International Journal of Students' Research in Technology & Management
eISSN: 2321-2543, Vol 8, No 3, 2020, pp 01-08
https://doi.org/10.18510/ijsrtm.2020.831

SOYBEAN LEAF DISEASES DETECTION AND CLASSIFICATION USING RECENT IMAGE PROCESSING TECHNIQUES

Arpan Singh Rajput[1*], Shailja Shukla[2], S. S. Thakur[3]
[1*]Research Scholar, JEC Jabalpur (M.P.), India; [2]Prof & Head, Department of Computer Science Engineering, JEC Jabalpur (M.P.), India; [3]Principal, JEC Jabalpur (M.P.), India.
Email: [1*]arpansinghrajput@gmail.com, [2]shailja270@gmail.com, [3]samajh_singh@rediffmail.com

Article History: Received on 25th May 2020, Revised on 22nd June 2020, Published on 1st July 2020

Abstract

Purpose: India is an agricultural country and soybean production is one of the major sources of earning. Due to the major factors like diseases, pest attacks, and sudden changes in the weather condition, the productivity of the soybean crop decreases. Automatic detection of soybean plant diseases is essential to detect the symptoms of soybean diseases as early as they appear on the growing stage. This paper proposed a methodology for the analysis and detection of soybean plant leaf diseases using recent digital image processing techniques. In this paper, experimental results demonstrate that the proposed method can successfully detect and classify the major soybean diseases.

Methodology: MatLab 18a is used for the simulation for the result and machine learning-based recent image processing techniques for the detection of the soybean leaf disease.

Main Findings: The main finding of this work is to create the soybean leaf database which includes healthy and unhealthy leaves and achieved 96 percent accuracy in this work using the proposed methodology.

Applications of this study: To detect soybean plant leaf diseases in the early stage in Agricultural.

The novelty of this study: Self-prepared database of healthy and unhealthy images of soybean leaf with the proposed algorithm.

Keywords: *Soybean, Soybean Plant Disease, Image Processing.*

INTRODUCTION

India is an agricultural country, and the agricultural area is the backbone of its economy. Its economy has traditionally been based on crop production. India occupies the fifth place of production of soybean (Gonzalez, R. C. & Woods, R. E.). According to the Soybean Processors Association of India, India occupies fifth place in the production of soybean and Madhya Pradesh is known as the "Soybean State" of India, comprising 55% of the total national area of soybean cultivation. That is because Madhya Pradesh's nature has blessed it with vast areas of fertile land and abundant success of water, which are the principal ingredients of an agro-based economy. Diseases and insect pests are the major problems in Soybean production. These require careful diagnosis and timely handling to protect the soybean crops from heavy losses. The major problem of soybean plant leaf diseases like Bacterial pustule, Bacterial blight, Rust, Frogeye leaf spot, Downey Mildew, Mung bean yellow mosaic etc. In table1 shows soybean production in the million metric tons in the Indian state.

Table 1: Soybean production in production in the million metric tons in the Indian state

S. No.	State	Sowing Area	Expected yield	Estimated Production
1.	Rajasthan	92012	1025	9445
2.	Madhya Pradesh	54099	1094	59170
3.	Maharashtra	36390	1054	38352
4.	Andhra Pradesh	1791	877	1571
5.	Chhattisgarh	1281	865	1108
6.	Gujarat	1342	925	1241
7.	Karnataka	3190	911	2906
8.	Others	1090	955	1041
Grand Total		**108395**	**1059**	**114834**

Source: http://www.sopa.org/st1.html

International Journal of Students' Research in Technology & Management
eISSN: 2321-2543, Vol 8, No 3, 2020, pp 01-08
https://doi.org/10.18510/ijsrtm.2020.831

Soybean Diseases and Symptoms

The figure of Soybean leaf Disease	Name of the Disease	Symptoms
	Bacterial pustule	In this disease tiny pale green spots with raised centers on both upper and lower leaf surfaces which develop raised pustules in lesion center; pustules usually form in lesions on lower leaf surface; mottled brown areas may develop on leaves if lesions coalesce; small red-brown spots may develop on pods of some varieties.
	Bacterial blight	In this disease water-soaked spots on leaves which enlarge and become necrotic; spots may be surrounded by a zone of yellow discoloration; lesions coalesce and give the plant a burned appearance; leaves that die remain attached to plant; circular, sunken, red-brown lesion may be present on pods; pod lesions may ooze during humid conditions.
	Frogeye leaf spot	Angular gray spots with purple to red-brown edges on leaves; brown to black fungal structures developing in the center of the spot; circular or elongated lesions where the inner membrane of pod contacts the seeds.
	Downey Mildew	Yellow or pale green spots on upper surfaces of leaves which enlarge and coalesce to form yellow patches; lesions may turn gray-brown to dark brown with a yellow margin.
	Mung bean yellow mosaic	Mungbean yellow mosaic disease is characterized by a bright yellow mosaic on the leaves of infected plants and causes significant losses to mung bean (Vigna radiata) crops in India.

RELATED WORK

Agriculture is the mother of India. Research in the agriculture domain is aimed towards increase the quality and quantity of the product at less expenditure with more profit. Farmers require constant monitoring of experts which might be prohibitively expensive and time-consuming. In agriculture, many systems have been proposed to solve or at least to reduce the problems, by making use of image processing and some automatic classification tools (Kanjalkar, H. P. & Lokhande, S. (2013)), (Priya, P. et al., (2015)).

Kutty, S. et al. (2013), proposed the process to classify Anthracnose and Downey Mildew, watermelon leaf diseases. To reduce noise and for segmentation, the median filter is used in this work. And for classification, a neural network pattern recognition toolbox is used. In this work achieved 75.9% of accuracy based on its RGB mean color component.

The goal of Sannaki, S. et al., (2013), is to diagnose the disease using image processing and artificial intelligence techniques on images of the grape plant leaf. This proposed work classifies mainly two diseases, downy mildew and powdery mildew of grape leaf. Masking is used to remove background to improve accuracy. This work used for segmentation is the k-means clustering method. After segmentation, feature extraction takes place by calculating the Gray Level Co-occurrence Matrix. And finally, classification is done by Feed Forward Back Propagation Network classifier in this work. They have used only the Hue feature which gives more accurate results.

Akhtar et al. (2013), have used the support vector machine approach for the classification and detection of rose leaf diseases as black spot and anthracnose. Authors have used the threshold method for segmentation and Ostu's algorithm was used to define the threshold values. In this proposed work, features of DWT, DCT, and texture-based eleven haralick features are extracted which are further used with the support vector machine approach and show efficient accuracy value.

The concept of detection and classification of apple fruit diseases, namely, scab, apple rot, and apple blotch. For that, segmentation is done using the K-means clustering technique. Then features are extracted from the segmented image. For classification, the Multiclass Support Vector Machine (SVM) is used (Jalal, S. et al., (2012)).

Usama Mokhtar et al. (2015), described the technique of Tomato leaves diseases detection and diseases are: Powdery mildew and Early blight. This work used Gabor wavelet transformation is applied in feature extraction for feature vectors also in classification. Cauchy Kernel, Laplacian Kernel and Invmult Kernel are applied in SVM for output decision and training for disease identification.

Khirade, S. & Patil A. B. (2015), discussed the main steps of image processing to detect disease in plant and classify it. It involves steps like image acquisition, image pre-processing, image segmentation, feature extraction and classification. For segmentation, methods like otsu's method, converting RGB image into the HIS model and K-means clustering are there. Among all, the k-means clustering method gives accurate results. In this proposed work classification is done using classification methods like Artificial Neural Network and Back Propagation Neural Network.

PROPOSED METHODOLOGY

Basic steps for the identification and classification of plant disease detection and classification are shown in the following figure 1.

A. Image Acquisition

B. Image pre-processing

C. Image segmentation

D. Feature extraction

E. Classification

F. Result

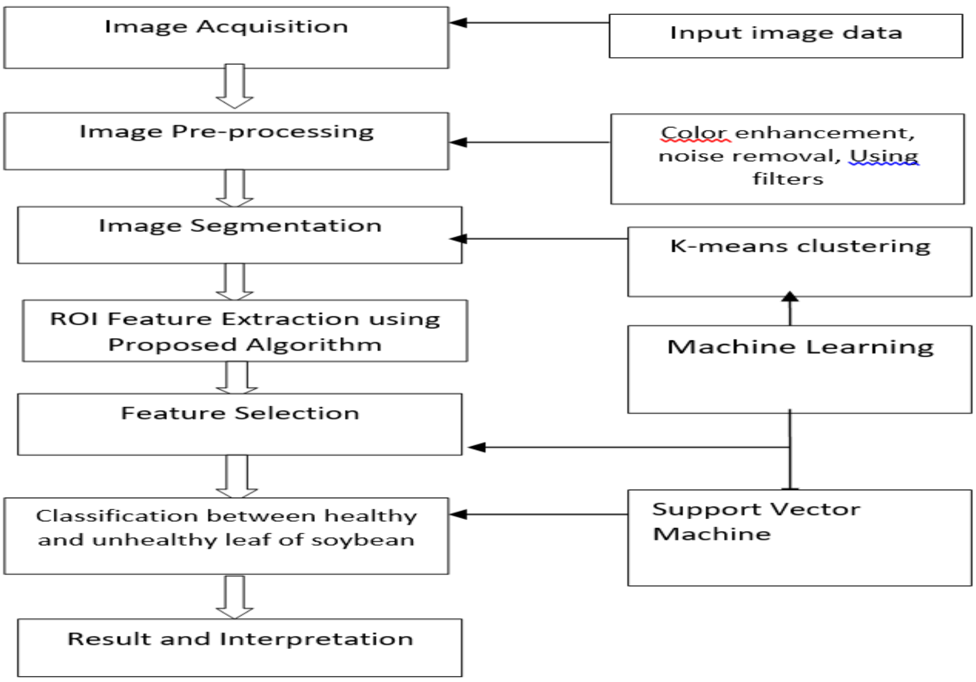

Figure 1: Block Diagram of proposed Algorithm

eISSN: 2321-2543, Vol 8, No 3, 2020, pp 01-08
https://doi.org/10.18510/ijsrtm.2020.831

A. Image Acquisition: This paper uses the self-prepared data set of soybean leaves from the sowing area of soybean in Jabalpur, Madhya Pradesh. This dataset contains different diseased as well as healthy leaves images of soybean. This data set containing 1250 different images of soybean diseases including healthy leaf. The data set shown in figure 2.

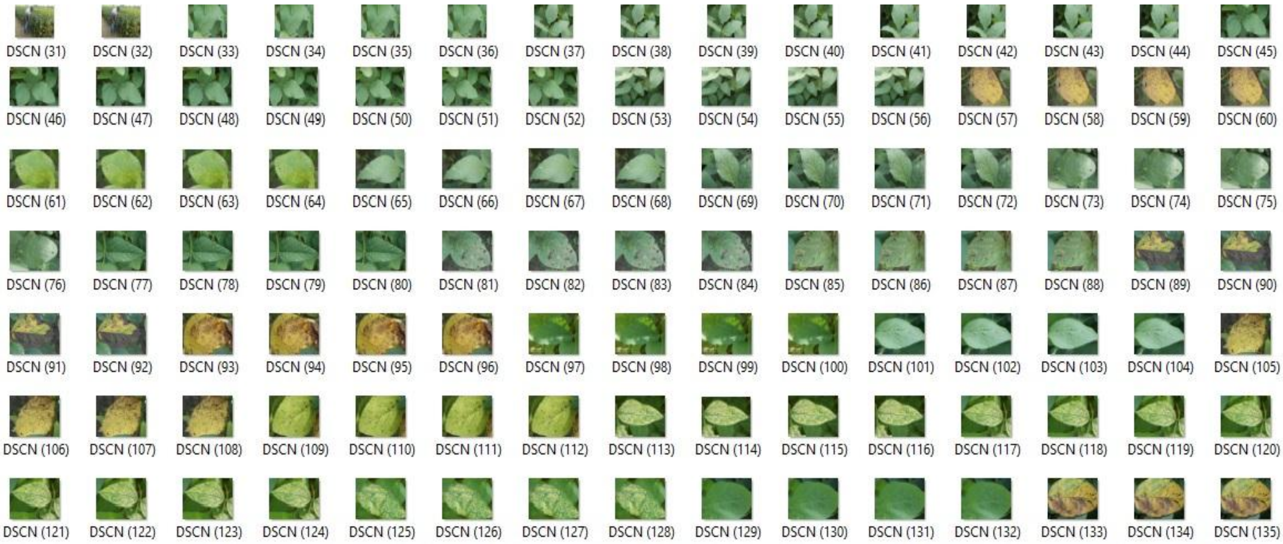

Figure 2: Self-prepared dataset of Soybean leaf

B. Image Pre-processing: After the first step of image acquisition the next step is to pre-process the image with the help of proposed algorithm for removal of excess green and background color and separating of green, blue and red part from background and removal of excess black and white part from the background (Bhong et al., (2016)). Histogram equalization**:** In the color histogram equalization process, the histogram equalization methods applied into separated RGB color components in the image (Gavhale, K. R. et al., (2014)). Color histogram equalization method can be achieved by converting a color plant leaf image into HSV/HSI image and enhancing the Intensity while preserving hue and saturation components (Haralick, R. M. (1973)). However, performing histogram equalization methods on components of R, G and B independently will not enhance the plant leaf image. At the end of this post part, check the histogram of before and after histogram equalization of a plant leaf image which is obtained by performing histogram equalization on the components (R, G and B) independently (Ferentinos, K. P. (2018)).

Steps to be performed:

1. First, convert RGB image into HSI Image.
2. Second, obtain the 'Intensity Matrix' from the HSI Image matrix.
3. Perform a Histogram Equalization method on the intensity Matrix.
4. The last step is to update the Intensity Matrix from the HSI Image matrix with the histogram equalized the Intensity matrix.
5. Convert HSI Image back to RGB Image (Shinde, B. T. S. (2018)).

Figure 3.a. shows the query image and figure 3.b shows the contrast-enhanced image

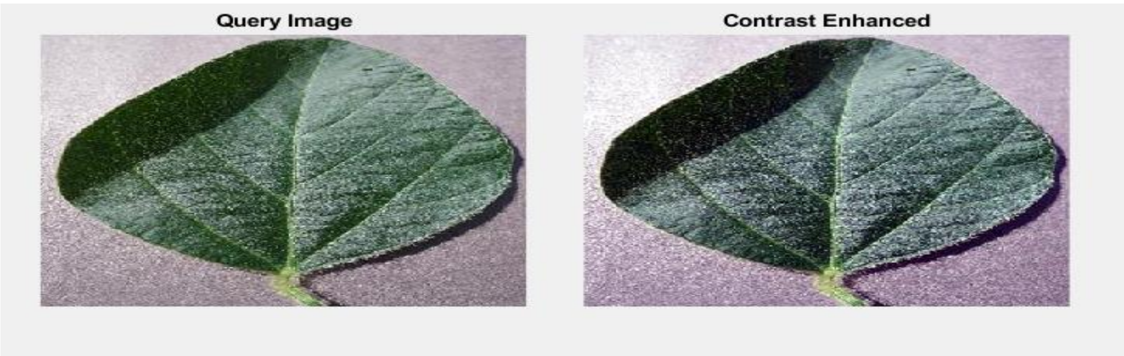

Figure 3: a. Input image **Figure 3: b.** Enhanced image

C. Image Segmentation: Image: Segmentation means partitioning of an image into various parts of the same features or having some similarity measure. The segmentation can be done using various methods like the otsu' method, K-means clustering unsupervised learning (Mollazade, K. et al. (2013)), (Yasikka, M. & Santhi, E. (2015)). In the proposed work K-means clustering is used for the segmentation. The resulting image after segmentation is shown in figure 4.

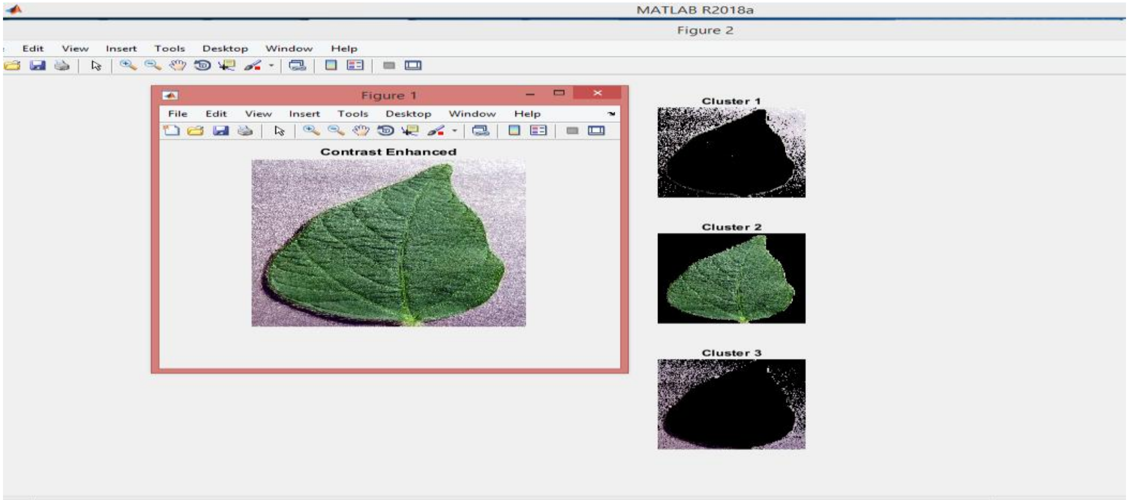

Figure 4: Input enhance the image and segmented image using the proposed algorithm

D. Feature Extraction: In the present paper, texture features i.e. gray level co-occurrence matrix (GLCM) features and statistical features are taken into consideration. GLCM features take the spatial relationship of pixels into consideration. It calculates such pairs of pixels having a specified relationship between the pixels. In this, features like contrast, co-relation, energy, entropy, smoothness and homogeneity were calculated. A gray-level co-occurrence matrix (GLCM) uses image texture to derive acquires features that represent the image (Guo, J. et al. (2011)). In statistical features, skewness, standard deviation, kurtosis and variance values are computed. Consider p (i, j) as the element of the co-occurrence matrix. It represents the probability of moving from a pixel with gray level i to a pixel with gray level j. Then Figure 5.a shows the values of extracted GLCM and statistical features of input images. Then parameters from GLCM can be represented by the below equations:

1. *Contrast:* Measures local variations of gray levels present in the image.

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j) \quad \text{(i)}$$

2. *Correlation:* Determines correlation of a pixel to its neighbour.

$$Correlation = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma i \sigma j} \quad \text{(ii)}$$

3. *Energy:* Provides sum of squared elements in GLCM

$$Energy = \sum_{i,j} p(i,j)^2 \quad \text{(iii)}$$

4. *Homogeneity:* Measures the closeness of distribution of elements in GLCM to GLCM diagonal.

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1=|i-j|} \quad \text{(iv)}$$

5. *Entropy:* The amount of information that must be coded is described by a quantity called as Image entropy.

$$Entropy = -\sum p_i Log p_i \quad \text{(v)}$$

Pi is the probability that the difference between 2 adjacent pixels is equal to *i*.

6. *Average:* Consider a region *R*. *a*[*m*, *n*] represents the pixel brightness. So, the sample mean of this over Λpixels will give the average brightness *ma* in that region.

$$ma = \frac{1}{\Lambda} \sum_{(m,n) \epsilon R} a[m,n] \quad \text{(vi)}$$

where $(m,n) \epsilon R$.

7. *Standard Deviation (S.D):* Consider the brightness of pixels in region *R* of an image. The sample standard deviation (*sa*), is the estimate of the standard deviation of this brightness and is given by:

$$sa=\sqrt{\frac{1}{A}\sum(a[m,n]-ma)^2} \quad \text{(vii)}$$

The other statistical parameters as mentioned earlier were also extracted from the diseased portion of the leaf.

Table 2: Shows the comparison between the feature set of Healthy and Unhealthy soybean leaf

Feature set	Healthy Leaf	Unhealthy leaf
Mean	32.71	40.02
S.D.	50.52	77.55
Entropy	3.41	3.51
RMS	8.02	9.15
Variance	1851.63	5649.28
Smoothness	1.00	1.00
Kurtosis	3.70	3.24
Skewness	1.30	1.35
IDM	255.00	255.00
Contrast	0.89	1.93
Correlation	0.78	0.82
Energy	0.42	0.42
Homogeneity	0.96	0.91

Table 2. shows the feature set of healthy and unhealthy leaf respectively. Likewise, we can get such feature values for other types of diseases. After analyzing these feature sets of the images, a decision can be made by comparing four properties of the healthy leaf with that of an infected leaf whether it is healthy or unhealthy. Thus, up to this stage, the only detection of soybean plant leaf disease is made (healthy or unhealthy). Figure 5. Shows the line graph comparison between the feature set of a healthy and unhealthy soybean leaf.

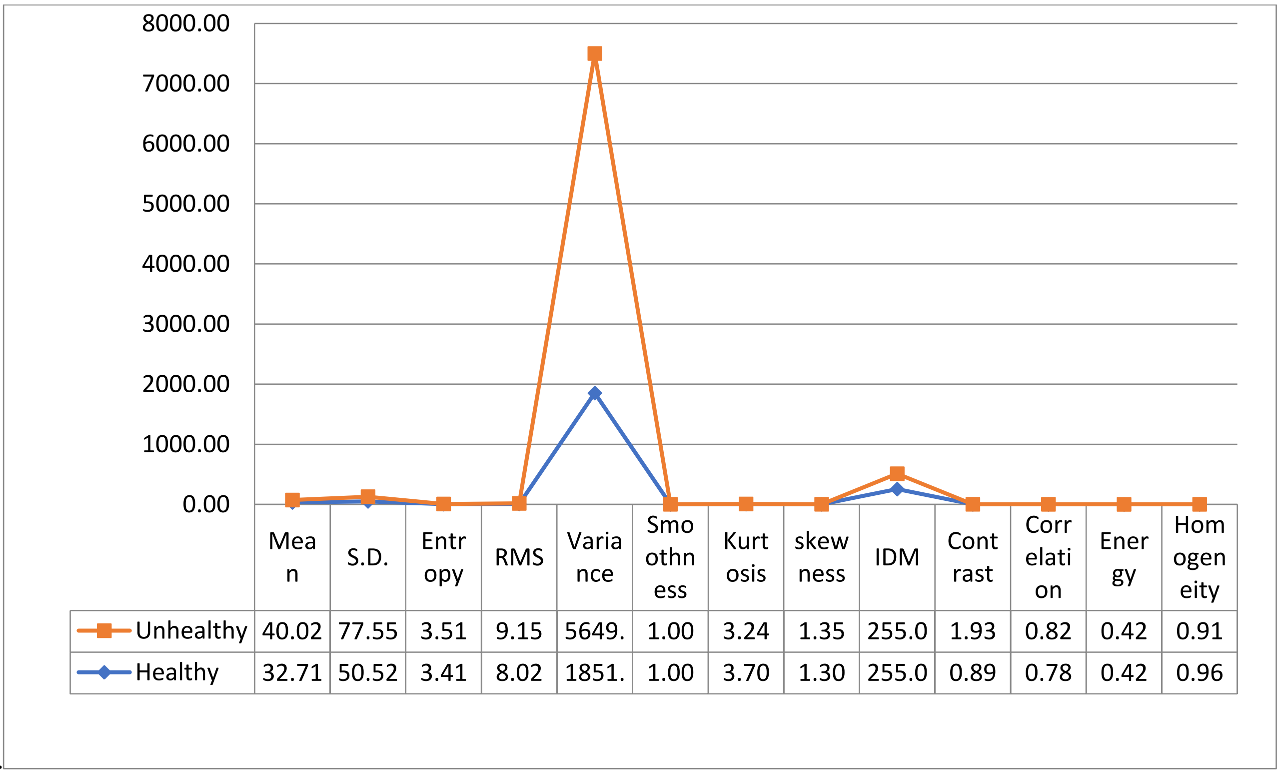

Figure 5: Graphical presentation of feature set between healthy and unhealthy soybean leaf.

E. Classification: This is the final stage of the proposed work where we used machine learning techniques for classification. This gives better and more accurate results. The classifier is tested using a different combination of various features. Using the statistical MATLAB18a commands the other properties find out in the process. Those are Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM for the given images (Al-Hiary, H. et al., (2011)). Figure 6 shows the input image; enhances image output and ROI segmented image number feature calculated through this work and the classified output that soybean leaf is healthy or unhealthy. Next step is to train the model with the following steps:

1. First, start with images of which classes are known for sure.
2. Next is to find the property set or feature set for each of them and then label suitable.
3. In this process take the next image as input and find features of this one as new input.
4. After that implement the binary SVM (Arivazhagan, S. & Newlin, S. R. (2013)), (Athanikar et al., (2016)) to multi-class SVM procedure.
5. Train SVM using the kernel function of choice. The output will contain the Support vector machine structure and information of support vectors, bias value etc.
6. Find the class of the input image.
7. After depending on the outcome species, the label to the next image is given and adds the features set to the database.
8. Steps 3 to 7 are repeated for all the images that are to be used as a database.
9. The testing procedure consists of steps 3 to 6 of the training procedure in the process. The outcome species is the class of the input image in the classification. Figure7 shows the detection of disease.
10. To find the accuracy of the system or the SVM, in this case, a random set of inputs are chosen for training and testing from the database. Two different sets for trains and tests are generated.

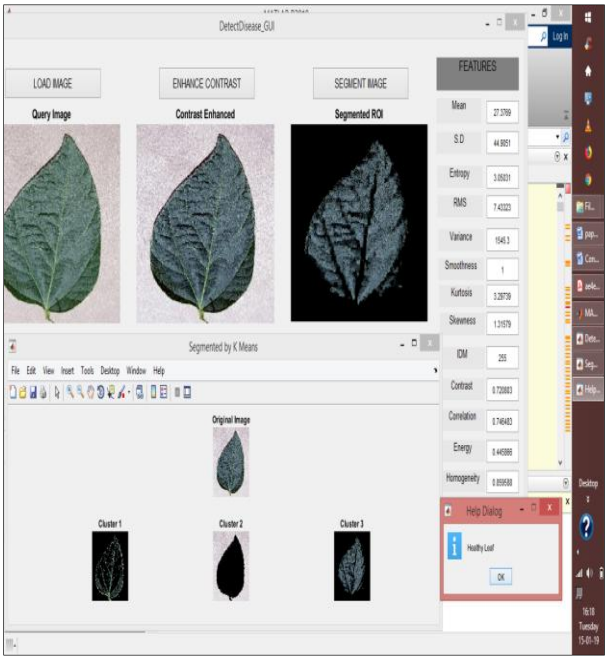

Figure 6: Input image, Enhance image output, ROI

Figure 7: Detection of Soybean disease Segmented image, Feature dialog box, and Final classified output.

CONCLUSION

This paper focused on soybean leaf disease detection and classification methods using image processing, machine learning. Image Acquisition is performed by available data set of 1250 soybean leaves. Image segmentation is performed by the K-means clustering algorithm which gives 3 clusters and Disease affected cluster is chosen and features such as Entropy, Variance, Kurtosis, Mean, Standard deviation and several others are considered to classify the disease present. Classification of the image is performed by SVM to identified healthy and unhealthy leaf. In the future, for segmentation different methods can be used like fuzzy clustering and for classification different combinations of classification methods like KNN classifier, naive bays etc for better accuracy and Deep learning technique will be designed with 100% accuracy for the detection and classification of plant leaf disease.

REFERENCES

1. Akhtar, Asma, Khanum, A., Khan, S.A. & Shaukat, A. (2013). Automated Plant Disease Analysis (APDA): Performance Comparison of Machine Learning Techniques. *IEEE International Conference on Frontiers of Information Technology (FIT)*, pp. 60-65. https://doi.org/10.1109/FIT.2013.19

2. Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M. & Rahamneh, Z. A. (2011). Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications*. 17(1), pp 31-38. https://doi.org/10.5120/2183-2754
3. Arivazhagan, S. & Newlin, S. R. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, *Agricultural Engineering Institute: CIGR journal*, 15(1).
4. Athanikar, Girish, & Badar, P. (2016). Potato leaf diseases detection and classification system. *International Journal of Computer Science and Mobile Computing*, 5.2, pp. 76-88.
5. Bhong, Vijay, S. & Pawar, B. V. (2016). Study and Analysis of Cotton Leaf Disease Detection Using Image Processing. *International Journal of Advanced Research in Science, Engineering and Technology*.ssss
6. Ferentinos, K. P. (2018), Deep learning models for plant disease detection and diagnosis. *Computer And Electronics in Agriculture, Elsevier*, 145, pp 311- 318. https://doi.org/10.1016/j.compag.2018.01.009
7. Gonzalez, R. C. & Woods, R. E., Digital Image Processing, Pearson Education, Third Edition.
8. Gavhale, K. R., Gawande, U. & Hajari, K. O. (2014). Unhealthy Region of Citrus Leaf Detection Using Image Processing Techniques. *IEEE, International Conference for Convergence of Technology*, Pune, pp 1-6.
9. Guo, J., Zhou, J., Qin, H., Zou, Q. & Li, Q. (2011). Monthly stream flow forecasting based on improved support vector machine model. *Expert Syst. Appl.* 38, 13073–13081. https://doi.org/10.1016/j.eswa.2011.04.114
10. Haralick, R. M. (1973). Textural Features for Image Classification. *IEEE transactions on Systems*, 3(6), pp 610-621. https://doi.org/10.1109/TSMC.1973.4309314
11. Jalal, S., Singh, A., Dubey, R. & Ram, S. (2012). Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns. *IEEE Third International Conference on Computer and Communication Technology*, pp. 978-0-7695-4872.
12. Kanjalkar, H. P. & Lokhande, S. (2013). Detection and Classification of Plant Leaf Diseases using ANN. *International Journal of Scientific & Engineering Research*, ISSN: 2229 5518.
13. Khirade, S. & Patil, A. B. (2015). Plant Disease Detection Using Image Processing. *IEEE, International Conference on Computing Communication Control and Automation*, Pune, pp 768-771. https://doi.org/10.1109/ICCUBEA.2015.153
14. Kutty, S. B., Abdullah, N.E., Hashim, A. H., Zraa, A., Rahim, A., Aida Sulinda Kusim,Tuan. Norjihan Tuan Yaakub, Puteri Nor Ashikin, Megat Yunus, Mohd Fauzi Abd Rahman (2013), Classification of Watermelon Leaf Diseases Using Neural Network Analysis, *IEEE, Business Engineering and Industrial Applications Colloquium (BEIAC), Langkawi*, pp 459 – 464.
15. Mokhtar, Usama, Alit, Mona A. S., Hassenian, A. E. & Hesham, H. (2015). Tomato leaves diseases detection approach based on support vector machines. *IEEE*, pp. 978-1-5090-0275-7/15.
16. Mollazade, K., Omid, M., Akhlaghian, F., Kalaj, Y. R., Mohtasebi, S. S. & Zude, M. (2013). Analysis of texture-based features for predicting mechanical properties of horticultural products by laser light. *Computers and Electronics in Agriculture, Elsevier*, 98, 34-45. https://doi.org/10.1016/j.compag.2013.07.011
17. Priya, P., Dony, A. & Dsouza, A. (2015). Study of Feature Extraction Techniques for the Detection of Diseases of Agricultural Products. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 3(2).
18. Sannaki, S., S., Rajpurohit, V. S., Nargund, V. B. & Kulkarni, P. (2013). Diagnosis and Classification of Grape Leaf Diseases using Neural Network. *IEEE, Tiruchengode*, pp 1 – 5.
19. Shinde, B. T. S. (2018). Improved K-means Algorithm for Searching Research Papers. *International Journal of Computer Science & Communication Networks*, vol. 4, pp. 197-202.
20. Yasikka, M. & Santhi, E. (2015). Foliage Measurement Using Image Processing Techniques. *International Journal for Trends in Engineering and Technology*, 5(2).

