

CLASSIFICATION AND SEVERITY DETECTION OF SOYABEAN FOLIAR DISEASES USING ARTIFICIAL NEURAL NETWORK

Swetha V S¹, Ashvant Ram²

BE, Computer Science and Engineering, College of Engineering-Guindy, Chennai, India¹

BE, Computer Science and Engineering, College of Engineering-Guindy, Chennai, India²

Abstract— *Soya bean is one of the most remunerative farming enterprises in India. Soya beans are commercially important as they are edible and a good source of nutrients. soya bean crops suffer from huge losses on account of foliar diseases like soybean-rust, bacterial blight, unexpected death syndrome, downy mildew, frog eye, and brown-spot. Plant diseases drastically affect the crops in quantity and quality. Naked eye observation of plant leaves and detection of diseases is the main approach followed widely. But this is extremely tiresome and time consuming in case of large farms. Expert opinion on disease identification is also expensive. This system proposes an automatic image based disease identification, classification and severity detection system on soya bean leaves. Identification and classification is done using Artificial Neural Networks. K-means algorithm is used for image segmentation. Severity detection is done using 3 parameters namely 'Infection per Region' (IPR), 'Disease Severity Index' (DSI), and 'Disease Level Parameter' (DLP). Thus we intend to build a comprehensive system with the effective implementation of all the above mentioned features.*

Index Terms—*Disease Severity Index (DSI), Disease Level Parameter (DLP), K-means algorithm and Infection per Region (IPR)*

I. INTRODUCTION

Soyabean is one of the most remunerative farming enterprises in India. Soyabeans are commercially important as they are edible and a good source of nutrients. Soyabean crops suffer from huge losses on account of foliar diseases like soybean-rust, bacterial blight, unexpected death syndrome, downy mildew, frog eye, and brown-spot. Plant diseases drastically affect the crops in quantity and quality. Naked eye observation of plant leaves and detection of diseases is the main approach followed widely. But this is extremely tiresome and time consuming in case of large farms. Expert opinion on disease identification is also expensive. This system proposes an automatic image based disease identification, classification and severity detection system on soyabean leaves. Identification and classification is done using Artificial Neural Networks. K-means algorithm is used for image segmentation. Severity detection is done using 3 parameters namely 'Infection per Region' (IPR), 'Disease Severity Index' (DSI), and 'Disease Level Parameter' (DLP). Thus we intend to build a comprehensive system with the effective implementation of all the above mentioned features.

II. RELATED WORKS

The work presented in [1] is focused on the problems associated with the cultivation and highlights the effect of different soya plant foliar diseases on its yield. It has also been presented a full automatic disease discovery and level assessment system that is based on color image sensing and processing. The methodology is based on illumination correction followed by background removal by the threshold intensity levels and red pixel filtering. Several new attributes namely Infection Per Region (IPR), Disease Severity Index (DSI) and Disease Level Parameter (DLP) for measuring disease severity level and for level classification have also been formulated and derived. This helps cultivators understand the severity of the disease and the necessary steps to be taken towards treating the infected plant.

The diagnosis of the disease is done using image processing and suitable artificial intelligence techniques on images of grape plant leaf. In the proposed system input is grape leaf images with complex background. Thresholding is deployed to mask green pixels and the resulting image is processed to remove the noise using anisotropic diffusion. The grape leaf disease segmentation is completed by K-means clustering and diseased portion from the segmented images is identified then. Feature extraction process is done and used for training the system. Best results were observed when Feed forward Back Propagation Neural Network was employed for classification. This work helps us in classification part of the problem statement. The methodology proposed in this work is adopted and more efficient algorithms are to be implemented in order to improve the correctness and effectiveness of the classification process.

A reduct formation rule of rough set theory is used for adapting K-medoid algorithm while membership values of these features are obtained using tenuous sets. A spatial segmentation is employed here wherein an image is divided into different parts having similar properties. The Choice of the initial cluster centers affects the performance of the K medoid algorithm, even though it is a straightforward and effective one. In this editorial, a customized K medoid algorithm is proposed consisting of two parts. In the former part, the preliminary cluster centers are optimized by rough set theory. In the latter part, the optimal

cluster centres are used to execute K medoid algorithm. The proposed scheme doesn't require any prior information on the number of segments. Five discrete state of the art image segmentation algorithms are used to compare the results and they are enriching.

Brain tumor detection is an important task in the medical field because it provides anatomical information on abnormal tissues in brain which in turn helps doctors in treatment planning and patient follow up. An approach to detect and specify anomalies present in brain images is proposed. The idea is to put together two allegories: Neural Network and Fuzzy Logic. These two allegories are united into one mutual system named Hybrid Neuro Fuzzy system which enjoys the reward of both Artificial Neural network system and Fuzzy Logic system, also eliminating their limitations. The Neuro-Fuzzy system combines the knowledge control of Artificial Neural Network system and the explicit information depiction of fuzzy inference system. The proposed system consists of four stages:

1. Data collection through various repository sites or hospitals
2. Pre-processing of various brain images
3. Feature extraction using Gray Level Co-occurrence Matrix (GLCM) and
4. Classification of brain images through Hybrid Neuro Fuzzy System.

Experimental results demonstrate capable results in terms of categorization accuracy, specificity and sensitivity.

An image of the leaf is taken and then processed to find out the condition of each plant. Proposed framework is modeled into four parts:

1. Image preprocessing including RGB to different color space conversion
2. Image enhancement
3. Fragment the section of interest using K-mean clustering for statistical handling to find out the defect and severity areas of plant leaves, feature extraction and categorization.
4. Texture feature is extracted using statistical GLCM and by means of mean values for color features.

This method ensures that chemicals only applied when plant leaves are detected to be affected with the disease. This paper makes a comparison on the different type of classification techniques and explains them. The scope of this work is restricted to artificial neural networks and knowledge based classifiers. The knowledge based classifier has proved better

though requires human expertise to set up rules and make the inference engine understand them in the classification process. The artificial neural network works based on machine learning nodes which help in classification without human intervention which though looks a bit less accurate, results in a completely automated system for classification. The work has considered 600 samples of leaf images and used 300 samples for training and 300 for testing and calculated the results of evaluation.

Anisotropic diffusion has been widely used in various biomedical imaging. Plenty of applications where non-linear diffusion filtering has also been used can be found. The goal of the work in [6] is to examine completely the role of the parameters on the quality of the outcome of the AD discrete scheme and to propose novel methodology for their automatic adaptation and a novel termination criterion for the whole iterative process, so that the final result is optimal. This paper has critical steps of the method and performs comparative evaluation of the various proposed options. The correct alternative and scaling of the conductance function and the technique for estimating the gradient threshold parameter are measured in order to come up with an optimal automatic discrete scheme. They also present a novel stopping criterion based on the quality of the images edges and also evaluate the scheme using a set of natural images. This helps in noise reduction in our work as our work is based on images taken at diverse environments and has a greater possibility of having higher range of noise values which have to be essentially reduced.

III. PROPOSED METHODOLOGY AND DISCUSSION

The proposed system aims at processing soya bean leaf images with complex background. Background removal is done using thresholding and masking. Anisotropic diffusion is used for noise removal. The existing system detects and estimates severity. We propose a system for disease identification and classification. The disease classification is done using artificial neural network. The diseases for which features are trained include Bacterial Blight, Frog eye Disease, Soya bean Rust, Sudden Death Syndrome.

The system as shown in figure 4.1 consists of two parts. First is severity detection and the second is classification of disease. Severity Detection is based on diseased pixel identification followed by calculation of Lesion Color Index (LCI), Ratio of Infected Area (RIA), Disease Severity Index (DSI). Second part is the classification of the disease. The system classifies leaves among the four diseases trained.

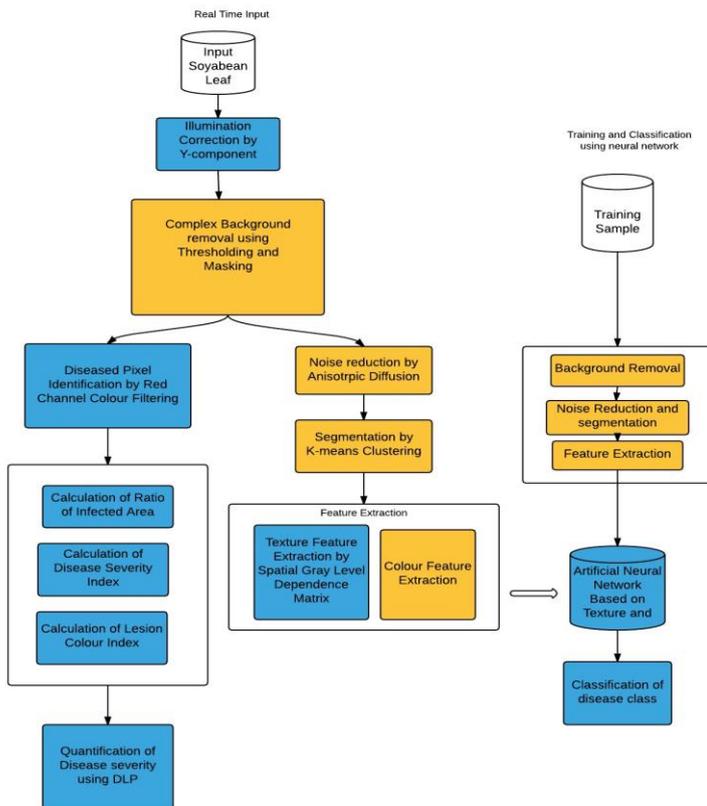


Figure 4.1 Basic Block Diagram

The diseases for which features are trained include Bacterial blight, Frog eye Disease, Soya bean Rust, Sudden Death Syndrome. The textures are extracted for the following diseases and trained to the Artificial Neural Network (ANN). A total of 9 texture features and 18 color features are extracted from each image and trained to the ANN.

A. Illumination Correction

Illumination Correction is done by normalizing the Y-parameter. The Y-component in a YCbCr image denotes the intensity values. Normalizing makes the values either 0 or 1 henceforth eliminating major differences in intensity in different parts of the image. This eliminates difference in image illumination due to sunlight.

Algorithm:

Input(image);

Extract Y component;

Calculate Mean Intensity $E[y]$;

Normalise Y Component: $Y(normalized) = Y/E[y]$;

Show (image);

B. Background Removal

Background removal is done by thresholding green pixels having intensity greater than 70 percent. A green masked image is obtained as a result of background removal.

Algorithm:

Input(image);

greenthreshold=70;

redthreshold= 140;

mask=(greenchannel < greenthreshold) &&

(redchannel > redthreshold);

maskedRed(mask)=0;

maskedGreen (mask)=255;

maskedBlue (mask)=0;

Show(masked image);

There are spots which are yellow which cannot be masked with thresholding green alone. A double threshold with both red and green given values form masking are used. This helps us efficiently mask diseased pixels which have green pixels above the threshold value and appear yellow in color for certain diseases.

C. Noise removal

Noise reduction is done using anisotropic diffusion. It is a space variant non-linear transformation of the original image. Modified Perona- Malik algorithm is used for edge detection and for anisotropic diffusion a smoothing method is used. Noise is characterized by presence of fine grains in the image which create difficulties in edge detection. Perona- Malik anisotropic diffusion smoothens the image by removing these grains and results in a smoothed image. The parameters for the Perona - Malik function are recommended experimental values for perfect noise removal in the gray scale images for macro images.

D. Segmentation

Segmentation is done using k-means clustering. Clustering is the process by which large sets of data are reduced to smaller sets of similar data. Here k-means clustering is used to segment images to 6 clusters. The diseased portions appear in different clusters

for different images. The cluster containing the affected pixels is manually selected by the user for further processing. The cluster containing the diseased portion is called lesion for which texture and color features have to be extracted.

Algorithm:

Read image.

Convert Image from RGB Color Space to a L*a*b* Color Space.

Use K-Means Clustering to classify the Colors in 'a*b*' Space.

Label Every i Pixel in the Image using results from 'KMEANS'.

Create Images that Segment the H and E Image by Color.

Segment the Lesion into a Separate Image.

E. Extraction of Texture and Color features

Grey Level Dependence matrix is used for Texture feature extraction. Color co-occurrence texture analysis is done using Spatial grey-level dependence matrix. Co-occurrence matrices measure the likelihood that a pixel at one particular gray-level given that pixel has a second particular gray-level, will come about at a discrete distance and orientation from any of pixel. The SGDMs are represented by the function P(i,j,d,) where i represents the gray level of location (x,y) in the image, and j represents the gray-level of the pixel at a distance d from location (x,y) at an orientation angle of .., where i is the row indicator and j is the column indicator in the SGDM matrix P(i,j,d).

Co-occurrence matrix is after this normalized. The equation for normalizing co-occurrence matrix is as follows

$$p(i, j) = \frac{P(i, j, 1, 0)}{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i, j, 1, 0)}$$

In the equation above, P (i, j, 1, 0) is the intensity co-occurrence matrix and Ng represents the total number of intensity levels.

Texture characteristics can be used as useful discriminator when target images do not follow well defined color or shape. Apart from these texture features color features are also extracted which are hue, saturation and intensity values of red, blue and green.

FORMULAE

No.	Features	Formula
1.	Contrast	$\sum_i \sum_j i - j ^2 p(i, j, d, \theta)$
2.	Uniformity(Energy)	$\sum_i \sum_j p(i, j, d, \theta)^2$
3.	Maximum probability	$\max_{i,j} p(i, j, d, \theta)$
4.	Homogeneity	$\sum_i \sum_j p(i, j, d, \theta) / (1 + i - j)$
5.	Inverse difference moment of order 2	$\sum_i \sum_j 1 / (1 + (i - j)^2) p(i, j, d, \theta)$
6.	Difference variation	Variance of $\sum_i \sum_j i - j p(i, j, d, \theta)$
7.	Diagonal variance	Variance of $p(i, j, d, \theta)$
8.	Entropy	$\sum_i \sum_j p(i, j, d, \theta) \log(p(i, j, d, \theta))$
9.	Correlation	$\frac{\sum_{i,j} (i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j}$

F. Training and Classification using Artificial Neural Network

Classification is done using artificial neural networks. The features act as input nodes and list of diseases are the output nodes. We make use of sigmoid features and hidden nodes for classification.

Algorithm:

Number of input nodes (i) = 27 (corresponding to 27 features)

Number of output nodes (o) = 4 (corresponding to 4 diseases)

Step 1: Collect the images of produce affected by bacterial blight symptom

blight symptom

Step 2: Extract different colour and texture features

Step 3: Train the knowledge-based and BPNN with extracted features

Step 4: Collect test images and perform step 2.

Step 5: Recognize and classify the images using knowledge based and BPNN.

G. Diseased Pixel Identification

This is done using red-channel colour filtering. Lesions have red component and this feature is used to identify lesion pixels. A threshold is set for red pixels and those pixels are alone taken into account as diseased pixels.

The image is converted to a binary system of image where the diseased pixel is 1 and the healthy portion is 1.

TABLE 4.1 TEXTURE EXTRACTION

Algorithm:

Input (image)

Define threshold for green

If threshold greater than 70

Set $r, g, b = 0$

Open(disk, image)

Show(image)

H. Disease severity**1. Ratio of Infected Area (RIA):**

Ratio of the Infected Area is indicated in percentage of pixels which are contaminated with any of the plant foliar diseases.

The RIA is calculated using formula (1)

$$RIA = \frac{\sum_x \sum_y I(x, y)}{\sum_x \sum_y I(x, y) + \sum_x \sum_y H(x, y)} = \frac{I_A}{I_A + H_A} \quad (1)$$

Where IA represent the infected and HA the healthy areas of the plant leaves, which are the pixel counts for the area of notice using region-props in MATLAB. $I(x, y)$ and $H(x, y)$ are the infected and healthy pixels for the respective regions after processing the soya leaf subsequent the algorithm in planned line of methods.

2. Lesion Color Index (LCI):

The Lesion Color Index indicates the cause for the prejudiced appraisal of the disease sternness level. To specify the disease severity level independently, the R-G color distribution is used to find the Lesion Color Index (LCI). LCI is evaluated using formula (2) on the basis of red and green color values of different pixels.

$$LCI = \frac{(R-G)}{\sqrt{R^2 + G^2 + B^2}} \quad (2)$$

The LCI histogram ranges are stuck between 1 and 1. In LCI histogram, the zero value on x-axis indicates the reference line.

3. Damage Severity Index (DSI):

It is observed that RSI and RIA depends on the total infected area or in supplementary terms, total infected pixels for the image under examination while being used for disease sternness quantification. Apparently higher count of infected pixels is the undeviating hint for sternness of the disease, which is not the case always because affected unhealthy spots may be concentrated o distributed. DSI is calculated using formula (3)

$$DSI = \frac{\max(A_i)_{i=0}^n}{\sum_x \sum_y I(x, y)} \quad (3)$$

$\sum_x \sum_y I(x, y) = \sum_n A_i$ and $1 \leq i \leq n$ where n is the total number of diseased regions in the infected soya plant leaf.

4. Disease Level Parameter (DLP):

The summation of the above measures gives the disease level parameter.

EHL - Extremely High Level

HL - High Level

ML - Medium Level

BMLL- Between Medium and Low level

ANLL- Almost Nil Level

TABLE 4.2 DLP LEVEL CLASSIFICATION

DISEASE LEVEL	DLP
EHL	57
HL	46
ML	34
BMLL	22
ANL	00

IV. EXPERIMENTAL EVALUATION**A. Dataset Description**

Images of leaves with diseases taken into consideration namely Bacterial Blight, Frogeye Disease, Sudden Death Syndrome,

Test4	Bacterial Blight	Nil
Test5	Bacterial Blight	Nil
Test6	Bacterial Blight	Bacterial Blight
Test7	Bacterial Blight	Bacterial Blight
Test8	Frogeye spot	Frogeye spot
Test9	Frogeye spot	Frogeye spot
Test10	Frogeye spot	Frogeye spot
Test11	Frogeye spot	Frogeye spot
Test12	Frogeye spot	Frogeye spot
Test13	Frogeye spot	Frogeye spot
Test14	Frogeye spot	Frogeye spot
Test15	Frogeye spot	Frogeye spot
Test16	Frogeye spot	Nil
Test17	Soyabean Rust	Soyabean Rust
Test18	Soyabean Rust	Soyabean Rust
Test19	Soyabean Rust	Bacterial Blight
Test20	Soyabean Rust	Soyabean Rust
Test21	Soyabean Rust	Soyabean Rust
Test22	Soyabean Rust	Nil
Test23	Sudden Death Syndrome	Frogeye spot
Test24	Sudden Death Syndrome	Sudden Death Syndrome
Test25	Sudden Death Syndrome	Sudden Death Syndrome

Test26	Sudden Death Syndrome	Nil
Test27	Sudden Death Syndrome	Sudden Death Syndrome
Test28	Sudden Death Syndrome	Soyabean Rust

Evaluation:

Here we took 28 sample test cases.

TN / True Negative= sample negative and predicted negative

TP / True Positive= sample positive and predicted positive

FN / False Negative= sample positive but predicted negative

FP / False Positive=sample negative but predicted positive

TABLE 5.2 EVALUATION FOR BACTERIAL BLIGHT

Cases	Predicted Negative	Predicted Positive
Negative Cases	TN = 20	FP = 1
Positive Cases	FN = 2	TP = 5

TABLE 5.3 EVALUATION FOR FROGEYE SPOT

Cases	Predicted Negative	Predicted Positive
Negative Cases	TN = 18	FP = 1
Positive Cases	FN = 1	TP = 8

TABLE 5.4 EVALUATION FOR SOYABEAN RUST

Cases	Predicted Negative	Predicted Positive
Negative Cases	TN = 21	FP = 1
Positive Cases	FN = 2	TP = 4

TABLE 5.5 EVALUATION FOR SUDDEN DEATH SYNDROME

Cases	Predicted Negative	Predicted Positive
Negative Cases	TN = 23	FP = 0
Positive Cases	FN = 2	TP = 3

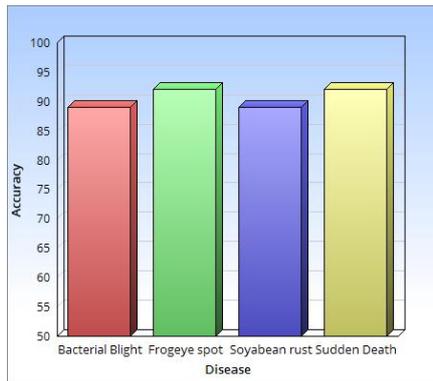


Figure 5.4

High precision as in Figure 5.4 indicates that most of the results are relevant than irrelevant.

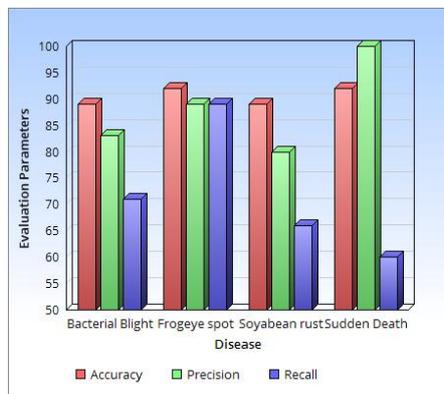


Figure 5.5

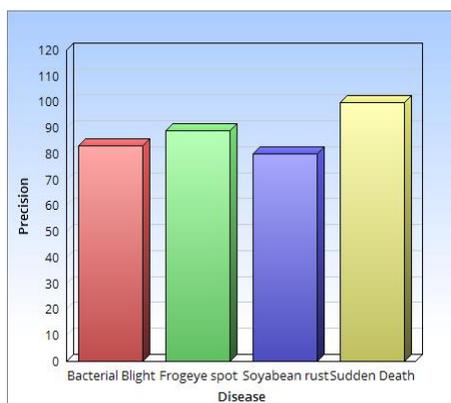


Figure 5.6

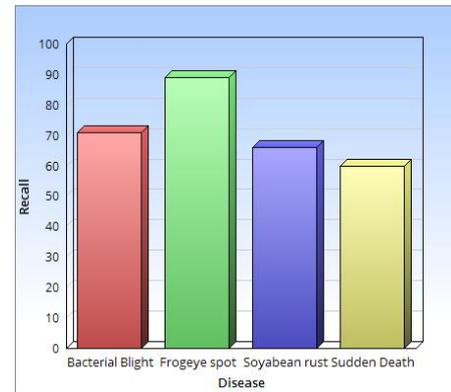


Figure 5.7

Higher precision than recall as in figure 5.7 indicates most of the retrieved result is related but not all related outcome are retrieved. High precision can mainly be attributed to the extensive feature extraction from the images. The 27 node neural network has been effectively trained to produce effective high precision. Recall as in figure 5.4 values are found to be lower in most cases. For all relevant results to be displayed the segment selected has to be done efficiently. Recall values can be improved by efficiently selecting the segment in cluster selection stage. Accuracy as in figure 5.6 appears to be high i.e., more than 90 percent in most diseases proves colour and texture based training of Artificial Neural Network is a reliable method for identification and classification of foliar disease using image processing. Effective background Removal i.e., processing Complex backgrounds is also a key factor towards better accurateness. As the categorization is texture based there are chances that background textures and colours could contribute to the training set. This has been handled efficiently to improve accuracy of the system.

2. Evaluation of Severity Estimation

Severity Estimation can be evaluated by generating quantification values for a set of data and checking for anomalies. Anomalies can be like RIA being high but LCI and DSI indicating very low disease severity. Any of these combinations result in an anomaly.

TABLE 5.6 VALUES OF RIA,DSI,DLP FOR SAMPLE IMAGES.

Input image	RIA	DSI	DLP
Test1	13.906	12.894	25
Test2	12.831	10.783	22
Test3	7.168	12.169	19
Test4	13.883	36.892	50
Test5	42.9665	82.0877	110
Test6	8.859	23.2926	32
Test7	15.3187	78.03	93
Test8	22.984	54.982	76
Test9	35.8	37.952	72
Test10	42.459	91.857	133

From the above values it is found that in most cases RIA, DSI and DLP increase proportionally. This is because they are all dependent on the characteristics of diseased pixels. If the lesions and pixel intensity are high naturally all values and eventually the DLP goes up for the particular leaf. Else it comes down in proportion to the other values.

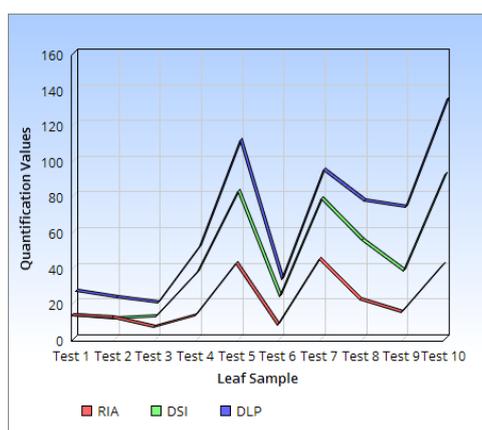


Figure 5.8

The graph in figure 5.8 shows that DLP increases as much as DSI and RIA increase. Hence this shows that these values which quantify the disease are proportional to the severity of the disease and the threshold values are set by experimentation prove that leaves which are not affected predominantly show very low DLP.

V. CONCLUSION

We have developed a system that classifies and quantifies in a soya bean leaf. We have classified 4 diseases and quantified using 3 parameters. We have used anisotropic diffusion for noise removal and k-means clustering for segmentation. We have also used colour and texture features for training.

As a future progress in this work, we could work on the space and time complexity of the algorithm used by replacing k-means

clustering by possibly more efficient algorithms. This tool requires the user to manually select the cluster for lesion extraction. A tool that automatically selects the lesion cluster would serve for complete automation of the process and can be used for unmanned monitoring of the fields for affected leaves. Though the disease severity has been quantified it is true that there is no effective means to evaluate the correctness of the quantification. Parameters that can be effectively evaluated by combining the parameters mentioned in this work can also be an extension.

REFERENCES

- [1] Shrivastava, Sourabh and Singh, Satish Kumar and Hooda, Dhara Singh "Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation", *Multimedia Tools and Applications*, Springer, vol. 74, pp. 1467–11484, 2015.
- [2] Sannakki, S.S.; Rajpurohit, V.S.; Nargund, V.B.; Kulkarni, P., Diagnosis and classification of grape leaf diseases using neural networks, in *Computing, Communications and Networking Technologies (ICCCNT)*, 2013 Fourth International Conference, pp.1-5, 4-6 July 2013.
- [3] Halder, A.; Dasgupta, A.; Ghosh, S., "Image segmentation using rough-fuzzy K-medoid algorithm," in *Communications, Devices and Intelligent Systems (CODIS)*, 2012 International Conference on, vol., no.38, pp.105-108, 28-29 Dec. 2012
- [4] Goswami, S.; Bhaiya, L.K.P., "A hybrid neuro-fuzzy approach for brain abnormality detection using GLCM based feature extraction", in *Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA)*, 2013 International Conference on, pp.1-7, 10-11 Oct. 2013.
- [5] Pujari, Jagadeesh D., Rajesh Yakkundimath, and Abdulmunaf S. Byadgi, "Recognition and classification of produce affected by identically looking powdery mildew disease," *Acta Technologica Agriculturae*, Volume 17, Issue 2 (Aug 2014).
- [6] Tsiotsios, Chourmouzios, and Maria Petrou, "On the choice of the parameters for anisotropic diffusion in image processing," *IEEE Transactions on*, 27(1), 250-263, 2015.