 ISSN (online) : 2277-5390
ISSN (Print) : 2277-5382

SOM based clustering for detecting bacterial spot disease in tomato field

X.Agnes Kala Rani [*1], R.Nagaraj[2]

[1]Professor, [2]Research Scholar, Department of Computer Applications, Karpagam University, Coimbatore, Tamil Nadu, India.

Abstract

Objectives: The main objective of introducing SOM based clustering method is to improve the classification accuracy and detection of bacterial spot disease in tomato field.
Methods: There are various image processing methods used to identify disease and severity of disease in plants. One of such methods uses visible spectrum Images for automatically detecting and classifying the severity of bacterial spot in tomato fields. Centroid-based K-means clustering was widely used for automatic segmentation.
Findings: Plant diseases are one of the major responsibilities for economic degradation in the agricultural industry. So regular monitoring of plant health and early detection of disease causing pathogens are required for minimizing disease spread and assist effective management practices. Centroid-based K-means clustering for segmentation always does not chose centroids that provide best results and also different initial set of centroids affect the shape and effectiveness of the final cluster.
Application/Improvements: To overcome the limitations of Centroid-based K-means clustering, Self-Organizing Maps (SOM) is introduced for achieving effective classification result and to improve the detection performance.

Keywords: Plant diseases, visible spectrum Images, centroid-based K-means clustering, Self-Organizing Maps.

1. Introduction

Image recognition has become more popular in the field of pattern recognition. Similar flow of ideas was applied on the field of pattern recognition of plant leaves, which was utilized for diagnosing plant leaf diseases. It should be noted that, observation of the human eye was comparatively weak and he cannot able to accurately diagnose the disease because minute variation in color pattern represents different types of disease in plants [1].

Visible spectrum images [2] are used at this instant. By obtaining this image, bacterial spot disease was automatically detected in tomato field and it does not require any laboratory procedures. Initially, images were acquired and undergo pre-processing steps. Classification method was then applied for detecting and measuring the disease severity based on segmenting the healthy leaves of the plants with respect to unhealthy leaves, exposed soil and fruits. Centroid-based K-means clustering was utilized to segment visual red and green wavelengths as well as leaves, fruits, soil. Post filtering was then applied to produce visible spectrum image. As a result, this method provides severity evaluation of the disease on the field. But clustering algorithm used in this method may affect the effectiveness of the classification result due to different initial set of centroids. This paper uses SOM to enhance the classification result.

In [3] identified the diseased part of Cotton leaf. The proposed method uses segmentation techniques for effective segmentation based on color and outline of disease spots. Texture and color feature extraction methods are then utilized for extracting features such as boundary, shape, color and texture for the disease spots to classify diseases. Then SVM classification method is used to recognize and classify the diseased part from healthy part. However, more complex structures are required for separation.

In [4] implemented an algorithm to segment the diseased spot in plant leaf. In this approach, color conversion of image was carried out for efficient segmentation result. After color transformation, median filter is utilized on an image for image smoothening. Otsu method is then applied in diseased spot segmentation for automatically selecting most desirable threshold for segmentation. CIELAB color model used in color conversion removed the noise from an image.

In [5] proposed a method which undergoes two segmentation methods to identify the severity of fungus disease in sugar cane leaves. The first segmentation used simple thresholding for distinguishing the leaves from the rest of

ISSN (online) : 2277-5390
ISSN (Print) : 2277-5382

the scene. In the second segmentation, the color transformation from RGB to the HSI color space was done, and a binarization was applied to separate the diseased regions. The threshold for the binarization was estimated by triangle thresholding method based on the gray-scale histogram of an image. The binary image was finally used for estimating the ratio of the infection related to the entire leaf.

In [6] proposed a system to identify two different types of diseases that attack rice leaves. The algorithm first converts the image from RGB to HIS color space. The K-means technique is applied to cluster the pixels into a number of groups. Those groups are then compared to a library that relates colors to the respective diseases. This comparison results in values that indicate the likelihood of each region being affected by each of the diseases.

In [7] utilized indirect methods like spectroscopic and imaging techniques in laboratory or in field-based conditions. Such techniques were generally for plant and disease specifics, since some symptoms may occur in specific wave bands based on the disease and the plant. The cost was similar to whether many hyper spectral wave bands were used or not. Some wave bands need more expensive equipment for obtaining data. If visual spectrum bands were used the cost of acquisition, since handheld consumer cameras could be used.

In [8] tested a multiplex PCR protocol for the purpose of simultaneous species identification, and they were able to identify four species of Xanthomonas strains. Although molecular methods are precise and efforts are being taken to provide kits for common plant diseases they are not fast and have limited capacity for automation in the fields.

In [9] utilized direct plant disease detection methods such as Molecular techniques for extracting plants samples and marking specific reactions to confirm changes when compared to healthy plants. Examples of molecular techniques are Genetic material profiling and antibodies injections. Besides molecular techniques being lab-based they were generally expensive and need more specialized laboring in the procedure steps.

In [10,11] developed an image processing system for detecting and classifying affected areas of leaves caused by early scorch, cottony mold, ashen mold, late scorch and tiny whiteness. Texture features of the regions were extracted and classified with a backpropagation Artificial Neural Network. There were more number of images tested but conditions of acquisition were not reported.

2. SOM based clustering for detecting bacterial spot disease in tomato field

Images are acquired from major areas and different crop sub areas of the fields. For each area, three samples of an image are taken. After image acquisition, image pre-processing is followed in which four procedures such as Gamma correction, Color space conversion, downsampling and Cutting for center parts is included.

Image pre-processing

Gamma correction means adjusting the luminosity levels closer to the human eyes perception. Gamma correction is given by,

$$\mathrm{I_{out}} = \mathrm{I_{in}^{\gamma}} \tag{1}$$

Where I represents the signal intensity

γ represents the real number ranges from 0 to 1

Digital Images acquired from image acquisition are in RGB color space. RGB images are transformed into CIEXYZ color space for defining all possible colors perceived by humans. CIElab color image obtained after transformation possess luminance channel (L), complementary chroma channels (a and b) where a channel represents the red to green conversion and b channel represents the yellow to blue conversion. The standard visualization of each channel is that black represents zero intensities and white represents maximum intensities. RGB to CIEXYZ color space conversion is given in transformation matrix.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [m] \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2}$$

and lab values (L, a and b) are estimated as,

ISSN (online) : 2277-5390
ISSN (Print) : 2277-5382

$$L = \begin{cases} 116 \times \frac{Y}{Y_w}^{1/3} - 16, & if \frac{Y}{Y_w} > 0.008856 \\ 903.3 \times \frac{Y}{Y_w}, & if \frac{Y}{Y_w} \leq 0.008856 \end{cases} \quad (3)$$

$$a = 500 \times \left\{ f\left(\frac{X}{X_w}\right) - f\left(\frac{Y}{Y_w}\right) \right\} \quad (4)$$

$$b = 200 \times \left\{ f\left(\frac{Y}{Y_w}\right) - f\left(\frac{Z}{Z_w}\right) \right\} \quad (5)$$

where X_w, Y_w and Z_w represents reference white points.

Digital Images with 1920 × 1080 pixels, 24 bits are down sampled to 250 × 180 pixels using Gaussian interpolation. Obtained images are cropped for visualizing the centered image area.

Segmentation

Classification method is used to detect and estimate the severity of disease based on classifying healthy leaves of the plants relatively to the unhealthy leaves, bare soil and fruits. It should be noted that, unhealthy leaves, bare soil and scalded fruits are the strong indicators of bacterial spot in tomato field.

Our proposed model enhances the classification result and also improves the detection performance using Self-Organizing Maps (SOM) algorithm on 'a' channel.

Input samples (x_i) are processed one at a time. For each pixel in the samples find Euclidean distance between randomly generated weight vector values and those of its corresponding immediate neighbors. The Euclidean distance is given by,

$$d(T) = \sum_{k=1}^{n} (x_{i,k} - w_{j,k}(T))^2 \quad (6)$$

Then the weights are updated based on the output obtained from (6) with minimum value. The weights are updated by applying following rule,

$$w_j(T+1) = w_j(T) + \alpha\,(x_i - w_j(T)) \quad (7)$$

As a result, similar color pixels are clustered.

Self-Organizing Maps (SOM) algorithm

1. Initialize distance of current neighborhood d (0) to a positive value
2. Initialize randomly generated weight values from inputs to outputs
3. Initialize time, T = 1
4. While computational iteration bounds are not exceeded do
5. Choose an input sample
6. Measure the square of the Euclidean distance from weight vectors (w_j) associated with each output sample
7. Select output sample that has minimum value of weight vector from (6)
8. Update weights to all samples within a topological distance given by d (T) from output sample, using the weight update rule (7)
9. Increment T
10. End while

ISSN (online) : 2277-5390
ISSN (Print) : 2277-5382

Then post filtering process is applied on L, a and b channel. Post filtering in L channel considered the unclassified pixels that have intensity values with too dark or too bright. Post filtering in clustered 'a' channel provides information about bare soil in the image. Visual description of reddish fruits is done by analyzing bothb channel and clustered channel. At last, the image pixels are labelled related to three regions considered. Then the severity of the disease based on the segmentation results is measured using,

$$A_c(\text{class}) = \frac{\text{RS (class)}}{\text{IS}} \quad (8)$$

$$R_c(\text{class}) = \frac{\text{RS (class)}}{\text{IS}-\text{RS }(A_3)} \quad (9)$$

Where Class = A_1, A_2 and $A_3$is one of the segmented areas

RegionSize represents the number of pixels in a particular class

ImageSize represents the size of the processed image

$A_c$represents the absolute coverage

R_c represents the relative coverage

Damage Coverage (D_c) is defined as the relative area of the processed image composed by the unhealthy area. It is given by,

$$D_c = 1 - R_c(A_1) \quad (10)$$

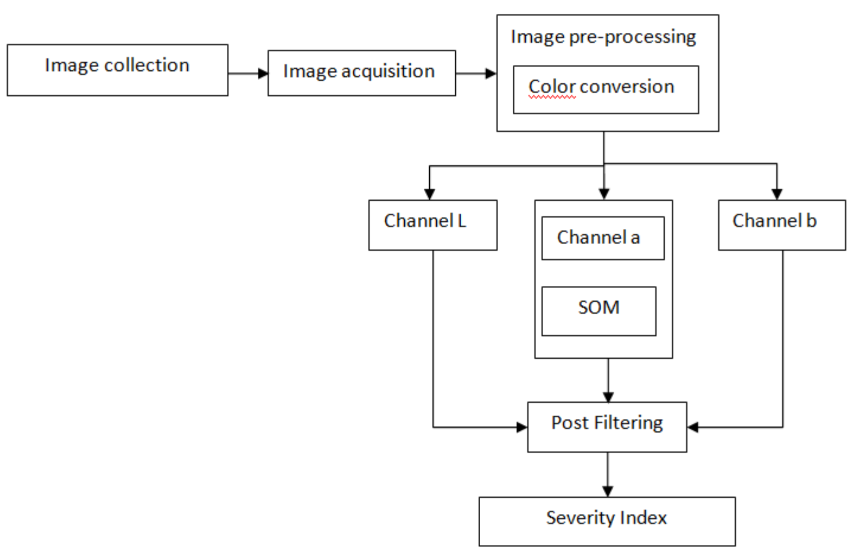
Severity index is measured by,

SI = 1+ 4* D_c

Where $D_c$represents Damage Coverage

The resultant image fields are graded by the Severity Index SI measurements. Linear fitting model with measures of R^2; R^2Adjusted; F-statistic and p-value are estimated for testing and comparing the models. (Figure 1)

Figure 1. Disease identification and classification system



3. Result and Discussion

Figure 2 and figure 3 shows the intermediate output images of both existing and proposed method for one sample input before classification. Initially, channels L,a and b (figure2 and figure3 (c,d and e))are estimated. Then in clustering result of a channel in existing system is shown in figure2 (f) and clustering result of a channel in our proposed system is shown in figure3 (f). It is proved that the output of our SOM algorithm outperforms than the existing algorithm. Then the post filtering result of channel L and b is given in (figure2 and figure3 (g and h)). It is observed that maximum features are seen as useful to distinguish yellowish and dry leaves, scalded fruits and bare soil.(Figure2 and figure3 (i)) represents the three regions A_1, A_2 and A_3 marked with yellow, green and white respectively.

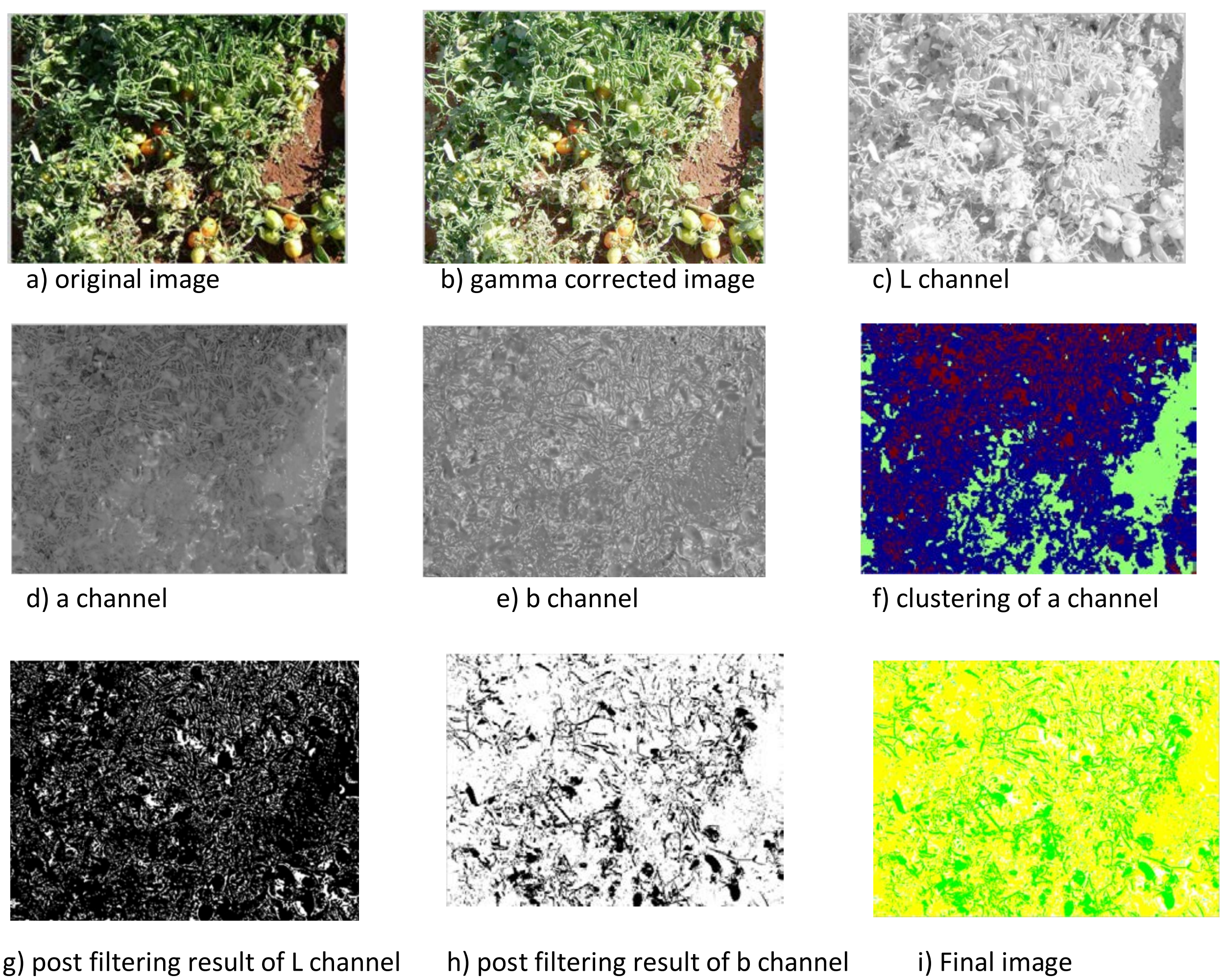
Performance Evaluation

R^2 comparison

R-squared is defined as the fraction by which the variance of the errors is less than the variance of the dependent variable

Figure4shows the R^2 comparison of the existing detection method with k means clustering algorithm, with the proposed detection method with SOM. It is proved from the result that SOM based detection system provides minimum error than k means clustering.

Figure 2. Sequence outputs of the image processing steps for one sample in existing system



a) original image b) gamma corrected image c) L channel

d) a channel e) b channel f) clustering of a channel

g) post filtering result of L channel h) post filtering result of b channel i) Final image

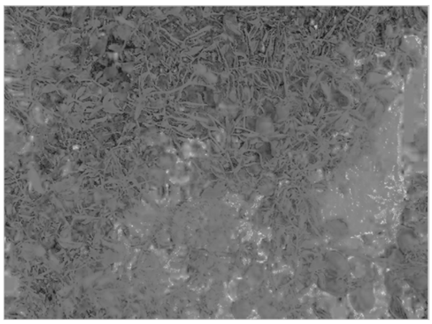
ISSN (online) : 2277-5390
ISSN (Print) : 2277-5382

Figure 3. Sequence outputs of the image processing steps for one sample in proposed system

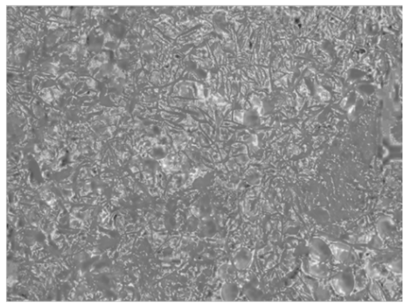
a) original image

b) gamma corrected image

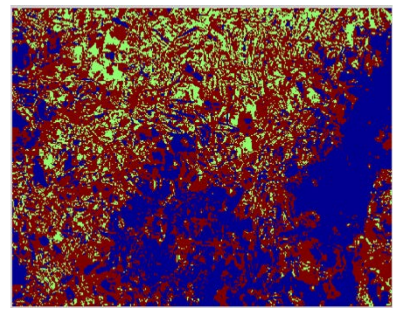
c) L channel



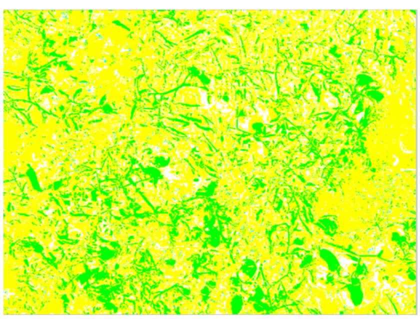
d) a channel



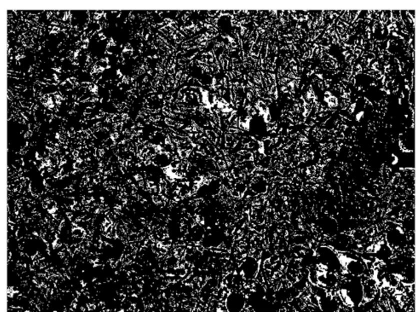
e) b channel



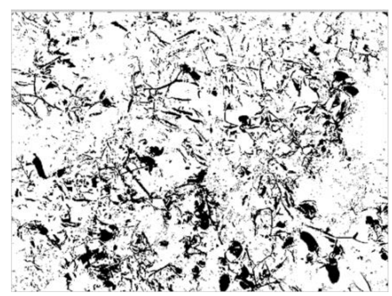
f) Clustering of a channel



g) post filtering result of L channel

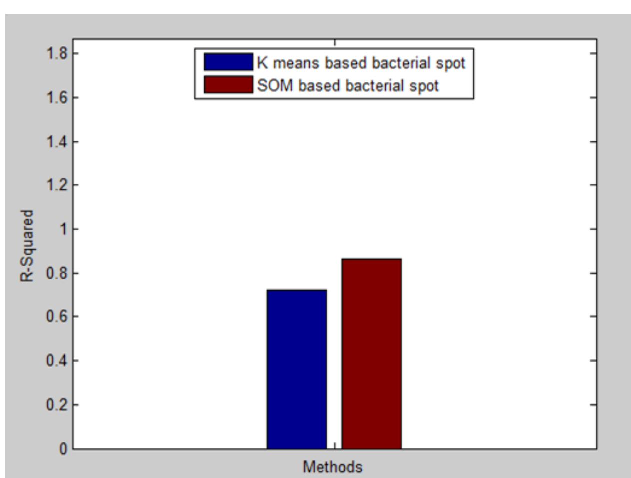


h) post filtering result of b channel



i) Final image

Figure 4. R^2 Comparison of K mean based clustering and SOM based clustering system

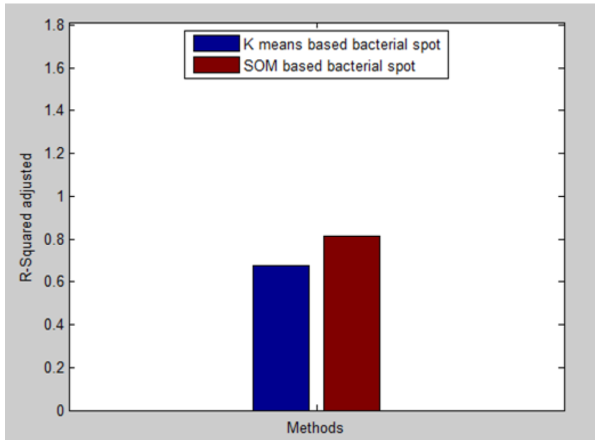


ISSN (online) : 2277-5390
ISSN (Print) : 2277-5382

R^2 Adjusted comparison

Adjusted R-squared means adjustment of the statistic based on the number of independent variables in the model Figure5 shows the R^2 adjusted comparison of the existing detection method with k means clustering algorithm with the proposed detection method with SOM. It is proved from the result that SOM based detection system manages the errors by adjusting independent variables than k means clustering.

Figure 5. R^2 adjusted comparison of K mean based clustering and SOM based clustering system

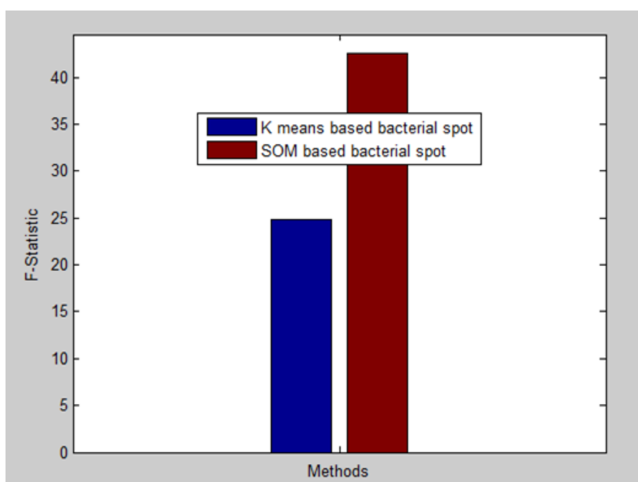


F-statistic comparison

F-statistic is defined as the test statistic for the analysis of variance (ANOVA) approach to test the significance of the model or the components in the model

Figure 6 shows the F-statistic comparison of the existing detection method with k means clustering algorithm with the proposed detection method with SOM. It is proved that, Our proposed model accurately tested the changes occurs in the sample images.

Figure 6. F-statistic comparison of K mean based clustering and SOM based clustering system

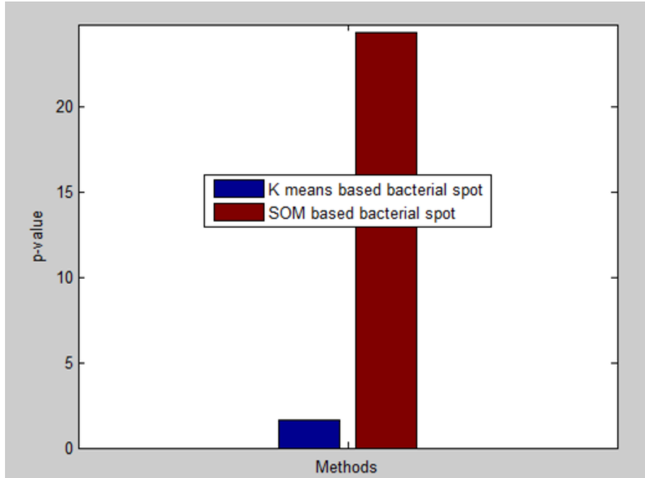


p- value comparison

p-value is defined as the level of marginal significance within a statistical hypothesis test representing the probability of the occurrence of a given event

Figure7shows the p-value comparison of the existing detection method with k means clustering algorithm with the proposed detection method with SOM. It is proved that, our proposed model accurately tested probability of the occurrence of diseases in the sample images.

Figure 7. p-value comparison of K mean based clustering and SOM based clustering system



4. Conclusion

This paper presented an efficient method for automatically detecting and grading the severity of bacterial spot disease in fields of tomato. SOM based Clustering and classification of three different areas with healthy leaves, unhealthy areas, bare soil and scalded fruits are performed automatically. Severity index measurement is used to identify the risk factor of diseases in plants. Results were demonstrated using R^2, R^2adjusted, F-statistic and p-value comparison. Comparison result shows that our proposed system outperforms than the existing system.

5. References

1. D. L. Borges, S. T. D. M. Guedes, A. R. Nascimento, P. Melo-Pinto. Detecting and grading severity of bacterial spot caused by *Xanthomonas* spp. in tomato (*Solanum lycopersicum*) fields using visible spectrum images. Computers and Electronics in Agriculture, 2016; 125, 149-159.
2. S. P. Patil, R. S. Zambre. Classification of cotton leaf spot disease using support vector machine. International Journal of Engineering Research and Applications.2014; 4, 92-97.
3. P. Chaudhary, A. K. Chaudhari, A. N. Cheeran, S. Godara. Color transform based approach for disease spot detection on plant leaf. International Journal of Computer Science and Telecommunications, 2012; 3(6), 65-70.
4. S. B. Patil, S. K. Bodhe. Leaf disease severity measurement using image processing. International Journal of Engineering and Technology, 2011; 3(5), 297-301.
5. R. A. D. Pugoy, V. Y. Mariano. Automated rice leaf disease detection using color image analysis. In 3rd international conference on digital image processing. International Society for Optics and Photonics. 2011, April, 80090F-80090F.
6. S. Sankaran, A. Mishra, R. Ehsani, C. Davis. A review of advanced techniques for detecting plant diseases. Computers and Electronics in Agriculture, 2010; 72(1), 1-13.
7. E. R. Araújo, J. R. Costa, M. A. S. V. Ferreira, A. M. Quezado-Duval. Simultaneous detection and identification of the Xanthomonas species complex associated with tomato bacterial spot using species-specific primers and multiplex PCR. Journal of applied microbiology, 2012; 113(6), 1479-1490.
8. M. M. López, E. Bertolini, A. Olmos, P. Caruso, M. T. Gorris, Llop,M. Cambra. Innovative tools for detection of plant pathogenic viruses and bacteria. International Microbiology, 2003; 6(4), 233-243.
9. D. Al Bashish, M. Braik, S. Bani-Ahmad. A framework for detection and classification of plant leaf and stem diseases. In Signal and Image Processing (ICSIP), 2010 International Conference on IEEE. 2010, Dec., 113-118.
10. D. Al Bashish, M. Braik, S. Bani-Ahmad. A framework for detection and classification of plant leaf and stem diseases. In Signal and Image Processing (ICSIP), 2010 International Conference on IEEE. 2010, Dec., 113-118.
11. V. Brindha, A. A. Mathew. Molecular characterization and identification of unknown bacteria from waste water. Indian Journal of Innovations and Developments, 2012; 1(2), 87-91.