

# K-Means Algorithm for ROI Segmentation from Diseased Plant Leaf

<sup>1</sup>T.Nagarathinam & <sup>2</sup>Dr. K. Rameshkumar

<sup>1</sup>Bharathidasan University, PhD Research Scholar, Department of Computer Science, MASS College of Arts & Science, Kumbakonam, Tamil Nadu (India)

<sup>2</sup>Bharathidasan University, PhD Research Guide, Department of Computer Science, MASS College of Arts & Science, Kumbakonam, Tamil Nadu (India)

---

## ARTICLE DETAILS

### Article History

Published Online: 10 November 2018

### Keywords

apple rot, block rot, grapes, k-mean

---

### Corresponding Author

Email: atnaga123[at]gmail.com

---

## ABSTRACT

Plants are one of the most important resources to living things in world. Every human and animal leads their life with the help of plants. Nowadays, most plants are affected by different diseases. When disease ratio is increase, the productivity from the plants decrease, sometimes it stops the productivity. To increase the productivity, need very fast disease detection techniques. For which we attempt to develop a automated segmentation algorithm for diseased portion segmentation from the plant leaves.

## 1. Introduction

K-means algorithms are used in various fields such as customer segmentation, image segmentation, medical image segmentation etc. In this article we attempt to develop and use the k-means algorithm for segment diseased portion from the leaf images. For this, we use three types of leaf images to test the segmentation. The plant leaves are grapes leaf, apple leaf and paddy leaf are used for performance evaluation of the k-means algorithm. K-clustering technique has been reviewed to easy references of the future researchers. The article begins with the introduction and the paper is organized as in section-II exhibit some k-means algorithm which is used by disease image segmentation. Section-2 dealt literature review of the papers, Section -3 dealt working of k-means algorithm. Section – 4 deals with SIFT features extraction and finally deals with conclusion with future work.

## 2. Review of Literatures

Di Cui, Oin Zhang & et al. in [1] proposed a method to fast & accurate detection & classification of plant diseases.

They uses both k-means and c-means algorithm to segments the diseased portion form the soybean leave images. Finally compare the performance of the algorithm.

IDheeb Al Bashish & et al. in[2] proposed a work in various stages, at first stage the k-means techniques is applied to in order to segment the images. The author select five kinds of leave diseases were selected; Cottony mold, tiny whiteness, Early scorch, Ashen mold and late scorch.

H. Al-Hiary, S. Bani-Ahmad et al is described in [3] Fast and Accurate Detection and Classification of Plant Diseases. In this article, the K-means algorithm for clustering the diseases so as to affect on plant leave images The proposed algorithm applied to segment the diseased portion from early and late scorch, ashen mold, cotton mold and tiny whiteness

R. L. Pugoy and V. Y. Mariano[4], proposed an automated system that uses k-means algorithm for cluster the pixels into a number of similar and dissimilar groups from rice leaf images.

In [5] the author title article as “A Hybrid Intelligent System for Automated Pomegranate Disease detection and Grading”. The authors of the paper use various techniques to recognize the diseases on the Pomegranate. They uses k-means technique would be applied to segment the pomegranate images.

In [6], Tushar H Jaware & et al. widen a Fast and accurate method for detection and classification of plant diseases. They have proposed algorithm which is tested on five diseases on the plants; the diseases were tiny whiteness, early scorch, Ashen mold, Cottony mold, late scorch. Secondly, K-means algorithm is making use of segment the images and then green pixels are masked due to it does not contain diseased portion. Next the pixels on the boundaries of the infected object were completely removed. Then the infected cluster was obtained as HIS format from RGB format and generates the SGDM matrices.

In paper [7] N.Valliammal and Dr.S.N.Geethalakshmi combine the K-means algorithm with Wavelet Transform to segment the diseased portion of the leaf images. Finally the concluded the combined segmentation algorithm also improve the performance of the segmentation result.

Wang H, Li G, Ma Z, Li X. in [8] tagged a article as “Application of neural networks to image recognition of plant diseases” proposed a method detect the diseases in wheat and grapevines. They used k-means algorithm for segment diseased portion from both grapes and wheat leaf images.

In [9] Jagadeesh Devdas Pujari<sup>1</sup>, Rajesh et al, has proposed a method which consists of two phases. In the first phase, they used various segmentation techniques to find percentage of affected and unaffected area of the images. The

segmentation techniques are k-means clustering, thresholding, region growing, and watershed.

In [10] Jagadeesh D. Pujari et al, uses K-means algorithm for segment the fungal diseases affected portion from cereals. The diseases are smut, leaf rust, leaf spot, leaf blight and Powderly mildew.

Niket Amoda et al.[11] proposed an automatic detection and classification of plant diseases. The developed a software for K-means clustering technique to segment the diseased portion from the image.

Rastogi, A et al.[12] are identify the leaf diseases using k-means clustering based segmentation algorithm. In this method k-number of clusters are generated using input image and the RGB space is converted into L\*a\*b space without consider of luminosity factor during processing it.

In [13] Amit Kumar Singh, Rubiya .A, B.Senthil Raja proposed an approach to identify rice plant disease Leaf blast. In their approach, they used the k-means algorithm to segment the infected portion of the rice leaf image.

In [14] Nikita Rishi1, Jagbir Singh Gill uses both Otsu method and k-means techniques to segment the diseased portion from the cotton and rice leaf images.

Megha P Arakeri, Malavika Arun, Padmini R K[15] proposed a novel computer vision system for detect and analysis of late blight disease in Tomato to utilize k-means algorithm. The system classifies the leaves of tomato in diseased or healthy using thresholding algorithm. The given image is identified as diseased, the k-means algorithm is applied on the leaf image to attain three clusters such as background, healthy part of leaf and diseased part of the leaf.

In the paper [16] Bed Prakash Amit Yerpude proposed a system to identify the mango leaf diseases and control prediction of the same. Here the authors uses K-means algorithm to segment the diseases portion from the mango leaf image and find the optimal number of cluster to produce the best performance with the accuracy of about 94%.

The following table- 1.shows the summary of the literature review

S.No	Name of Plants	Names of the diseases
1	Soybeans	----
2	Cotton	Cottony mold, tiny whiteness, Early scorch, Ashen mold and late scorch
3	Cotton	early and late scorch, ashen mold, cotton mold and tiny whiteness
4	Rice leaf	Brown spot etc.,
5	Pomegranate	-----
6	Cotton	tiny whiteness, early scorch, Ashen mold, Cottony mold, late scorch
7	Wheat and grapevines	-----
8	Cereals	Smut, leaf rust, leaf spot, leaf blight and Powderly mildew.
9	Cotton	-----
10	Rice	Leaf blast
11	Cotton and Rice	Leaf blight, smut etc.,
12	Tomato	late blight
13	Mango	-----

### 3. K-Means algorithm

Clustering is a technique to split and lay down of data into a various number of groups. It's a one of trendy technique. In k-means clustering, it divides a compilation of data into a k number group of data. It classifies a prearranged set of data into k number of disjoint cluster. K-means algorithm consists of two separate steps. In the first step it computes the k centroid and in the second step it acquires each point to the cluster which has nearest centroid as of the particular data point. There are various methods to define the distance of the adjacent centroid and one of the most used method is Euclidean distance. Once the grouping is done it recalculate the new centroid of every cluster and based on that centroid, a new Euclidean distance is calculated between each center and each data point and allocates the points in the cluster which have minimum Euclidean distance. Every cluster in the partition is defined by both its member objects and its centroid. The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. Thus K-means is an iterative algorithm in which it reduces the amount of distances from every object to its cluster centroid, on the whole clusters.

Algorithm for the K-Means image segmentation:-

**Input:** cropped images( $cm_1, cm_2 \dots cm_n$ )

- Step 1: Read cropped image.
- Step 2: Convert cropped image as RGB
- Step 3: Classify colors using K-Means clustering.

Step 3.1: Define the threshold value for masking.  
 Step 3.1: Green color pixels are masked, if the pixel intensity of the green component is less than the pre-computed threshold value, then the zero value is assigned to the components of green, blue and red pixel.

Step 4: Eliminate the masked cells present inside the edges of the infected cluster.

Step 5: Crop the diseased portions.

**Output:** diseased portions ( $dp_1, dp_2, \dots dp_n$ )

The above algorithm is implemented in MatlabR2014a version to segment the diseased part and the result of cropped view and K-means segmented algorithm segmented is

illustrated in figure-3.1. The segmented portion of the image is easy to process in mobile devices with minimum amount of

computation and with minimum amount of memory.



Fig-3.1 shows the cropped view of the images & K-means algorithm segmented diseased portion.

#### 4. SIFT Features extraction

The SIFT computation encompass of four notable phases: Scale- space detection, Key point localization, Orientation Assignment and Representation of a key point descriptor. The elements are positioned at maximum and minimum of DoG capacities connected in scale space. Then, the descriptors are processed as an arrangement of orientation histograms on 4x4 pixel neighborhoods, and every histogram contains 8 bins. This prompts a SIFT feature vector with 128 measurements on

every patch. The summary of the Sift features extraction steps are

- a. Detection of Scale-Space Extrema
- b. Key Point Localization
- c. Orientation Assignment
- d. Keypoint Descriptor.

The following figure -4.1a and b shows the some keypoints and SIFT features

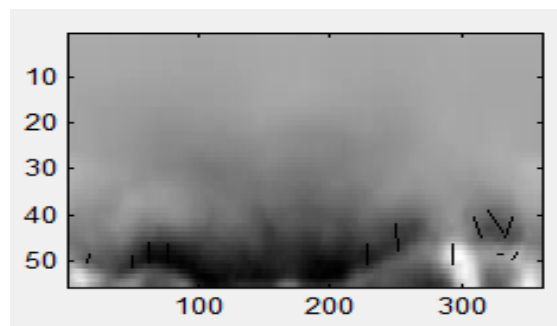


Fig- 4.1 a SIFT Keypoints

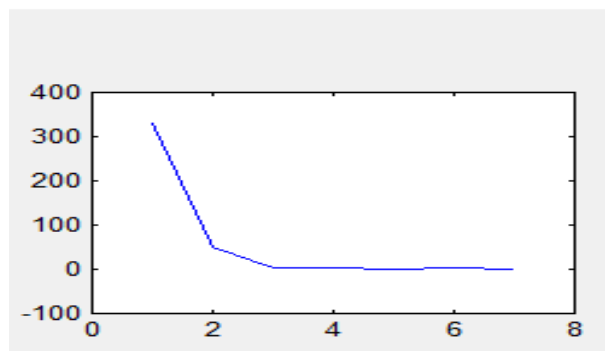


Figure-4.1 b SIFT features

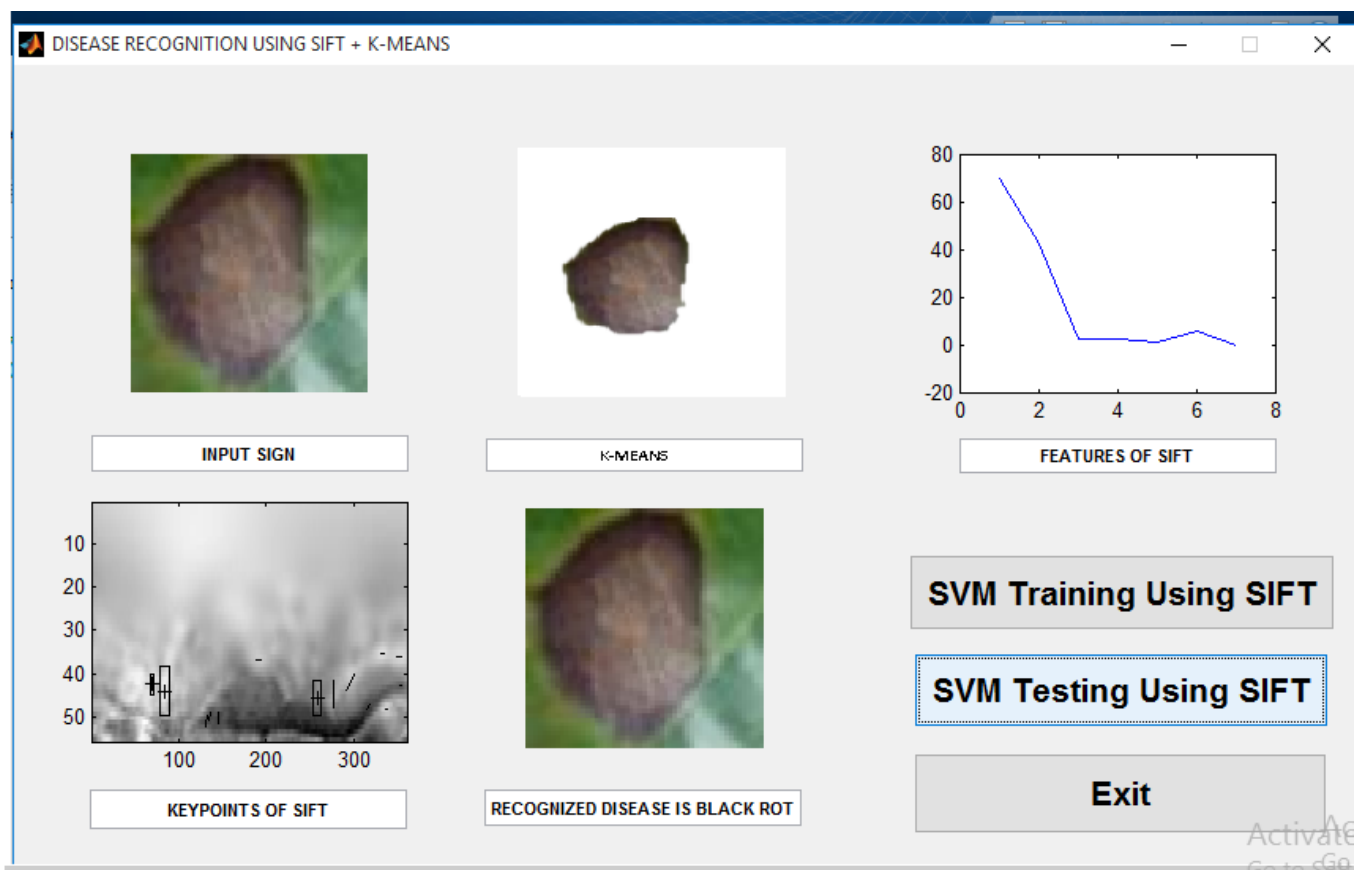
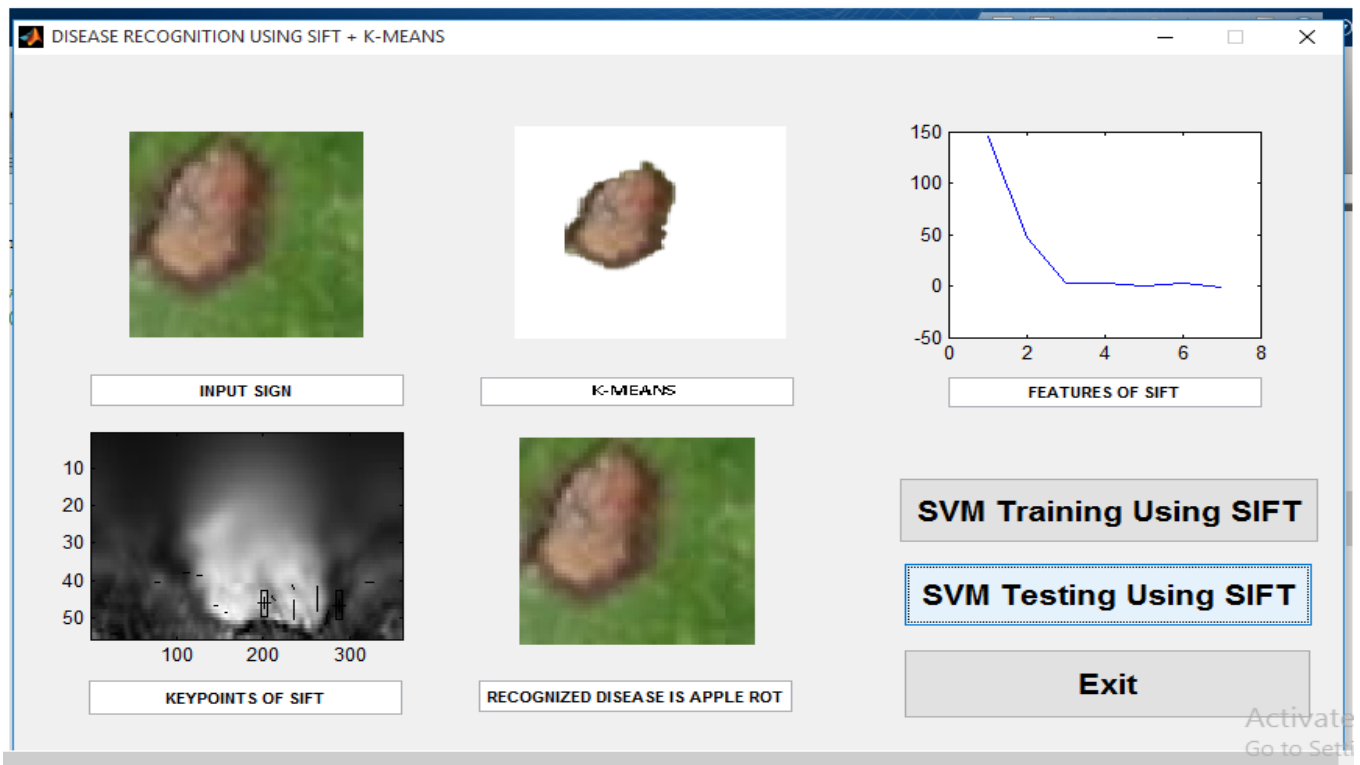
#### 5. SIFT features with SVM

Support Vector Machine is used to construct the optimal separating hyper plane for various plant leaf disease features. For diagnosis a disease, plant leaf diseases features are extracted using SIFT from the input images for the three disease categories. In the training phase, seven dimensional

feature vector is extracted from each diseased image and is given as input to the SVM model. The seven features are x position, y position, scale(sub-level), size of feature on image, edge flag, edge orientation, curvature of response through scale space. For training, seven features per diseases are extracted, but number of keypoint varies and it depends on the image complexity.

For recognition, seven disease features fed into the SVM model and the distance between each of the feature vectors and the SVM hyperplane is derived. The average distance is calculated for each model. The average distance gives better result than using distance for each feature vector. The recognition of the disease is decided based on the maximum

distance. The entire application is implemented in MATLABR2014a version and significant result is obtained. The screen shot of the developed application is shown in figure 5.1 a, b and c which found apple rot disease from apple leaf, black rot and spanesh measles diseases from grapes leaf respectively.



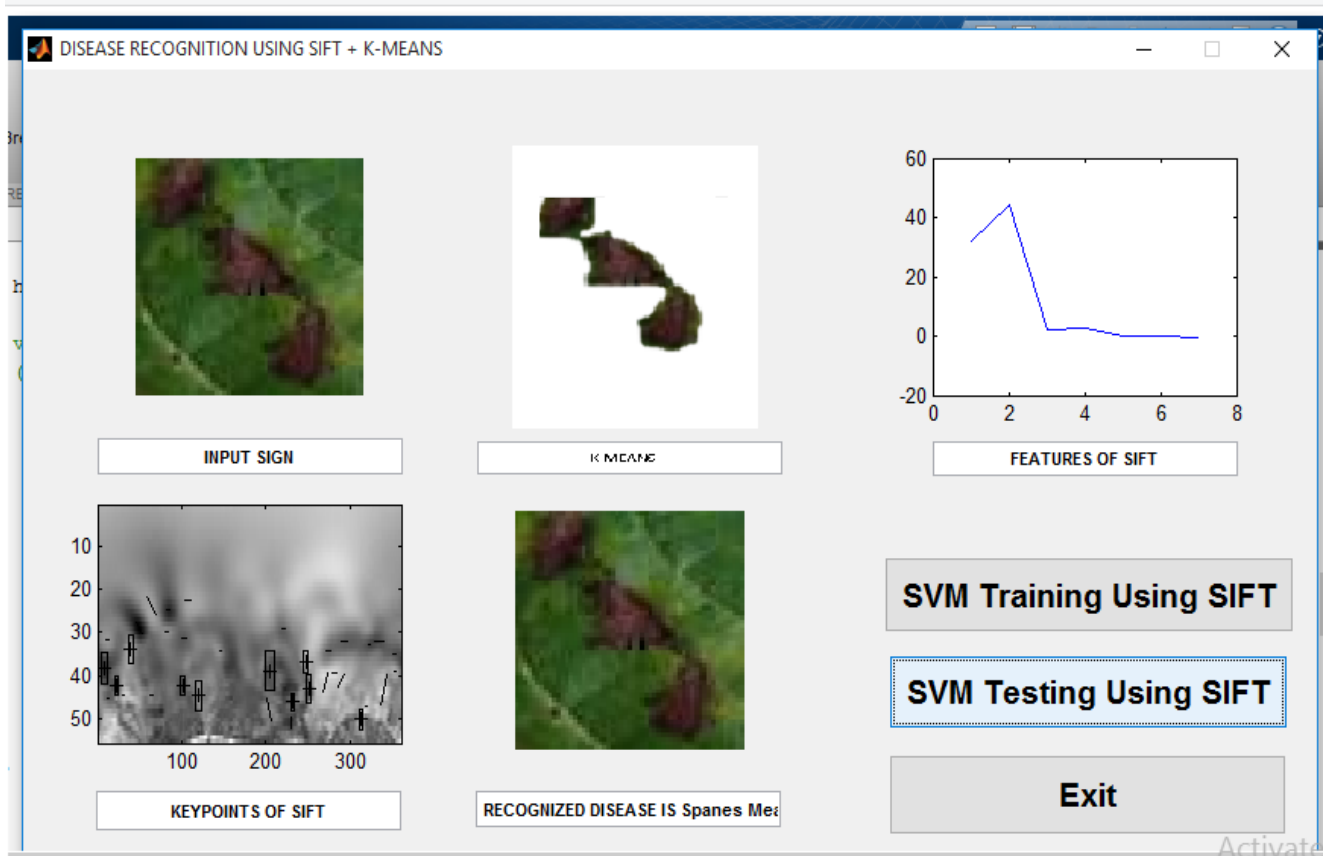


Figure 5.1 a, b and c which found apple rot disease from apple leaf, black rot and Spanish measles diseases from grapes leaf respectively.

The following table 2 shows the Disease-Wise Classification accuracies of the K-MEANS+SIFT+SVM models

Model	Disease type	Training			Testing		
		No., of images	MIS-classification	Classification Accuracies	No., of images	MIS-classification	Classification Accuracies
KMEAS+SIFT+SVM	BR	60	-	100	40	1	97.5
	SM	60	-	100	40	2	95
	BR(A)	60	-	100	40	1	97.5

BR- Black Rot, SM – Spanish Measles, BR (A) - Black Rot (Apple)

The following table 3 shows the performance of the SVM models

Classifiers	Precision (%)	Recall (%)	Accuracy (%)	F-Score (%)
SIFT+SVM	96.74	96.74	97.82	96.74

### 6. Conclusion and Future work

Since this review, we can conclude that there are number of ways by which we can detect diseases in plant. This paper evaluates the K-means techniques in data mining which is used for segmentation. The result of the algorithm is compared with various segmentation algorithms, it produces better segmentation result. SIFT features were extracted from the diseased portion and inputted to the classifier SVM to obtained

the overall classification accuracy of 97.82. Still lot of research is going on for using various techniques to produce automated plant diseases detection via mobile. Our upcoming research article concentrates image classification algorithm for mobile devices with limited memory, higher processing speed and with more accuracy.

### References

- Di Cui, Oin Zhang, Minzan Li, Youfu Zhao and Glen L. Hartman[2009] Detection of soybean rust using a multispectral image sensor, *SpringerLink –Journal article*.
- Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad ,”A Framework for Detection and Classification of Plant Leaf and Stem Diseases”, *2010 International Conference on Signal and Image Processing*.
- H. Al-Hiary, S. Bani-Ahmad et al, Fast and Accurate Detection and Classification of Plant Diseases, *International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011*

4. R. L. Pugoy and V. Y. Mariano "Automated rice leaf disease detection using color image analysis", *Proc. SPIE 8009, Third International Conference on Digital Image Processing (ICDIP 2011), (July 08, 2011)*
5. Sannakki S.S., Rajpurohit V.S., Nargund V.B., Arun Kumar R. and Yallur P.S., "A Hybrid Intelligent System for Automated Pomegranate Disease detection and Grading", *International Journal of Machine Intelligence ISSN: 0975-2927. 2011.*
6. Tushar H Jaware, Ravindra D Badgujar and Prashant G Patil, "Crop disease detection using image segmentation", *Proceedings of Conference on Advances in Communication and Computing (NCACC'12), April 21, 2012.*
7. N.Valliammal and Dr.S.N.Geethalakshmi., "Leaf Image Segmentation Based On the Combination of Wavelet Transform and K Means Clustering", *(IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 1, No. 3, 2012*
8. Wang H, Li G, Ma Z, Li X. "Application of neural networks to image recognition of plant diseases", *Proceedings of the 2012 International Conference on Systems and Informatics (ICSAI) Yantai: IEEE; 2012.; pp. 2159-2164.*
9. Jagadeesh Devdas Pujari<sup>1</sup>, Rajesh et al, Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical
10. Jagadeesh D. Pujari, Rajesh Yakkundimath and Abdulmunaf S. Byadgi. "Classification of Fungal Disease Symptoms affected on Cereals using Color Texture Features", *International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.6, No.6 (2013), pp.321-330 <http://dx.doi.org/10.14257/ijcip.2013.6.29> . ISSN: 2005-4254 IJSIP.*
11. Vivek Chaudhari, C.Y. Patil, "Disease detection of cotton leaves using advanced image processing", *International journal of advanced computer research, June-2014 (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4 Number-2 Issue-15, pg 653-659, June-2014.*
12. Niket Amoda, Bharat Jadhav, Smeeta Naikwadi. "Detection and Classification of Plant Diseases using Image Processing", *International Journal of Innovative Science, Engineering and Technology 2014;1- 7.*
13. Rastogi, A.; Arora, R.; Sharma, S., "Leaf disease detection and grading using computer vision technology & fuzzy logic," *in Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference, vol., no., pp.500-505, 19-20 Feb. 2015.*
14. Amit Kumar Singh, Rubiya .A, B.Senthil Raja- " Classification of Rice disease using digital image processing and SVM Classifier", *International Journal of Electrical and Electronics Engineers ISSN- 2321-2055 (E) <http://www.arresearchpublication.com> IJEEE, Volume 07, Issue 01, Jan- June 2015.*
15. Nikita Rishi, Jagbir Singh Gill- "An Overview on Detection and Classification of Plant Diseases in Image Processing", *International Journal of Scientific Engineering and Research (IJSER) [www.ijser.in](http://www.ijser.in) ISSN (Online): 2347-3878, Impact Factor (2014): 3.05 Volume 3 Issue 5, May 2015.*
16. Megha P Arakeri, Malavika Arun, Padmini R K " Analysis of Late Texture Features, International Journal of Advanced Science and Technology Vol. 52, March, 2013
17. Techniques", *I.J. Engineering and Manufacturing, 2015, 4, 12-22 Published Online November 2015 DOI: 10.5815/ijem.2015.04.02 Available online at <http://www.mecs-press.net/ijem>.*
18. Bed Prakash , Amit Yerpude – " Identification of Mango Leaf Disease and Control Prediction using Image Processing and Neural Network", *IJSRD - International Journal for Scientific Research & Development/ Vol. 3, Issue 05, 2015 | ISSN (online): 2321-0613*