RESEARCH ARTICLE OPEN ACCESS

Detection of Affected Part of Plant Leaves and Classification of Diseases Using CNN Technique

A.Blessy[1], Dr. D.C. Joy Winnie Wise[2]

[1]A.Blessy, PG Scholar, Francis Xavier Engineering College, Tirunelveli.
[2]Dr. D.C.Joy Winnie Wise, DEAN, Francis Xavier Engineering College, Tirunelveli

Abstract—

Plant pathology is the scientific study of plant diseases caused by pathogens and environmental conditions. It includes pathogen identification, diseases cycles, economic impact, management of plant diseases, etc. In existing, to detect the diseases they used the spectroscopic techniques. These techniques are very expensive and can only be utilized by trained persons only. This project presents the detection of diseases which are detected using CNN (ConvolutionNeuralNetwork) technique. First the sample leaf image is given as input. Then, color channels are separated from the leaf image from these the green pixels are masked from the original image. The masking is done to avoid the processing of the green area of the leaves, since, it is healthy. By removing the green area from the original area the remaining infected area is calculated. Then the features are extracted from the affected area. Finally these features are given to the CNN to classify the disease. After finding the disease the solution for those diseases are send the corresponding user mobile using GSM device.

Keywords — **Leaf disease, Fuzzy C-means clustering, GLCM, Convolution Neural Network**

I. INTRODUCTION

Developing countries like India the economy is greatly rely on agriculture. The quantity and quality of agricultural product is reduced as a result of disease. Plant disease is caused by micro-organism like fungi and microorganism .the lifecycle of micro-organism is unable to predict .some disease do not have visibility throughout early stage it solely seem that finish. The prediction of plant sickness by eye is employed in apply however results area unit subjective and disease extent isn't exactly measured. Nowadays automatic detection of plant sickness is a very important analysis topic and therefore mechanically detects the diseases from the symptoms that seem on the plant leaves. Plant disease is one among the necessary issue that causes important reduction within the quality and amount of plant production. Detection and classification of plant diseases are necessary task to increase plant productivity and economic process. Detection and classification are one of the attention-grabbing topics and far a lot of mentioned in engineering and IT fields.

P. R. Rothe and R. V. Kshirsagar [1] proposed associate degree Active Contour model (Snake segmentation) technique for segmenting the pathological region from the cotton leaf. Hu's moments [26] are used as the options for the classification. For training and classification, it uses a set of seven moments and Back Propagation Neural network has been used for classification with an accuracy of eighty five.52%. Back- propagation neural networks are extremely economical for finding Multiple category issues. Its weight is updated using Levenberg Marquardt optimisation. The proposed ways will be applied to alternative crops like orange, citrus, wheat, corn and maize etc. AakankshaRastogi, RitikaArora and Shanu Sharma [2] suggested a Fuzzy system for leaf illness detection and grading. K-means clustering technique has been used for segmentation. which teams similar pixels of associate degree image. RGB color area is reborn to L*a*b space, where L is the luminance and a*b square measure the colour area. The reason for this conversion is that luminosity issue isn't necessary for the colour image. GLCM [27] matrix including distinction, correlation, energy and homogeneity has been measured for disease grading. Artificial Neural Networks as been used for training the knowledge. Fuzzy logic is employed for grading the disease. SmitaNakwadi and NiketAmoda[3] recommended a k-means bunch technique for segmentation. RGB has been converted to HIS, where H is the hue, I indicate the intensity and S indicate the saturation value. Color Co-occurrence methodology or CCM method has been used for color feature extraction. Plant disease is detected victimization bar graph matching. The Threshold value for the picture element is computed victimization Otsu's methodology.

S. S. Sannakki and V. S. Rajpurohit [4] suggested a Back-propagation Neural Network based mostly classifier (BPNN) for sleuthing the malady in Pomegranate leaf. Features have been designated as color and texture. BPNN detects and classifies the diseases with a precision of around ninety seven.30 %. Dr. K. Thangadurai and K. Padmavathi [5] recommended pc vision image improvement for leaf malady identification. It includes color conversion and Histogram deed. Histogram deed will increase the image clarity. RGB to Grayscale conversion is used to retain the luminance data

instead of Hue and Saturation data. For encoding of linear intensity values, Gamma expansions are used. Cumulative Gaussian distribution operate distributes the intensity worth of the image. Histogram deed provides the higher quality image in Grayscale. YuanTian, Chunjiang Zhao, ShenglianLuANdXinyuGuo[6] proposed an SVM-based Multiple Classifier System (MCS)[25] for wheat leaf diseases. It uses a stacked generalization structure to join the classification choices obtained from 3 sorts of support vector machines (SVM) based mostly classifiers. The features like color, texture and shape options area unit used as coaching sets for classifiers. Firstly, features area unit classified mistreatment a classifier in low-level of MCS to corresponding mid-level classes, which will part notice the symptom of crop diseases in line with the information of plant pathology. Then the mid-level features area unit generated from these mid-categories generated from low-level classifiers. Finally, high-level SVM has been trained and correct errors made by the colour, texture and shape SVM [24] to improve the performance of detection. Compared with other classifiers, it can offer higher success rate of detection. The classifiers like SVM Artificial Neural Network classifier, k-nearest neighbor (k-NN) classifier's, the MCS can acquire higher recognition accuracy than others classifiers. Color, texture and shape SVMs [24] to improve the performance of detection. Compared with other classifiers, it can offer higher success rate of detection. The classifiers like SVM, Artificial Neural Network classifier, k-nearest neighbor (k-NN) classifier's, the MCS can acquire higher recognition accuracy than others classifiers.

NeetuChahal and Anuradha[7] proposed a C means that bunch Approach for the identification leaf sickness. Neural Networks are used for the classification. GodliverOwomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa [9] recommended a Machine learning system that converts RGB to HSV, RGB to L*a*b*. The Shape was conjointly taken as property gap [11] that square measure accustomed calculate all the parts in every threshold image. Shape options elect as space of minimum introduction parallelogram, elongation, small compactness, small perimeter and Moment of Inertia. It uses seven classifiers such as Nearest Neighbors [12], Decision tree [13], [14], Random forest [15], [16], Extremely irregular Trees [17], Naive Bayes [18] and support vector classifier (Linear SVM and RBF SVM) [19], [20], [21], [22].The splitting of dataset into coaching and testing was done by the k-fold cross-validation or known as rotation estimation methodology. The dataset was randomly split into reciprocally exclusive subsets (folds) of equal size of ten [23].Comparing the seven classifiers, Extremely irregular Trees yield a terribly high score. RatihKartikaDewi and R. V. HariGinardi[9] recommended associate image pattern classification techniques to spot the rust sickness in sugarcane leaf. Features have been select as color and Texture. Shape options embody solidity, extent, minor axis length and eccentricity. Texture features square measure extracted from distinction, correlation, energy and homogeneity. It converts RGB to science lab as a result of LAB color channel is consistent in terms of human perception.SVM classifiers are used for the classification of coaching sample in support vector machine is dissociable by a hyperplane. Eccentricity is used to work out whether the sickness form could be a circle or line phase. It is the ratio of the gap between axis length and therefore the foci. An conic whose eccentricity is zero will represent as a circle; but, an conic whose eccentricity is one will represent a line phase. Minor axis length is employed to work out the length of the axis of the diseased region. Extent is {the space|theworld|the realm} of the diseased region that's divided by the area of the bounding box. The Extent can be computed because the space divided by the world of the bounding box. Solidity is used to work out the world of the diseased region Divided by the pixels within the planoconvex hull. From Gray-Level Co-occurrence Matrix (GLCM) the texture features [10] will be extracted. Texture features square measure distinction, correlation, energy, and homogeneity. The Otsu-threshold is used to extract the form feature. Mean,skewness and kurtosis square measure used to represent color as options. For this, it transforms RGB to LAB. SVM is the binary classifier because it will classify accurately [6]. There are completely different varieties of kernel perform in SVM classifier.

The main objective of the paper is to identify the diseases in the plants. Plant pathology is the scientific study of plant diseases caused by pathogens and environmental conditions. It includes pathogen identification, diseases cycles, economic impact, management of plant diseases, etc. In existing, to detect the diseases they used the spectroscopic techniques. These techniques are very expensive and can only be utilized by trained persons only. This project presents the detection of diseases which are detected using CNN (Convolution Neural Network) technique. First the sample leaf image is given as input. Then, color channels are separated from the leaf image from these the green pixels are masked from the original image. The masking is done to avoid the processing of the green area of the leaves, since, it is healthy. By removing the green area from the original area the remaining infected area is calculated. Then the features are extracted from the affected area. Finally these features are given to the CNN to classify the disease. After finding the disease the solution for those diseases are send the corresponding user mobile using GSM device.

The remainder of the paper is organized as follows: In Section II, the overview of proposed method is presented. In Section III, the proposed method is specifically depicted, including its design idea and practical implementation approach. In Section IV, the performance of the proposed method is evaluated. Finally, conclusions are made in Section V.

II. CROP DISEASE DETECTION FROM LEAF IMAGES

The work flow of proposed method is shown in Fig.1.In the first module, the disease affected part of the input image is segmented. In the second module, the texture features are

extracted from the segmented part of the input image. In the last module, the classification approach is used for finding the affected disease in the input image. The further details of these modules are discussed below:

CROP DISEASE DETECTION FROM LEAF IMAGES PROCEDURE

The proposed method has three modules. They are

1. Affected Part Segmentation
2. Feature Extraction
3. Classification

A. Affected Part Segmentation

This is the first module of this project. In this module from the given input image the disease affected part is segmented. Then only find the type of disease and its severity can also be easily measured. To segment the affected part the FCM Clustering Algorithm is used.

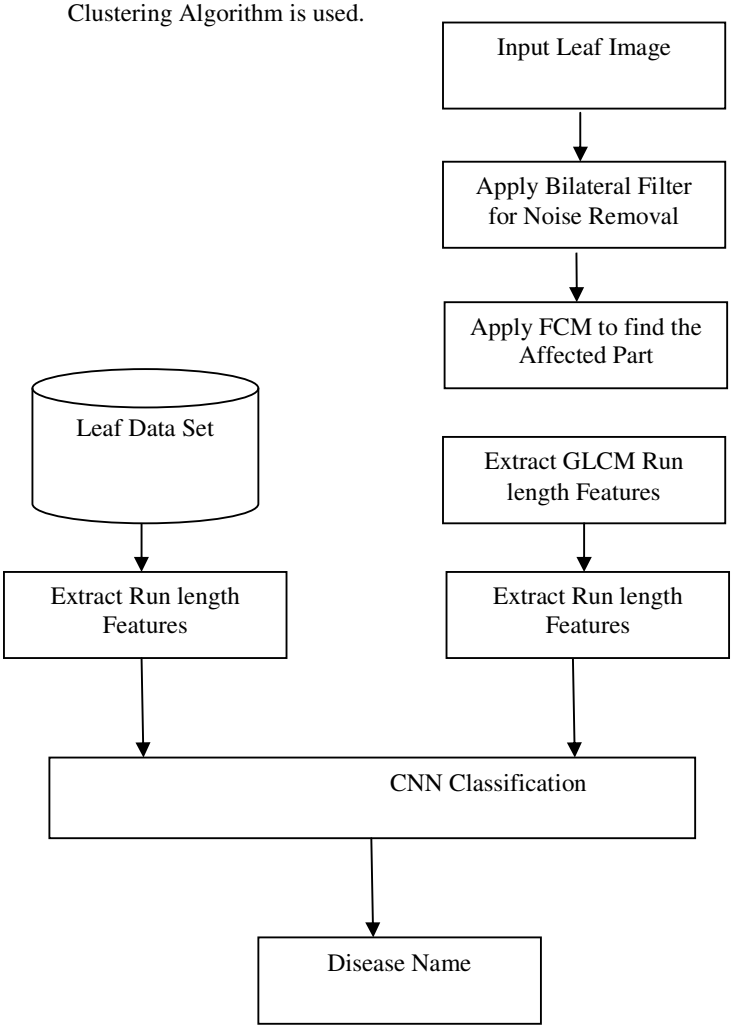


Fig. 1. Overall Block Diagram of Proposed Methods

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data is partitioned into two or more clusters. This method is developed byDunn in 1973and improved by Bezdek in 1981. This method is frequently used in pattern recognition. It is based on minimization of the objective function: The algorithm proceeds as follows:

1. Initialize Z=[zij] matrix, Z(0)
2. At k-step: calculate the centers vectors B(k)=[bj] with Z(k),

$$B_j = \frac{\sum_{i-1}^{n} v_{ij}^{m} \quad y_i}{\sum_{i-1}^{n} v_{ij}^{m}}$$

3. Update Z(k) , Z(k+1),

$$Z_{ij} = \frac{1}{\sum_{k-1}^{c} \left(\frac{\|x_i - \ b_j\|}{\|x_i - \ b_k\|} \right)^{\frac{2}{m-1}}}$$

4. If || Z(k+1) - Z(k)||< then STOP; otherwise return to step 2.

where Z is the membership function and B is the cluster center and k is the iteration number.

B. Feature Extraction

After segmenting process the next step is to calculate the features from the part. The features are used to uniquely identify the disease name and its severity. To extract the features in this project uses Gray Level Co-Occurance Matrix (GLCM) and Runlength Matrix (RLC) are used.

A GLCM is a histogram of co-occurring grayscale values at a given offset over an image. The GLCM is created from a gray-scale image. „The GLCM is calculates how often a pixel with gray with gray-level (grayscale intensity or level (grayscale intensity or Tone) value i occurs either horizontally, occurs either horizontally, vertically, or diagonally to adjacent pixels vertically, or diagonally to adjacent pixels with the value with the value j.

Run-length encoding is used to represent the strings of symbols in an image matrix. For a given image a gray level run is defined as a set of consecutive, collinear pixels having the same gray level. Length of the run is the number of pixels in the run. Run-length encoding (RLE) is an extremely straightforward type of information pressure in which keeps running of information (that is, groupings in which similar information esteem happens in numerous back to back information components) are put away as a solitary information esteem and check, as opposed to as the first run.This is most useful on data that contains many such runs. Consider, an example for, simple graphic images such as icons, line drawings, and animations. It is not useful with files which don't have many runs as it could greatly increase the file size. RLE may likewise be utilized to allude to an early illustrations document organize bolstered by CompuServe for compacting high contrast pictures, however was generally supplanted by their later Graphics Interchange Format. RLE additionally alludes to a little-utilized picture organize in

Windows 3.x, with the extensional, which is a Run Length Encoded Bitmap, used to pack the Windows 3.x start up screen. Ordinary uses of this encoding are the point at which the source data involves long substrings of a similar character or paired digit.

C. Classification

The final process is to classify the disease name and its severity. To do this processes the Convolution Neural Network. CNN is one of the classification technique. It has two stages, training and testing phase.

In machine taking in, a convolutional neural system (CNN, or ConvNet) is a class of profound, bolster forward fake neural systems that has effectively been connected to breaking down visual symbolism. CNNs use a difference of multilayer perceptrons designed to require minimal preprocessing.They are otherwise called move invariant or space invariant fake neural systems (SIANN), in view of their common weights design and interpretation invariance qualities . Individual cortical neurons respond to stimuli only in a constrained region of the visual field called as the receptive field. The receptive fields of dissimilar neurons partially overlap such that they cover the entire visual field. CNNs use relatively little pre-processing compared to other image classification algorithms.This implies the system takes in the channels that in conventional calculations were hand-designed. This freedom from earlier information and human exertion in include configuration is a noteworthy preferred standpoint.

III. PERFORMANCE ANALYSIS

A. Experimental Images

In this project the images taken from the real cameras are used. The sample images are shown in the below Fig.4.1 The image size is 512 x 512 color images. All of these images are images which are affected by any one of the disease. These images are used for experimental purposes.

B. Performance Analysis

To evaluate the performance of the crop disease detection techniques several performance metrics are available. This project uses the Classification Accuracy, Precision Rate, Recall Rate and Error Rate to analyses the performance.

Classification Accuracy

Accuracy is the measurement system, which measure the degree of closeness of measurement between the original disease and the detected disease.

Accuracy= (TP+TN)/(TP+FP+TN+FN)

where, TP – True Positive (equivalent with hit)
FN – False Negative (equivalent with miss)
TN – True Negative (equivalent with correct rejection)
FP – False Positive (equivalent with false alarm)

Error Rate

Error Rate is the measurement system, which measure no of falsely identified diseases name form the given input images.

$$Error\ Rate = \frac{Number\ of\ images\ of\ falsely\ identified\ diseases}{Total\ number\ of\ images}$$

Precision Rate

The precision is the fraction of retrieved instances that are relevant to the find.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Where TP = True Positive (Equivalent with Hits)
FP = False Positive (Equivalent with False Alarm)

Recall Rate

The recall is the fraction of relevant instances that are retrieved according to the query.

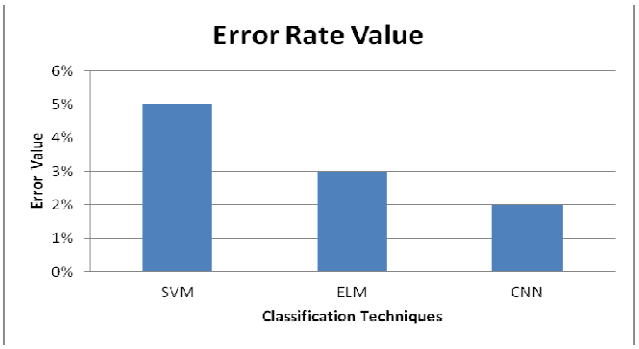
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

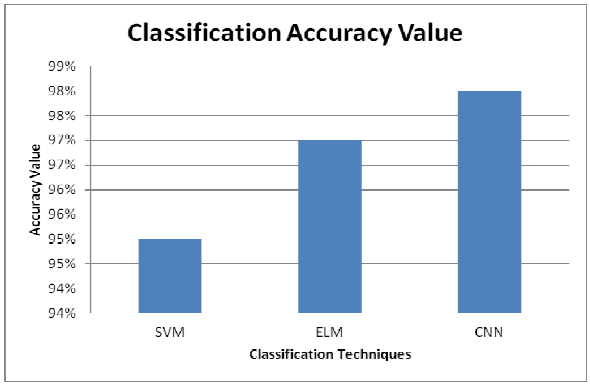
Where TP = True Positive (Equivalent with Hits)
FP = False Negative (Equivalent with Miss)

Fig. 5. Experimental Images

Crop Disease Detection Methods	Classification Accuracy Value
SVM	95%
ELM	97%
CNN	98%

Table 5.1 Classification Accuracy value



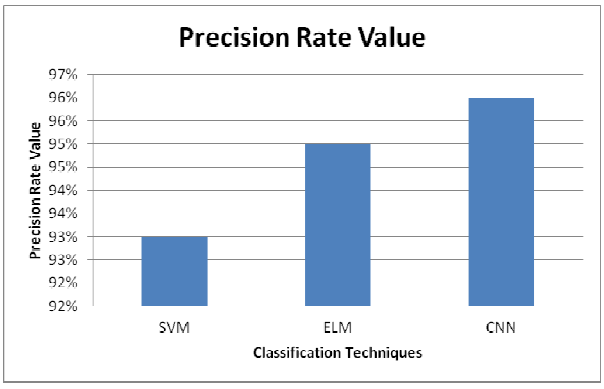


Crop Disease Detection Methods	Precision Rate Value
SVM	93%
ELM	95%
CNN	96%

Table 5.3 Precision Rate Value

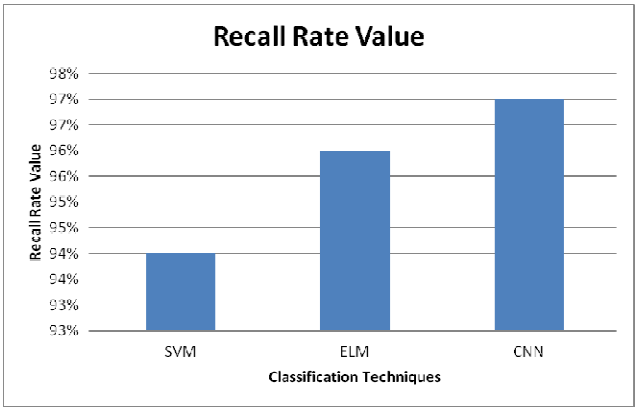
Crop Disease Detection Methods	Error Rate Value
SVM	5%
ELM	3%
CNN	2%

Table 5.2.Error Rate Value



Crop Disease Detection Methods	Recall Rate Value
SVM	94%
ELM	96%
CNN	97%

Table 5.4 Recall Rate Value



IV. CONCLUSION

In the paper the diseases in the plants are detected. Plant pathology is the scientific study of plant diseases caused by pathogens and environmental conditions. It includes pathogen identification, diseases cycles, economic impact, management of plant diseases, etc. In existing, to detect the diseases they used the spectroscopic techniques. These techniques are very expensive and can only be utilized by trained persons only. This project presents the detection of diseases which are detected using CNN (Convolution Neural Network)technique. First the sample leaf image is given as input. Then, color channels are separated from the leaf image from these the green pixels are masked from the original image. The masking is done to avoid the processing of the green area of the leaves, since, it is healthy. By removing the green area from the original area the remaining infected area is calculated. Then the features are extracted from the affected area. Finally these features are given to the CNN to classify the disease. After finding the disease the solution for those diseases are send the corresponding user mobile using GSM device.

References

[1] P. R. Rothe and R. V. Kshirsagar," Cotton Leaf Disease Identification using Pattern Recognition Techniques", International Conference on Pervasive Computing (ICPC),2015.

[2] AakankshaRastogi, RitikaArora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.

[3] SmitaNaikwadi, NiketAmoda," Advances In Image Processing For Detection Of Plant Diseases," International Journal of Application or Innovation in Engineering & Management (IJAIEM), Vol2, Issue 11, November 2013.

[4] S. S. Sannakki and V. S. Rajpurohit,"Classification of Pomegranate Diseases Based on Back Propagation Neural Network," International Research Journal of Engineering and Technology (IRJET), Vol2 Issue: 02 | May-2015.

[5] Dr.K.ThangaduraiandK.Padmavathi," Computer Visionimage Enhancement For Plant Leaves Disease Detection," World Congress on Computing and Communication Technologies2014.

[6] Yuan Tian,Chunjiang Zhao, Shenglian Lu and XinyuGuo," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Proceedings of 2010 Conference on Dependable Computing (CDC'2010),November 20-22, 2010.

[7] NeetuChahal and Anuradha," A Clustering Adaptive Neural Network Approach for Leaf Disease Identification," International Journal of Computer Applications (0975 – 8887)Vol 120 – No.11, June 2015.

[8] GodliverOwomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa," Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease

and Black Sigatoka Disease ",Proceding of the 1'st interational conference on the use of mobile ICT in Africa ,2014.

[9] RatihKartikaDewi and R. V. HariGinardi," Feature Extraction for Identification of Sugarcane Rust Disease", International Conference on Information, Communication Technology and System, 2014.

[10] Bernardes, J. G. Rogeri, N. Marranghello, and A. S Pereira,"Identification of foliar diseases in cotton crop," Topics in Medical Image Processing and Computational Vision, vol. 8, pp. pp 67–85, 2013.

[11] Ronse, "Set-theoretical algebraic approaches to connectivity in continuous or digital spaces," Journal of Mathematical Imaging and Vision, 1998.

[12] Z. Ma and K. Ata, "K-nearest-neighbours with a novel similarity measure for intrusion detection," UKCI'13, pp. 266–271, 2013.

[13] O. Maimon and L. Rokach, "Data mining and knowledge discovery handbook, second edition," April 2010.

[14] L. Ruey-Hsia and B. Geneva G, "Instability of decision tree classification algorithms," Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 570–575, 2002.

[15] B. S. Leo and L. Breiman, "Random forests," Machine Learning, pp. 5– 32, 2001.

[16] L. Andy and W. Matthew, "Classification and regression by randomforest," R News, vol. 2, no. 3, pp. 18–22, 2002.

[17] H. Zhang, "The optimality of naive bayes," Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004), 2004.

[18] S. R. Gunn, "Support vector machines for classification and regression," University of Southampton, Technical Report, 1998.

[19] S. Keerthi, O. Chapelle, and D. DeCoste, "Building support vector machines with reduced classifier complexity," Journal of

Machine Learning Research, vol. 7, pp. 1493–1515, 2006.

[20] wei Hsu, C. chung Chang, and C. jen Lin, "A practical guide to support vector classification," National Taiwan University, Taipei 106, Taiwan, 2010.

[21] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines: And other kernel-based learning methods," Cambridge University Press, 2000.

[22] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2, pp. 1137–1143, 1995. [Online]. Available:

[23]
http://dl.acm.org/citation.cfm?id=1643031.1643047
.

[24] Tian Y W, Wang B, Tang X M. Recognition of maize disease based on texture feature and support vector machine. Journal of Shenyang Agricultural University, 2005, 36(6): 730-732. (in Chinese)

[25] Roli, F. Andgiacinto,G. 2002. Design of multiple classifier systems. In Hybrid Methods in Pattern Recognition, H. Bunke, and A. Kandel, eds. World Scientific, River Edge, NJ.

[26] Dr. K.Sri Rama Krishna ,B.Vanajakshi "Classification Of Boundary And Region Shapes Using Hu-Moment Invariants ",B.Vanajakshi et al / Indian Journal of Computer Science and Engineering (IJCSE),Vol. 3 No. 2 Apr-May 2012.

[27] MiroslavBenco, Robert Hudec, PatrikKamencay, Martina Zachariasova and SlavomirMatuska,"An Advanced Approach to Extraction of Colour Texture Features Based on GLCM",International Journal of Advanced Robotic Systems, 14 May 2014