# IMAGE SEGMENTATION USING SUPPORT VECTOR MACHINE AND FUZZY C-MEANS CLUSTERING FOR ANALYZING PLANTS DISEASES

[1]D.Angayarkanni,[2] Dr.L.Jayasimman,
[1]Research Scholar, PG and Research Department of Computer Science, Bishop Heber College Trichy,Tamilnadu,India.
[2]Assistant Professor, Department of Computer Applications, Bishop Heber College, Trichy, Tamilnadu, India.

***Abstract :*** Image segmentation performs a major role in the agricultural field due to its extensive range of applications, to identify plants parts and classifying the different diseases present in plant parts. Leaf spot diseases weaken trees and shrubs by interrupting chemical change, the way by that plants generate energy that sustains defence systems and growth and influences survival. This difficulty occurrence can be solved by providing the remedial action within the time and it can be achieved through the introduction of technology in the system. Several methods are used to segment the images. Different kinds of algorithms have been proposed for segmenting the image data using clustering process. Selecting a classification method is always a hard task because the quality of the result is differ from various input data. This paper, compare and analyze two algorithms such as SVM (Support Vector Machine) and Fuzzy C-Means on the basis of their performance for the noisy Leaf images and to find which algorithmic approach presents better results in image segmentation. These segmentation algorithms are compared in terms of Segmentation Accuracy, Noise level and Time Complexity.

***IndexTerms* - Image Segmentation Techniques, Clustering Algorithm, Fuzzy C-Means Algorithm, Support Vector Machine algorithm, Edge Detection Technique, Plant Disease Management.**

## I. INTRODUCTION

In agricultural field, computer vision systems would assist to tackle the diseases problems. Development for agricultural applications using computer vision systems is recognition of food products in food processing, sorting of fruits in fruit processing, detection of seeds, detection of diseases, classification of grains, medicinal plant recognition, and so on [1]. Using these techniques, digital images are obtained using digital camera (in given domain) in which image processing techniques are applied to extract useful characteristics necessary for further analysis. The image segmentation is used to segment an image into different region such that each region is considered as homogeneous. Several image segmentation methods are available, however still none of them are suitable for all the applications [2]. Edge-based segmentation technique is one of the broadly used techniques for image segmentation [3]. Edge-based scheme, partitions or subdivide an image depends on abrupt changes in the intensity level of images. Edge detection methods consist of making decision as to check whether pixels are presented an edge or not. It includes adaptive local operations, high emphasis spatial frequency filtering, gradient operations, line, and curve flittering and functional approximations [4], [5]. These basic techniques are still utilized in new segmentation techniques. Edge detection is performed as a simple and basic method used in segmentation methods [6]. It is considered as a useful segmentation technique for simple images that have a few characteristics, like grayscale image. Therefore, this paper deal with edge based segmentation methods.In this paper, the edge based segmentation techniques are used to compare and analyze with two algorithms such as FCM and SVM algorithms. These algorithms can be used on various data sets. The performance of image data classification depends on a variety of factors around test mode, dissimilar nature of data sets and size of data set.

### 1.1 VARIOUS TYPES OF LEAF DISEASES

The challenging task in agricultural field is plant disease management. Most of the diseases are observed on the stems or leaves of the plant. Here, precise quantification of visually observed diseases, traits has not yet analyzed due to the complexity of visual patterns. Thus, it has been increasing demand for more sophisticated image pattern understanding in an efficient manner [7]. Most of the leaf diseases are affected by viruses, fungi, and bacteria. Fungi are recognized primarily from their morphology, with emphasis placed on their reproductive structures. Bacteria are more primitive than fungi and generally have simpler life cycles. Bacteria survive in single cells and increase in numbers by dividing into two cells during a process known as binary fission. Viruses are extremely tiny particles consisting of protein and genetic material with no associated protein [9]. In biological science, sometimes thousands of images are generated in a single experiment. There images can be required for further analysis like classifying lesion, scoring quantitative traits, calculating area eaten by insects, etc. Almost all of these tasks are processed manually or with distinct software packages. It not only considers tremendous amount of work but also suffers from two major issues such as excessive processing time and subjectiveness rising from different individuals. Hence to conduct high throughput experiments, efficient computer software is needed to automatically extract and analyze significant content of plants. Here, the image processing plays an important role [8]. This paper presents a comparative study in different image processing techniques used for studding leaf diseases.

### 1.2 EDGE DETECTION

Edge detection is mostly used for analyzing meaningful discontinuities in gray scale images. Images are represented by edges with rapid intensity disparities. First order derivative and second order derivative are used for the detection of edges in an image. Various categories of edges such as Ramp edge, Step edge, Roof edge and Line edge are present in an image. An ideal edge is a step function from one pixel to the next pixel. However, many factors should be taken into consideration for detecting edges such as noise, pixel clarification, brightness of the image as a whole, corners.In image processing analysis, a problem of fundamental importance is the edge detection technique. Edges characterize object boundaries and are functional utilization of segmentation, registration, and identification of objects. Edge detection techniques are a typical application using automatic character recognition. There are three fundamental steps in edge detection [9].They are smoothing an image for noise removal, edge point detection and edge linking and thinning.

Generally, edge are detected with gradient and laplacian operators.

- Gradient operator is used to detect the edges with the minimum and maximum value of the first order derivative of the image. Prewitt, Roberts, Sobel operators are used to detect both horizontal and vertical edges. Sharp edges can be divided out through suitable thresholding.
- Laplacian operator explores edges with zero crossings during the second order derivative of the image, for example Laplacian of Gaussian, MarrHildreth [10], etc.

The initiative behind edge detection is to discover places in an image in which the intensity modifies rapidly, using one of two general conditions[11] The places where the magnitude of the gradient(first order derivative) of the image is greater than a specified threshold. And the places where the laplacian(second order derivative) operator has a zero crossing.

The general matlab syntax for edge function is [g, t] = edge (f, 'method', parameters)
where f is the input image, method is one of edge detection technique and parameters are additional parameters.

## II. METHODOLOGY

Many popular edge based image segmentation mechanisms are used for plant leaf classification. Plant leaves are classified based on its different morphological aspects. There are various classification techniques used such as SVM and FCM. Plant leaf disease classification has extensive application in Agriculture.

### A. SVM (SUPPORT VECTOR MACHINE)

SVM was first initiated very effective technique for general pattern recognition, classification and regression, by Vapnik [12] [13]. SVM technique is one of the most helpful and standard techniques for data classification [14]. It is mainly used to identify plant disease horticulture crops/affecting agriculture. Due to its efficient performances and implementations proved to be excellent for high dimensional issues and small data sets. Viewing training input vector in an n-dimensional space, SVM builds a hyper-plane in the space, which can be used for classification that has the highest perpendicular distance to the closest data point of the classes (functional margin).To find the maximum margin,two parallel hyper-planes at the margin level are constructed, one on each side of the isolating hyper-plane with no data in between them.The aim is to determine which class a new data point belongs to, based on data points associated with the classes. In SVM, a data point is represented as a p-dimensional vector (a list of p numbers) and to check whether they can be separated by a (p−1) dimensional hyper-plane. This is known as maximum margin classifier or a linear classifier. A hyper plane is known as a point in a one dimensional space. In two dimensional spaces, it is a line. As the dimensions are increasing it become a hyper-plane. The main goal is to separate the hyper planes with maximum margin.

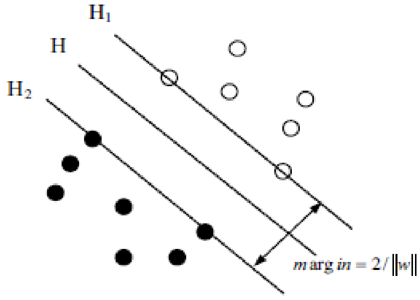


**Figure 1: SVM classifier with hyper planes**

In Figure 1 illustrates that two samples such as solid and hollow dots point, the classification line represents H. The classification line is parallel to H1 and H2 and both cross the samples nearest the classification line in two classes, and the distance between H1 and H2 is called as classification interval. Not only the optimal classification line separate the two class properly (training error is 0), but also make the class interval largest. Even if H satisfies the conditions for optimal separating hyper-plane [15], the training sample points on H1 and H2 called as support vectors.

A training sample set $\{ (x_i , y_i ), i = 1,2,...,l$, with $y_i \in \{ +1, -1\}$, which denotes two types of identified classes respectively. If $x_i \in R^n$ belong to the first category, the corresponding output is marked as positive ($y_i = +1$), and if it belongs to the second category, the corresponding output is marked as negative ($y_i = -1$). The main aim is to construct a decision line correctly that classify the test data as possible, and its original problem is described as:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

$$\text{s.t.} \quad y_i(w^T \varphi(x_i ) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, \ldots, l$$

In above formula, C means the penalty function; the bigger C denotes a greater penalty for misclassification that is the only parameters can adjust the algorithm. Using the Lagrange multiplier method to resolve quadratic programming issues in above formula with a linear constrained, its corresponding dual problem is:

$$\min_{a} \frac{1}{2} a^T Q_a - e^T a$$

$$\text{s.t. } 0 \leq a_i \leq C, i = 1, \ldots, l \qquad y^T a = 0$$

Let the unit vector represents e, the upper bound is C > 0, a $l$ x $l$ positive semi-definite matrix is Q, and $Q_{ij} = y_i y_i k (x_i, x_j)$.

**B. FUZZY C-MEANS (FCM) ALGORITHM**

The FCM algorithm is mainly used to image segmentation by partitioning the space of image into various cluster regions with similar image's pixels values. The most suitable clustering type is fuzzy clustering for leaf images segmentation. This FCM clustering technique was first initiated by Dunn and later it was developed by Bezdek [16]. For the analysis of image and data models are used by unsupervised Fuzzy clustering. Here, the vector space of a sample point divided into several sub-spaces according to a distance measure. FCM has proffered a broad domain of applications like astronomy, agricultural engineering, geology, chemistry, medical diagnosis, image analysis, target recognition and shape analysis [17]. FCM clustering is also considered as a soft clustering technique that allows partial belongingness of pixels into different clusters. This partial membership is analyzed using membership functions. The sum of all membership degrees for any given data point is equal to 1. This technique performs better applicability to segment the applications compare than hard clustering algorithm. The algorithm discovers 'c' clusters by reducing the objective function given by below equation.

$$J_{FCM} = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{\ a})^{q} d^{2}(X_k, V_i) \ldots\ldots\ldots\ldots$$

where, the data points represent as $x_k = \{x_1, x_2, \ldots, x_3\}$, the number of data items denotes as n, the number of clusters signifies as c, the degree of membership of $x_k$ in the $i^{th}$ cluster is signified as $u^i_k$, q is the weighting exponent on each fuzzy membership, $v_i$ represents the centre of cluster i, $d^2(x_k, v_i)$ is the distance between data point $x_k$ and cluster centre $v_i$. FCM algorithms form a basis for clustering techniques that dependent on objective functions.

**Algorithm:**

**Step 1:** Consider that M-dimensional N data points signified by $x_i$ (i = 1, 2…N) are to be initially clustered.

**Step 2:** Assume that the number of clusters to be made, that is, C, where $2 \leq C \leq N$.

**Step 3:** Select an appropriate level of cluster fuzziness $f > 1$.

**Step 4:** Initialize the N × C ×M sized membership matrix U, at random, such that $U_{ijm} \in [0, 1]$ and $\Sigma\ U_{ijm} = 1.0$, for each i, j and a fixed value of m.

**Step 5:** Determine the cluster centers $CC_{jm}$, for $j^{th}$ cluster and its $m^{th}$ dimension by using the following expression:

$$CC_{jm} = \frac{\sum_{i=1}^{N} U_{ijm}^{f} x_{im}}{\sum_{i=1}^{N} U_{ijm}^{f}}$$

**Step 6:** Compute the Euclidean distance between $i^{th}$ data point and $j^{th}$ cluster center with respect to, $m^{th}$ dimension similar to the following expression:

$D_{ijm} = ||(x_{ijm} - CC_{ijm})||$

**Step 7:** Update fuzzy membership matrix U according to $D_{ijm}$. If $D_{ijm} > 0$, after that

$$U_{ijm} = \frac{1}{\sum_{c=1}^{C}\left(\frac{D_{ijm}}{D_{icm}}\right)^{\frac{2}{f-1}}}$$

If $D_{ijm} = 0$, then the data point overlaps with the corresponding data point of $j^{th}$ cluster center $CC_{jm}$ and it has the full membership value, that is, $U_{ijm} = 1.0$.

**Step 8:** Repeat from Step 5 to Step 7 until the changes in $U \leq \varrho$, where $\varrho$ is a pre-specified termination criterion.

### III. EXPREMENTAL RESULTS

The objective of extracingt the principal edge features of the image is to produce a clean edge map. The edge detection based image segmentation techniques were implemented using MATLAB R2012a on a PC with Pentium-4 3.20 GHz CPU, 1GB of RAM running Windows XP, and tested with the LEAF image. The images are in JPEG format. Input image is manually collected using digital camera and then the captured image is preprocessed. The preprocessed image then identify and classifies the pest infected region of leaves by using FCM , SVM edge based image segmentation techniques provided. Both FCM and SVM methods detect the infected part of the leaf diseases effectively. Two data sets named training data set and testing dataset are used. The training data set include images which already have been processed and the disease about them have been detected. One image from testing data set is obtained and its characteristics are matched to images in the training dataset. The main aspect of this study is to find the area that is affected by disease would be computed in terms of percentage (%) and then the disease is also detected. Finally, the results of work are implemented in an efficient manner. The effective performance of SVM and FCM techniques are evaluated based on parameters like Standard deviation, mean, RMS, Entropy, Smoothness, Variance, Skewness, Kurtosis, Contrast, IDM, Energy, Correlation, Homogeneity. Finally the accuracy percentage, time complexity of infected region is also computed for FCM and SVM techniques.

**Table 1: Parameter values of Leaf Image**

| PARAMETERS | VALUES | |
|---|---|---|
| | SVM | Fuzzy C-Means |
| Mean | 23.956 | 37.11 |
| Standard Deviation | 64.1271 | 61.5127 |
| Entropy | 1.82979 | 3.551 |
| RMS (Root Mean Square) | 5.91658 | 8.9364 |
| Variance | 3792.75 | 3840.71 |
| Kurtosis | 8.11280 | 3.5232 |
| Skewness | 2.45277 | 1.62921 |
| Contrast | 0.336132 | 0.320692 |
| Correlation | 0.901203 | 0.947059 |
| Energy | 0.696369 | 0.775901 |
| Homogeneity | 0.924350 | 0.838 |

### Accuracy

The accuracy percentage is considered as the ratio of suitably identified image samples to the total number of test image samples. The accuracy percentage is provided by below Equation.

$$Percentage\ accuracy\ (\%) = \frac{Correctly\ Recognized\ Image\ Samples}{Total\ Number\ of\ Test\ Image\ Samples} * 100$$

Initially, the classification is completed using the Minimum Distance Criterion (MNC) with SVM Clustering and demonstrates its efficiency with accuracy of 86%. In FCM algorithm, the accuracy detection is developed into 92.63%. In the second phase classification is completed using SVM classifier and illustrates its efficiency with accuracy of 94%. Currently, the accuracy detection is progressed into 98% through FCM algorithm. The SVM, FCM approaches and its accuracy classification information are exposed in Table 1. As a result, when compare to SVM schemes the FCM classification accuracy was significantly performed better.

**Table 2: Accuracy**

| S. no | Approaches | Accuracy (%) |
|---|---|---|
| **1** | **SVM** | **94** |
| **2** | **FCM** | **98** |

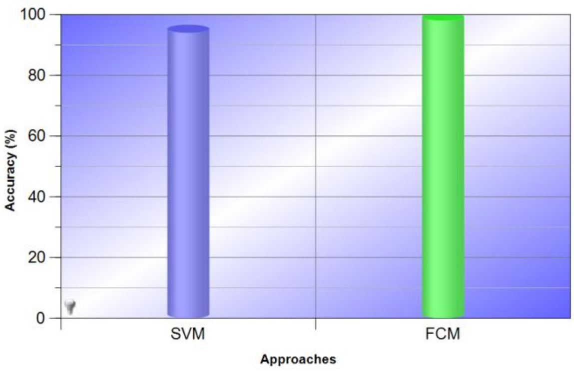


**Figure 2: Comparison of approaches by accuracy**

### Noise Level

The noise level of SVM and FCM techniques are tabularized in Table 2. FCM illustrates good clustering irrespective of noise and contains edge details. It indicates that the sensitiveness of clustering methods to noise data's. As a final point of the result, the noise level of the FCM algorithm is performed better than other techniques SVM.

**Table 3: Noise Level**

| S. no | Approaches | Noise Level (%) |
|---|---|---|
| 1 | SVM | 20 |
| 2 | FCM | 15 |

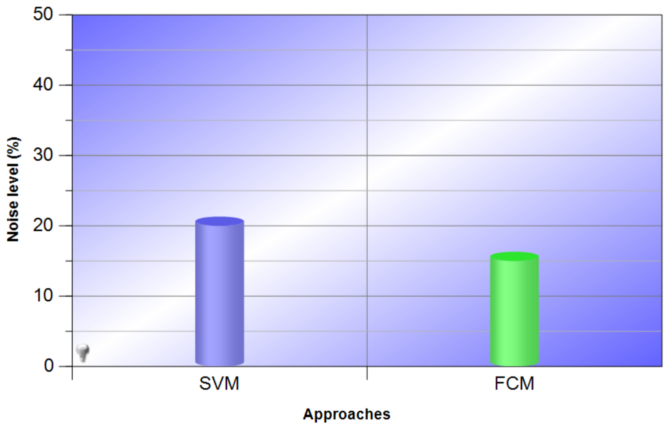


**Figure 3: Comparison of Noise level**

## Evaluation Time

In the evaluation time for SVM and FCM algorithm, it was found that the time consumption of clustering process is less in SVM, but it is very high in FCM techniques, from Table 3. Therefore, accuracy detection is very high in FCM techniques although time complexity is also very high. The evaluation time for SVM and FCM algorithm demonstrate in figure 4.

**Table 4: Time Complexity**

| S. no | No of iterations | SVM | FCM |
|---|---|---|---|
| 1 | 5 | 2000 | 3000 |
| 2 | 10 | 4000 | 6000 |
| 3 | 15 | 8000 | 9000 |
| 4 | 20 | 10000 | 12000 |
| 5 | 25 | 12000 | 15000 |

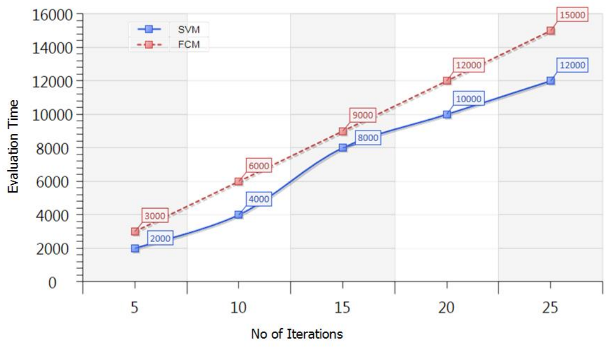


**Figure 4: Comparison of algorithm using Evaluation time**

## IV. CONCLUSION

The challenging task in agricultural field is plant Disease management process and that most diseases are detected on the leaves or stem of the plant. In general, the plants have an effect on fungal diseases, Bacterial diseases, and viral leaf diseases. The major aspect of detection of diseases is accuracy rate. From this paper, Fuzzy C-Means is analyzed and compared with SVM for image segmentation process, in which FCM gives higher accuracy than SVM. FCM provides accuracy 98%, although SVM provides only the accuracy 94%. In evaluation time based comparison, FCM achieves higher time than SVM. Also, it has analyzed that, in SVM technique, even if the number of clusters increased, the computational time decreased automatically.

Therefore, it shows that number of clusters is inversely proportional to the computational time. Fuzzy based image clustering gives better performance than SVM algorithm.

**V. REFERENCES**


[1] Yan Cheng Zhang, Han Ping Mao, Bo Hu, Ming Xili ,"features selection of Cotton disease leaves image based on fuzzy feature selection techniques", IEEE Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. 2007.

[2] Ajay A. Gurjar, Viraj A. Gulhane," Disease Detection On Cotton Leaves by Eigen feature Regularization and Extraction Technique", International Journal of Electronics, Communication & Soft Computing Science and Engineering(IJECSCSE)Volume 1, Issue 1.

[3] Ghaiwat Savita N, Arora Parul "Detection and classification of plant leaf diseases using image processing techniques: a review." International Journal Recent Adv Eng Technology 2014;2(3):2347–812. ISSN (Online).

[4] Dhaygude Sanjay B, Kumbhar Nitin P. "Agricultural plant leaf disease detection using image processing", Int J Adv Res Electr Electron Instrum Eng 2013;2(1).

[5] Mrunalini R Badnakhe, Deshmukh Prashant R. "An application of K-means clustering and artificial intelligence in pattern recognition for crop diseases". Int Conf Adv Inf Technol 2011;20. 2011 IPCSIT.

[6] P.Revathi, M.Hemalatha, "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques" ISBN, pp 169-173, 2012 IEEE.

[7] Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases" IJCA, vol. 17(1), pp. 31-38, March 2011, IEEE-2010.

[8] Jayamala K. Patil, Raj Kumar, "Advances In Image Processing For Detection of Plant Diseases" JABAR, vol. 2(2), pp. 135-141, June-2011.

[9] Gonzalez, R.C., and Woods, R. E. "Digital Image Processing" . 3rd ed., Pearson Education (Singapore) [1992]

[10] Maar, D., Hildreth E., "Theory of edge detection" , Proceedings Royal Soc. London, vol. 207, 187-217, 1980

[11] S.Lakshmi, Dr. V.Sankaranarayanan "A study of Edge Detection Techniques for Segmentation Computing Approaches" IJCA Special issue on CASCT, 2010

[12] V. N. Vapnik, ―Editor, The Nature of Statistical Learning Theory‖, New York, Springer-Verlag, (1995).

[13] N. Chistianini and J. S. Taylor, "An Introduction to Support Vector Machines and other kernel-based learning methods", Cambridge University Press, (2000).

[14] R. S. Gunn, ―Support vector machines for classification and regression‖, ISIS technical report, vol. 14, (1998).

[15] D. Tomar and S. Agarwal, "A comparison on multi-class classification methods based on least squares twin support vector machine", Knowledge-Based Systems, vol. 81, (2015), pp. 131-147.

[16] Y. Yang, Ch. Zheng and P. Lin,"Fuzzy C-Means Clustering Algorithm with a Novel Penalty Term for Image segmentation", *Opto-Electronics Review*, vol. 13(4), pp. 309–315, (2001).

[17] Keh-Shih Chuang, Hong-Long Tzeng, Sharon Chen, Jay Wu and Tzong-JerChen, Fuzzy C-Means Clustering with Spatial Information for Image Segmentation, Journal of Computerized Medical Imaging and Graphics, vol. 30, pp. 9–15, (2006).