IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Comparative Analysis on Segmentation Approaches for Plant Leaf Disease Detection

S.Vijayalakshmi[1*], D.Murugan[2]

[1*] Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelvlei, Tamilnadu, India
[2] Department of Computer Science and Engineering, Manonmaniam Sundaranar University,Tirunelvlei,Tamilnadu,India

[*]Corresponding Author: sviji2005@gmail.com

Available online at: www.ijcseonline.org

Accepted: 15/May/2018, Published: 31/May/2018

***Abstract*-** Plant pathology is the scientific analysis of plant diseases caused by pathogens and different environmental conditions. The leaf is one of the significant plant parts which highlight the presence of diseases. Existing methods use spectroscopic techniques to detect the diseases present in plants. These techniques are very expensive and can only be utilized by trained persons only. The method mentioned in this paper is an easy and cost-effective way which utilizes the leaf image of the plant. This input image is subjected to segmentation of disease part, feature extraction and classification in order to identify the disease. The main objective of this paper is to compare the clustering approaches FCM, Artificial Bee Colony and K-Means which are useful in disease part segmentation and to identify the best approach yields accurate results for identifying the plant disease. This work utilizes the GLCM, Run Length, Color Moment and Color Histogram features for feature extraction. Once these features are extracted from the segmented disease part, the disease present in the leaf is identified using the KNN (K-Nearest Neighbor) technique. The experimental result shows that the Artificial Bee Colony approach segments the diseased part of the leaf in a better way than the other two approaches.

Keywords — Leaf disease, FCM, ABC, k-means clustering, GLCM, Support Vector Machine, K-Nearest Neighbor Approach

I. INTRODUCTION

The backbone of Indian economy is agriculture. Most of the population present in India mainly depends on agriculture. There are many research problems in agriculture, which are mainly aimed to increase the productivity and quality at reduced expenditure and with increased profit. Plant leaf disease detection is a research area which identifies the disease present in a plant by analyzing its leaf. The successful plant cultivation of a farm mainly depends on monitoring health and disease in a plant. The endemic of plant diseases could cause several losses and affect agricultural production and food security.

Most of the plant disease is on their leaves and on the stem of the plant. Mainly the diseases are classified into viral, bacterial, fungal, diseases due to insects, rust etc. on the plant. The prediction of plant disease by the naked eye is used in practice, but the results are subjective and disease extent is not precisely measured. Moreover, visual identification is labor intensive, less accurate and can be done only in small areas. The important task for farmers is to find out these diseases as early as possible. But the younger generation is not having enough knowledge about agriculture. Even though the government is having many supporting centers for the awareness of farmers, direct inspection of each and every farmer's land is impossible. But without directly seeing the crops one cannot confirm the disease present in the plant. The automatic detection techniques can help in finding the plant diseases in an early and efficient way. Nowadays image processing techniques and neural networks are used to detect the type of diseases automatically. This enables machine vision that is to provide image-based automatic inspection, process control, and robot guidance.

The paper is organized into the following sections: Section II includes a discussion of related work and Section III gives detailed inside view on segmentation techniques taken to comparing the plant disease classification. Section IV describes experimental results of classification performance of plant disease detection using various segmentation techniques and Section V presents the conclusion achieved from the comparative analysis.

II. Related Work

Agricultural research is a major field which includes various research problems. Plant disease detection is one such research which attains farmer's interest as it can result in increased productivity. Various researchers achieved significant results that contribute to the betterment of existing systems. In this section, various techniques for image processing, for plant disease detection are discussed.

Recognition and segmentation are techniques to extract appropriate features from the plant leaf. To achieve the results that, identification and the outcomes from the feature extraction are classified. The authors have proposed the marker-controlled watershed segmentation algorithm for leaf extraction and disease diagnosis [1]. On the basis of work, this method found various types of HSI images separately. Finally, the results sometimes failed in leaf extraction when solidity measure is used to select the best leaf extraction.

Aakanksha et.al [2] proposed a fuzzy system for leaf disease detection. K-means clustering is used has been used for the segmentation. RGB color is converted into L*a*b color space, where L is the Luminance and a*b square measure the color area. This conversion is that luminosity issue isn't necessary for the color image.

The author has proposed a method to identify the disease on plant leaves using Image Processing and clustering techniques. RGB color is converted into L*a*b color space for image preprocessing and then hierarchical clustering is used for the segmentation. GLCM method used for the feature extraction. These features are given as input into the SVM classifier [3] to identify the leaf disease. The recognition accuracy received 92% from the comparison of three segmentation techniques.

Dubey et.al [4] has presented a novel approach for detecting defects apples based on color features. Apple scab, apple rot, and apple blotch diseases are considered for experimentation by a k-means clustering technique. Evaluating the quality of defects apples was observed by the proposed method. The experimental results reveal that the proposed approach is effective in terms of precision and computational efficiency.

Lange et.al [5] has proposed an approach for detecting maize crop diseases in image processing system as well as to detect the color features from the diseased leaf. The findings were got that some leaf disease symptoms, namely brown stripe downy mildew and stem borer found on maize crop. Finally, classification is done using an ANN classifier to detect the disease type.

Another recent approach for a specific type of recognized disease spots present on the leaves of sugarcane [6]. To identify the disease leaves which are spots present based on several features, colors, ratio, eccentricity, circularity, and moment's analysis have been tested. This system could be given a classification accuracy of 95.25% using minimum classifier.

The author suggested methods to identify the fungal disease on cereal leaves. The disease affected regions are segmented using the K - means technique. A color Co-occurrence matrix is used for feature extraction, to extract the affected parts of leaves. These features are used as inputs to SVM and ANN Classifier [7]. The outcome of the classification accuracy 87% using ANN Classifier and 91.16% using the SVM classifier. An SVM classifier originates to be better than ANN classifier based on this 2 classifier results.

Al-Bashish et.al. [8] Developed a fast and accurate method of identifying and classifying leaf disorders by K-means segmentation and the classification is done by the neural network. The automatic classification of leaf disorders is based on high-resolution multispectral and stereo images [9]. This approach uses sugar beet leaves.

The clustering algorithm highly opts for a lot of research problems; but still, it suffers from several drawbacks due to the choice of initializing. Giuliano Armano and Mohammed Reza Germany have explicated clustering analysis, which is based on a combination of artificial bee colony algorithm and k-means technique [10]. The result reveals that ABCk algorithm will definitely converge to an optimal solution in almost all experimental.

Feature extraction is one of the important works in image processing. In this paper [11], P. Mohanaiah et.al has proposed image texture feature extraction using GLCM approach to detect the features. Identify the features highly depends on the application. An application of gray level co-occurrence matrix (GLCM) has been working with many experimental based on Entropy, Inverse Difference Moment, Angular Second Moment, and Correlation.

This work compares various segmentation techniques for plant disease in order to identify which segmentation technique provides better classification accuracy in the diseased leaf images. This work compares the segmentation techniques, namely, Fuzzy C-means, K-means clustering and Artificial Bee Colony algorithm.

III. METHODOLOGY

CROP DISEASE DETECTION FROM LEAF IMAGES

The overall architecture of the crop disease detection is shown in Figure.1. It has three modules, namely Affected Part Segmentation, Feature Extraction, Feature Extraction, and Classification. Initially, the input image is segmented into disease and normal leaf part. 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.

3.1 Affected Part Segmentation

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.

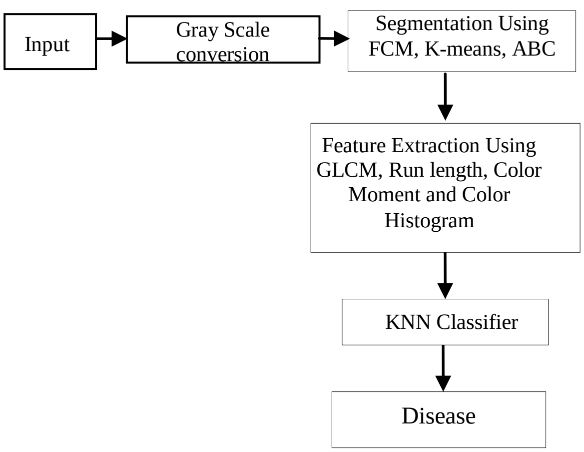

Figure 1. Overall Block Diagram of Proposed Methods

To segment, the affected part the FCM Clustering, Artificial Bee Colony and K-Means Clustering approach are used. FCM clustering is an iterative, data-partitioning algorithm that assigns and observations to exactly one of the k clusters defined by centroids, where k is chosen before the algorithm starts.

3.1.1 Fuzzy c-means Algorithm

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the objective function: The algorithm proceeds as follows:

1. Initialize $U = [U_{ij}]$ Matrix, U (0)
2. At k-step: Calculate the center's vectors $C(k) = [c_j]$ with U(k)

$$c_j = \frac{\sum_{i-1}^{n} U_{ij}^{m} x_i}{\sum_{i=1}^{n} U_{ij}^{m}}$$

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

$$U_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|X_i - C_j\|}{\|X_i - C_k\|}\right)^{\frac{2}{m-1}}}$$

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

Where U is the membership function and C is the cluster center and k is the iteration number

From the segmented infected part of the leaf image texture, shape and color features are extracted using the techniques GLCM, run length, color moment and color histogram techniques respectively.

3.1.2. K- Means Clustering Algorithm

1. Place K points in the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroids.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated. It is calculated by using the below formula

$$J = \sum_{J=1}^{K}\sum_{I=1}^{X}\left\|X_i^{(j)} - C_j\right\|^2$$

Where $\left\|x_i^{(j)} - c_j\right\|^2$ Is a chosen distance measure between a data point $x_i^{(j)?}$ And the cluster center c_j Is an indicator of the distance of the *n* data points from their respective cluster centers.

3.1.3 Artificial Bee Colony Algorithm

1. Initialize the bee colony $X = x_i$ {I = 1,2, ... , n}, where n denotes the population size, x_i Is the it bee.
2. According to the fitness function, calculate the fitness few of each employed bee x_i , and record the maximum nectar amount as well as the corresponding food source.
3. Each employed bee produces a new solution view in the neighborhood of the solution in its memory by $V_{i=}X_i + (X_i - X_k) \times \Phi$, where k is an integer near to me,k ≠ i, and Φ is a random real number in [−1, 1].
4. Use the greedy criterion to updateX_i. Compute the fitness of. If V_i is superior toX_i, X_i is replaced with; $V_i$otherwise X_i remained.
5. According to the fitness I of xi, get the probability value Pi via formulas (1) and (2).

$$P = \frac{fit_i}{\sum_{i=1}^{n} fit_i} \quad \text{-----} \quad (1)$$

$$fit_i = \begin{cases} \frac{1}{1+f_i}, & if\ f_i \geq 0 \\ 1 + abs(f_i) & if\ f_i < 0 \end{cases} \quad \text{------} \ (2)$$

6. Depending on the probability Pi, onlookers choose food sources, search the neighborhood to generate candidate solutions, and calculate their fitness.
7. Use the greedy criteria to update the food sources.
8. Memorize the best food source and nectar amount achieved.
9. Check whether there are some abandoned solutions or not. If true, replace them with some new randomly-generated solutions by x_i = min + (max−min) × (, where (is a random real number in [0, 1], min and max stand for lower and upper bounds of possible solutions respectively.
10. Repeat steps (3) –(9), until the maximum number of iterations (Kmax) is reached or stop conditions are satisfied.

3.2. Feature Extraction

After segmenting process the next step is to calculate the features of the part. The features are used to uniquely identify the disease name and its severity. To extract the features in this paper uses Gray Level Co-Occurrence Matrix (GLCM) and Run length Matrix (RLC) is used. A GLCM is a histogram of co-occurring grayscale values at a given offset over an image. As far as the leaf of the plant is considered, the significant features can be obtained by 1. The color of the leaf 2. The texture of the leaf 3. The shape of the leaf.

The GLCM is a one of the feature extraction technique which extracts different combinations of pixel brightness values (gray levels) occur in a leaf disease image. After changing the color image to gray-level image by using the GLCM texture feature extraction we extracted the 5 features such as energy, entropy, correlation, skewness and kurtosis of each leaf image. Our initial assumption in characterizing image texture in leaf image is that all the texture information is contained in the gray-level Co-occurrence matrices. Hence all the textural features here are extracted from these gray-level Co-occurrence matrices.

Color is one of the most widely used features. Color features can be obtained by various methods like color histogram, Color moment. The Color moment method has the lowest feature vector dimension and lowers computational complexity. Hence it can be considered as a suitable parameter to generate feature vectors which can be further used for the leaf disease classification purpose. The color of the diseased leaf is used to generate the features of color histogram. The Color features extractions are applied to samples that are a healthy leaf of the plant and the diseased leaf of the plant. Plant diseased is detected by using histogram matching. The histogram matching is based on the color feature.

A color histogram of an image represents the distribution of the composition of colors in the image. The histogram is generated for both healthy leaf sample and diseased leaf sample and saves in the memory; these histograms are displayed, when we generate the histogram for the testing image. In the second phase, after the training process, the histogram of the testing sample is created or generated suddenly. Once the histograms are generated from both samples and the testing image, immediately we will apply the comparison technique based on the histogram. The comparison is first with the testing sample and the healthy sample if the testing sample is diseased, it compares the testing sample with the diseased sample and these steps take a few minutes to display the comparison result that is the testing sample is diseased or not. A color histogram is the most commonly used color features in image retrieval systems. However, this feature cannot effectively characterize an image, since it only captures the global properties. To make the retrieval more accurate, this paper introduces a run length feature. The feature integrates the information of color and shape of the objects in an image. It can effectively discriminate the directions, areas and geometrical shapes of the objects.

3.3. Classification

The final process is to classify the disease name and its severity. To do this process the K Nearest Neighbor Classifier is used. In K-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

Crop disease detection methods	Classification Accuracy Value	Error Rate Value	Precision Rate Value	Recall Rate Value
K Means	89%	11%	90%	88%
FCM	91%	9%	92%	90%
ABC	93%	7%	94%	92%

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is more frequent among the k training samples nearest to that query point. A commonly used distance metric for continuous variables is the Euclidean distance. The Euclidean distance between points p and q is calculated by using the below formula.

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

IV. PERFORMANCE ANALYSIS

4.1 Experimental Images

In this paper, the images taken from the real cameras are used. In total 200 images are taken. From that 100 images are used for training and remaining 100 images are used for testing. The sample images are shown in the below Fig.2 The image size is 512 x 512 color images. All of these images are images which are affected by any one of the diseases. These images are used for experimental purposes.

4.2 Performance Analysis

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

The following table shows the classification performance obtained using the various segmentation techniques. As a first step of the comparison, classification is done by segmenting the diseased leaf part is done with K-means clustering technique. After segmenting the diseased leaf part features are extracted using GLCM, Run length, Color moment and Color Histogram. Extracted features are given as input to a K - NN classifier to detect the disease. It achieves a classification accuracy of 89%. From the same methodology is applied to FCM and ABC segmentation techniques. Classification with FCM and ABC segmentation achieves a classification accuracy of 91% and 93% respectively. Performance metrics like Error rate, Precision rate, and recall rate also evaluated and Tabulated below.

Table: Comparative results on various segmentation approaches

From the table, it is clear that classification with ABC segmentation achieves better classification accuracy with reduced error rate.

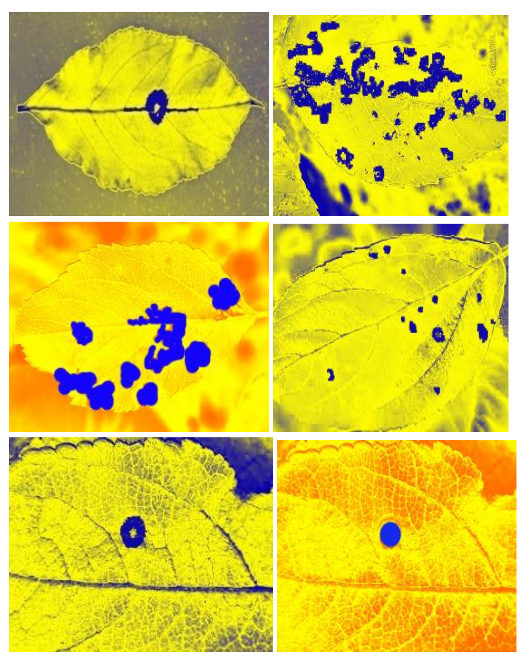

Figure 2 Experimental Images

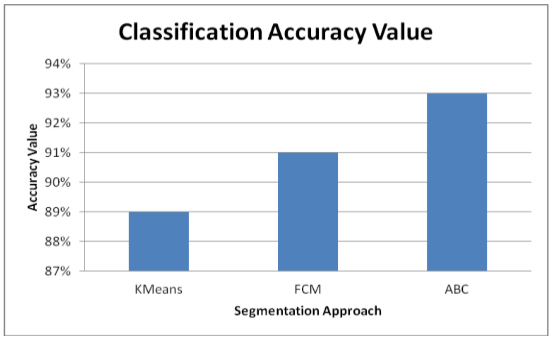

Figure 3. Experimental results in Classification Accuracy value

This paper analysis the classification accuracy value of the K-Means, FCM and ABC. The results are captured and showed in Fig. 3. From the above graph it is shown that the detection accuracy value of the ABC method is higher than the other existing approaches. So the ABC method is better than the existing approaches.

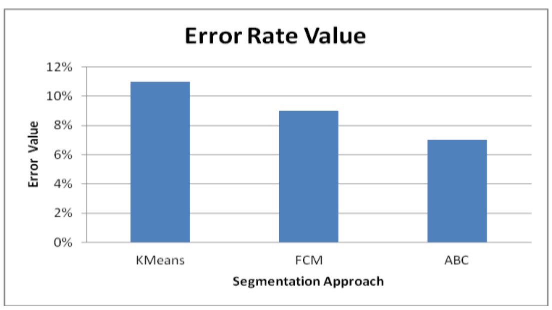

Figure 4. Experimental results in Error Rate value

This paper analysis the error rate value of the K-Means, FCM and ABC. The results are captured and showed in Fig. 4. From the above graph it is shown that the error rate value of the ABC method is lower than the other existing approaches. So the ABC method is better than the existing approaches.

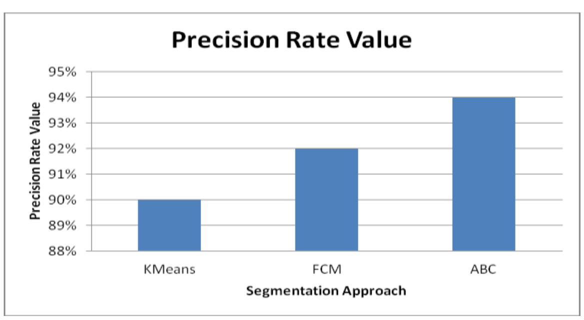

Figure 5. Experimental results in Precision Rate value

This paper analysis the precision rate value of the K-Means, FCM and ABC. The results are captured and showed in Fig. 5. From the above graph it is shown that the precision rate value of the ABC method is higher than the other existing approaches. So the ABC method is better than the existing approaches.

This paper analysis the recall rate value of the K-Means, FCM and ABC. The results are captured and showed in Fig. 6.

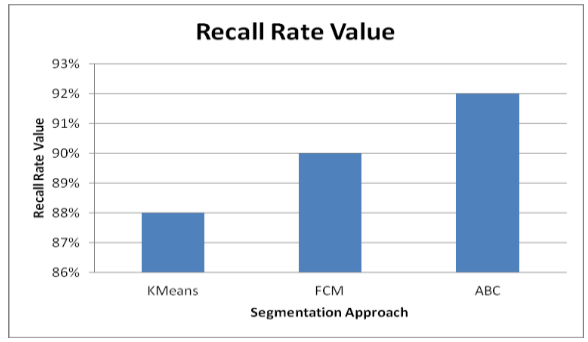

Figure 6. Experimental results in Recall Rate value

From the above graph it is shown that the recall rate value of the ABC method is higher than the other existing approaches. So the ABC method is better than the existing approaches.

V. CONCLUSION

In this paper, the brief comparative analysis is made between the three different clustering techniques to identify the diseases in the plant leaves. We have taken clustering techniques such as KNN, FCM clustering Approach, ABC algorithm respectively to segment the affected portion of the leaves. Here we have used the strong vector features such as GLCM, Run Length, Color Moment and Color Histogram respectively as the input for the classifiers. The performance factors such as accuracy, precision rate, recall rate and error rate are measured for these three different techniques. From the performance analysis, it is found that the ABC algorithm yields the better outcome of other two techniques for identifying the leaf diseases.

REFERENCES

[1] Shiddalingappa Kadakol, Jyothi B Maned, "Intelligence System for leaf Extraction and disease Diagnostic", International Journal of computer sciences and Engineering " Vol.4, Issue 3, May 2016, PP.62-66.

[2] Aakanksha Rastogi, Ritika Arora, and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" (SPIN) 2015, pp 500-505.

[3] N. Swetha and N. Sasirehka,"Prediction of Leaf Disease uses Segmentation with Hierarchical Clustering", International Journal of Engineering Technology, Science and Research, Vol. 3, Issue.6, June 2016, pp. 37-42.

[4] Dubey S R., Dixit P., Singh N. And Gupta, J P, "Infected Fruit Part Detection uses K-Means Clustering Segmentation Technique",

International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 2, Issue.2, 2013, pp. 65-72.
[5] Lange P S., Patil S A., Khot D S., Otari O D And Malakar U G, "Automatic Detection and Classification of Plant Disease through Image Processing", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue.7, July 2013, pp. 798-801.
[6] Sungkur R K., Baichoo S. And Poligadu A, 2013, "An Automated System to Recognize Fungi-caused Diseases on Sugarcane Leaves", Proceedings of Global Engineering, Science and Technology Conference, Singapore, 3-4 Oct. 2013, pp. 1-11.
[7] Jagadeesh D. Pujari, Rajesh Yakkundimath and Abdulmunaf S. Byadgi, " International journal of signal Processing, Image Processing and Pattern Recognition, Vol.6, Issue.6 (2013).
Pp. 321-330.
[8] Al-Bashish, D., M. Braik, and S. Bani-Ahmad. "Detection and classification of leaf diseases using K-means-based segmentation and neural networks based classification". *Information Tech. Journal*, Vol 10, Issue 2, 2011, pp 267-275.
[9] Bauer, S. D., F. Koch, W. Forstner. "The potential of automatic methods of classification to identify leaf diseases from multispectral images". *Precision Agriculture*, Vol 12, Issue 3, 2011, pp 361-377.
[10] Giuliano Armano and Mohammad Reza Farmani," Clustering Analysis with Combination of Artificial Bee Colony Algorithm and k-Means Technique", International Journal of Computer Theory And Engineering", Vol.6.No.2, April 2014, pp. 141-145.
[11] P. Mohanaiah*, P. Sathyanarayana**, L. GuruKumar," Image Texture Feature Extraction Using GLCM Approach", International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013, pp. 1-5.

AUTHOR PROFILE

S. Vijayalakshmi is active Researcher in the field of Image Processing. She has received Master of Computer Applications, from the Department of Computer Applications, Madurai Kamaraj University, Madurai. Currently, she is doing a Doctoral degree from Manonmaniam Sundaranar University, Tirunelveli, India.

D. Murugan is an active analyst and Researcher in the field of Face Recognition and Intelligent systems. He has received a Ph.D. in Computer Science & Engineering. Currently, he is working as Professor under the Department of Computer Science & Engineering, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India.