 2 : Input image

The major components of the proposed method are:
_ Leaf image preprocessing and segmentation
_ Feature extraction
_ Feature selection
_ Classification

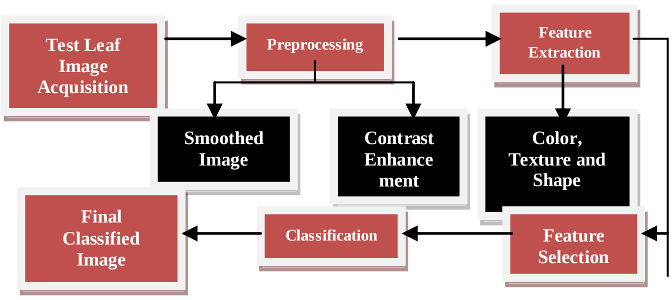


Fig 3:Detailed explanation of process

2.2. Leaf image pre-processing and segmentation

The images of the leaves collected requires the preprocessing stage due to the noise obtained during the image acquisition step. The leaf images are attained from natural geographical locations, where the changing illumination conditions over the day affects the quality of the image. The noise in the plant image is removed using the cellular automata filter, which is the preprocessing filter. The noise such as the shadows in the leaf imagesare removed using the preprocessing filters such as the cellular automata filter. A preprocessing or filtering step is applied to minimize the degradation related to the noise. There has been a lot of work in structuring the efficient noise suppression filters. This stage is necessary to enhance the leaf image quality and made the feature extraction component more reliable for the improvement of broad and narrow leaf image pruning.

2.2.1 Smoothing Image:

After gray image process, smoothed image is obtained. In image processing, a Gaussian blur (also known as Gaussian smoothing) is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise and reduce detail.

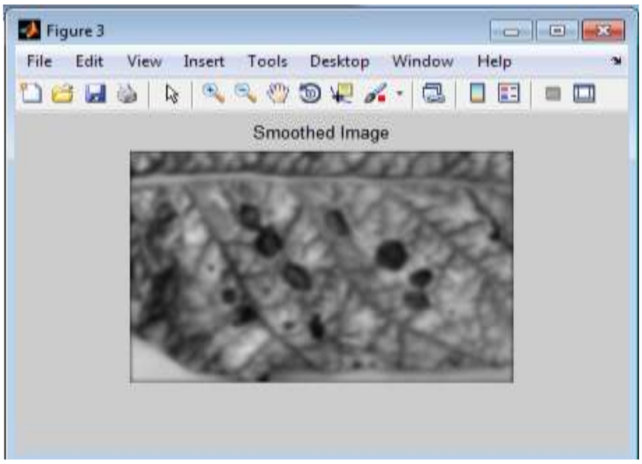


Fig 4:Smoothed image

2.2.2. Histogram equalization

The histogram equalization techniques are adopted to enhance the quality of the leaf image. It eliminates the background information, redundant and hidden details to process in a fast and easy manner. It also deals with enhancing the contrast of suspicious areas in the leaf image. It is used to increase the contrast of the image without affecting the information structure. Depicts the results of contrast enhancement after applying the histogram equalization technique.

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 5, ISSN No. 2455-2143, Pages 120-125
Published Online March-April 2017 in IJEAST (http://www.ijeast.com)

2.3. Feature extraction

The preprocessed images are analyzed using several textures such as the shape, and color features. In the proposed work, the texture features are extracted using Gabor wavelets.

2.3.1. Gabor-based Haralick texture feature extraction

Gabor filter is an outstanding technique to extract the features for analyzing the textures and also for detecting the edge. It determines the leaf image using Gabor functions that are similar to the System. In feature extraction stage, contrast enhanced image are processed. In this stage Color, Texture and shape features are extracted in enhanced image. In our proposed work, the texture features are extracted using Gabor filter. Finally encoding the extracted information into feature vectors for further process.

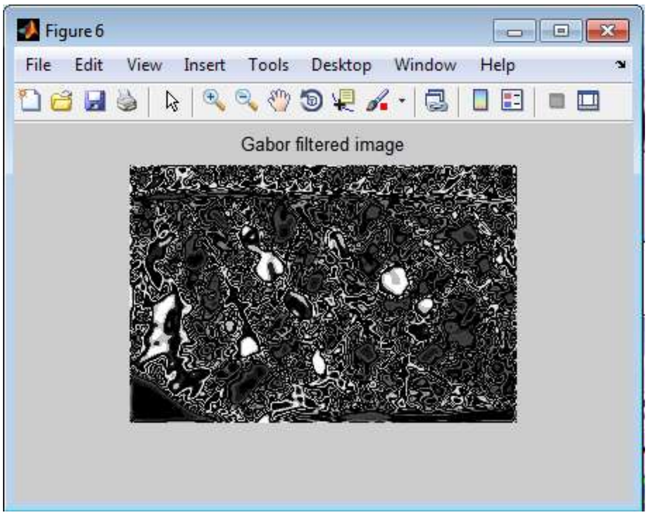


Fig 5: Gabor filtered image for texture extraction

Table 1 : Values computed by the Haralick texture features from Gabor applied images for a single leaf.

S.No.	Haralick texture features	Computed values based on the Haralick textures
1.	Energy	0.163696
2.	Correlation	0.306612
3.	Contrast	6.7495
4.	Entropy	2.79773
5.	Inverse difference	0.8655004
6.	Sum of average	5.02234
7.	Sum of variance	21.3322
8.	Sum of entropy	2.09952
9.	Difference in variance	6.787
10.	Difference of entropy	1.62541
11.	Information measures of correlation 1	-0.0805082

2.3.2 Color Feature Extraction:

The color spaces are used for producing the color histograms. Each color is represented as a mixture of the three primary color channels (Red, Green and Blue) in the RGB color space. The shortcoming of this scheme is the sensitivity to illumination changes. The RGB color space is converted into the HSV (Hue, Saturation, and Intensity) color model, which separates the intensity from the chromaticity. In color based, color HSV histogram features such as mean and standard deviation are extracted for preprocessed leaf image.

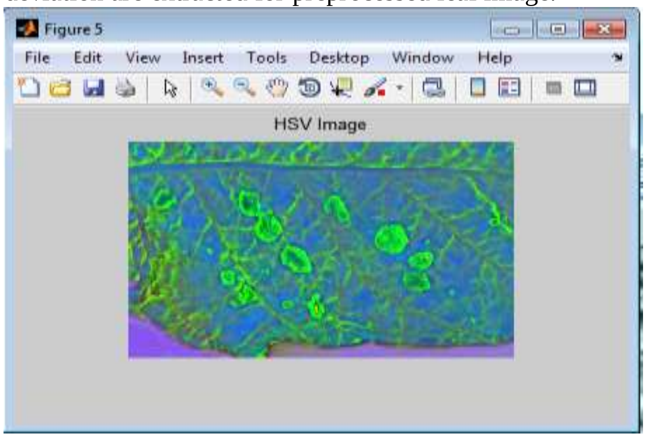


Fig 6:HSV image for color extraction

Color based Feature Extraction Algorithm

Input: RGB color image of plant leaf
Output: Color histogram
Begin
  Read the input plant leaf image in RGB color space;
  Convert it into HSV color space;
  Extract the color features Hue (H), Saturation(S), Intensity value (V) components;
  Use HSV bins and obtain the color histogram;
  Normalize the color histogram;
End

2.3.3 Texture Feature Extraction:

The texture of a plant is due to parallel lines of distinct colors and utilizes the Gabor wavelets for extracting the leaf textures. From the output of the Gabor filter, eight different orientations are obtained. Then, the Haralick feature is determined for each orientation of the image, which holds 8*13 texture features. From this obtained result, the mean value of image texture features is estimated, which holds 1*13 texture features. The 13 texture features involved in our proposed work is the energy, correlation, inertia, entropy, inverse difference, the sum of average, the sum of the variance, the sum of entropy, difference on average, difference in variance, difference of entropy, information measures of correlation 1, and information measures of correlation 2.

2.3.4 Shape Feature Extraction

The shape feature used in our proposed system is the area, centroid, eccentricity, equivalent diameter, extent, major axis length, minor axis length, orientation, and perimeter.

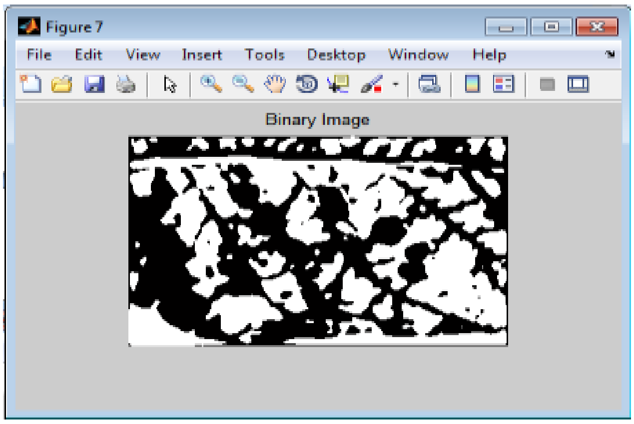


Fig 7:Binary image for shape extraction

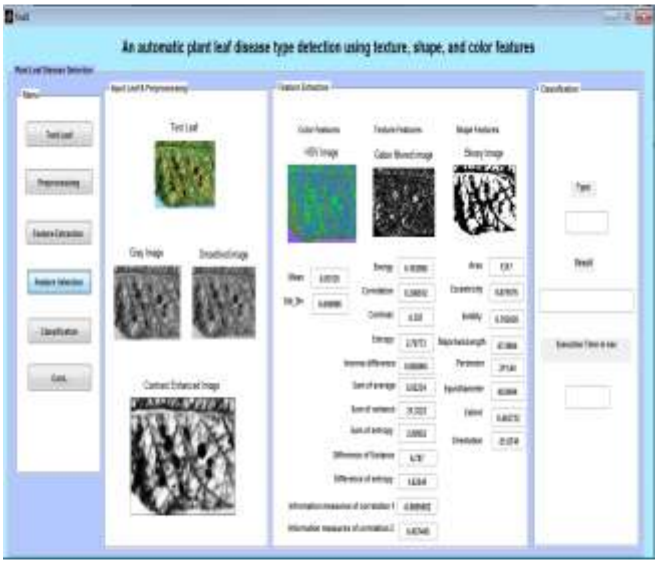


Fig 8:After extracting shape,color,texture features.

2.4 Feature Selection

As many numbers of features are extracted from each image, it may increase the time complexity. It is overcome by the propose PCA.

Feature Reduction using PCA

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated,

$$m_x = E\{x\} = \frac{1}{K}\sum_{K=1}^{K} X_K$$

Matrix A in Eq. (1) is determined by the covariance matrix Cx. Rows in the A matrix are formed from the eigenvectors e of Cx ordered according to corresponding eigenvalues in descending order. The evaluation of the Cx matrix is possible according to relation

$$C_X = E\{(x - m_x)(x - m_x)^T\} = \frac{1}{K}\sum_{k=1}^{K} x_k x_k^T - m_x m_x^T$$

As the vector x of input variables is n-dimensional it is obvious that the size of Cx is n x n. The elements Cx(i, i) lying in its main diagonal are the variances

$$C_x(i,i) = E\{(x_i - m_i)^2\}$$

of x and the other values Cx(i, j) determine the covariance between input variables xi, xj.

$$C_x(i,j) = E\{(x_i - m_i)(x_j - m_j)\}$$

between input variables xi, xj . The rows of A in Eq. (1) are orthonormal so the inversion of PCA is possible according to relation

$$x = A^T y + m_x$$

The kernel of PCA defined by Eq. (1) has some other interesting properties resulting from the matrix theory which can be used in the signal and image processing to fulfill various goals.

2.5. Classification

The classification process is done by extracted color, texture, and shape features in leaf. The main novelty here is the adoption of Random Forest.

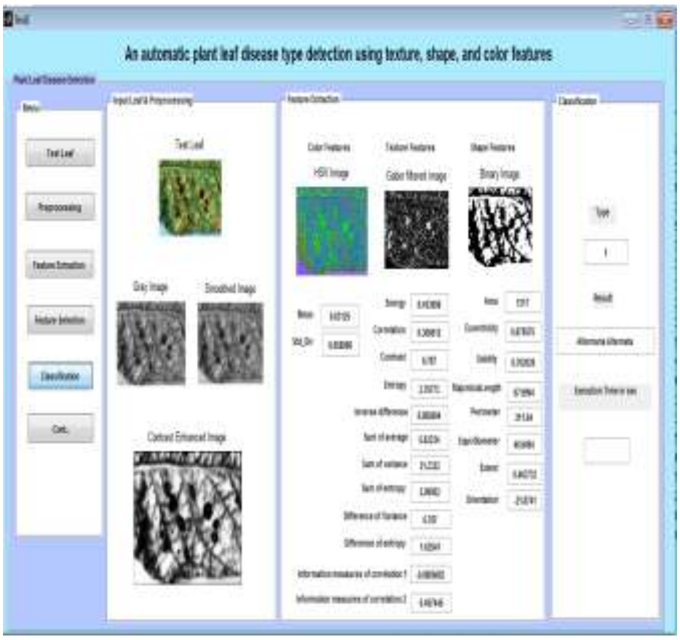
Random Forest

One of the foremost common ways or frameworks utilized by knowledge scientists at the 'rose knowledge science skilled follow cluster' is Random Forests. The Random Forests formula is one in every of the simplest among classification algorithms -able to classify giant amounts of information with accuracy. Random Forests are associate degree ensemble learning technique (also thought of as a kind of nearest neighbor predictor) for classification associate degreed regression that construct variety of call trees at coaching time and outputting the category that's the mode of the categories output by individual trees (Random Forests may be a trademark of Leo Bremen and Adele monger for an ensemble of call trees).

Features of Random Forests

International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 5, ISSN No. 2455-2143, Pages 120-125
Published Online March-April 2017 in IJEAST (http://www.ijeast.com)

- It is unexcelled in accuracy among current algorithms.
- It runs efficiently on large data bases.
- It can handle thousands of input variables without variable deletion.
- It gives estimates of what variables are important in the classification.
- It generates an internal unbiased estimate of the generalization error as the forest building progresses.
- It has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.
- It has methods for balancing error in class population unbalanced data sets.
- Generated forests can be saved for future use on other data.
- Prototypes are computed that give information about the relation between the variables and the classification.
- It computes proximities between pairs of cases that can be used in clustering, locating outliers, or (by scaling) give interesting views of the data.
- The capabilities of the above can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier detection.



III. CONCLUSION

In this paper, automated plant leaf type detection is proposed using the features of texture with Gabor, shape, and color. At first, the input image is preprocessed using the CA filter. The contrast of the leaf images are enhanced using the histogram equalization, and the image quality is improved by the segmentation. Furthermore, the features are extracted from each leaf image to maximize the time complexity. The PCA is used for selecting the best features, among various features. The selected features are classified corresponding to the category usingthe RF. The performance analysis shows that the proposed system offers higher accuracy, sensitivity, and specificity than the existing SVM technique. The proposed RF method is faster in classifying the types of the leaves, when compared to the existing methods. The true positive value is higher than the false positive value. The value of the area, when calculating the AUC curve is found to be 0.91. It describes various execution times of the existing classification techniques and the proposed RF classification techniques. It is clear from the table that the proposed classification scheme is fast and more accurate than the existing classification schemes. The main aim of the proposed classification technique is to classify the type of leaf. There are several issues during this research work, which is listed as below:

- The leaves with the formal structure and similar shapes are difficult to classify.
- If the leaves are affected by shadow or any disease, which change the color of the leaf is difficult to classify the type of leaf.
- The leaf with complicated backgrounds are difficult to identify.

In future, the proposed system can extend to classify the disease of the plant leaf by determining the effects of adding multiple leaf features and also extend the proposed work to classify the medicinal leaf images, which is used in the Ayurvedic medicines for curing the human diseases.

IV. REFERENCE

1) Ab Jabal, M.F., Hamid, S., Shuib, S., Ahmad, I., 2013. Leaf features extraction and recognition approaches to classify plant. J. Comput. Sci. 9, 1295–1304.
2) Agrawal, S., Shimpi, R., Modified Particle Swarm Optimization.
3) Ahmed, F., Al-Mamun, H.A., Bari, A.H., Hossain, E., Kwan, P., 2012. Classification of crops and weeds from digital images: a support vector machine approach. Crop Prot. 40, 98–104.
4) Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., ALRahamneh, Z., 2011. Fast and accurate detection and classification of plant diseases. Int. J. Comput. Appl. 17, 31–38.
5) Arribas, J.I., Sánchez-Ferrero, G.V., Ruiz-Ruiz, G., Gómez-Gil, J., 2011. Leaf classification in sunflower crops by computer vision and neural networks. Comput. Electron. Agric. 78, 9–18.
6) Arun, C., Sam Emmanuel, W., Christopher Dhurairaj, D., 2013. Texture feature extraction for identification of medicinal plants and comparison of different classifiers. Int. J. Comput. Appl. 62, 1–9.

7) Arun Priya, C., Balasaravanan, T., Thanamani, A.S., 2012. An efficient leaf recognition algorithm for plant classification using support vector machine. In: International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME), pp. 428–432.
8) Bama, B.S., Valli, S.M., Raju, S., Kumar, V.A., 2011. Content based leaf image retrieval (CBLIR) using shape, color and texture features. Ind. J. Comput. Sci. Eng. 2, 202–211.
9) Chaki, J., Parekh, R., 2012. Designing an automated system for plant leaf recognition. Int. J. Adv. Eng. Technol. 2, 149–158.
10) Du, J.-X., Zhai, C.-M., Wang, Q.-P., 2013. Recognition of plant leaf image based on fractal dimension features. Neuro Comput. 116, 150–156.
11) Elhariri, E., El-Bendary, N., Hassanien, A.E., 2014. ‘‘Plant classification system based on leaf features”. In: International Conference on Computer Engineering & Systems (ICCES), pp. 271–276.
12) Gwo, C.-Y., Wei, C.-H., Li, Y., 2013. Rotary matching of edge features for leaf recognition. Comput. Electron. Agric. 91, 124–134.
13) Hu, R., Jia, W., Ling, H., Huang, D., 2012. Multiscale distance matrix for fast plant leaf recognition. IEEE Trans. Image Process. 21, 4667–4672.
14) Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I., 2011. Leaf classification using shape, color, and texture features. Int. J. Comput. Trends Technol., 225–230.
15) Kalyoncu, C., Toygar, Ö., 2015. Geometric leaf classification. Comput. Vis. Image Underst. 113, 102–109.
16) Kebapci, H., Yanikoglu, B., Unal, G., 2010. Plant image retrieval using color, shape and texture features. Comput. J., 1–14.
17) Larese, M.G., Namías, R., Craviotto, R.M., Arango, M.R., Gallo, C., Granitto, P.M., 2014. Automatic classification of legumes using leaf vein image features. Pattern Recogn. 47, 158–168.
18) Lee, K.-B., Hong, K.-S., 2013. An implementation of leaf recognition system using leaf vein and shape. Int. J. Bio-Sci. Bio-Technol. 5, 57–66.
19) Wenjiang Huang, Qingsong Guan, Juhua Luo, Jingcheng Zhang, Jinling Zhao, Dong Liang, Linsheng Huang, and Dongyan Zhang, “New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases”, IEEE journal of selected topics in applied earth observation and remote sensing,Vol. 7, No. 6, June 2014
20) Dr.K.Thangadurai, K.Padmavathi, “Computer Visionimage Enhancement For Plant Leaves Disease Detection”, 2014 World Congress on Computing and Communication Technologies.
21) Monica Jhuria, Ashwani Kumar, and Rushikesh Borse, “Image Processing For Smart Farming:Detection Of Disease And Fruit Grading”, Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013
22) Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff Rohani Binti S Mohamed Farook, “Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques”, 2012 Third International Conference on Intelligent Systems Modelling and Simulation.
23) Mrunalini R. Badnakhe, Prashant R. Deshmukh, “Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.