



## Vectored Machine Learning Rearing Process: Early Detection of Leaf Diseases

B Rajendra Prasad<sup>1\*</sup>, T Ramashri<sup>2</sup> and K Rama Naidu<sup>1</sup>

<sup>1</sup>Department of ECE, JNTUA, Anantapuramu, Andhra Pradesh, India

<sup>2</sup>Department of ECE, S V University, Tirupati, Andhra Pradesh, India

Received 12 April 2019; revised 25 June 2019; accepted 5 May 2020

Over the past years, the plant leaf analyses through image processing have drawn a remarkable approach in assessing leaf disease severity through accurate and precise conclusions. We proposed, 'Scale Invariant Feature Transform' (SIFT) based Distinctive Scale Invariant Mapping Procedure (DSIMP) for training images. Random Separation Propagation (RSP) Procedure and Redundant multiclass Support Vector Machine (RM-SVM) are implemented to detect the rice and groundnut leaf diseases at its early stages. Discriminative Gray Level Co-occurrence Matrix (DGLCM) and K means clustering is used for recognition and quantification to give the best color analysis. Experiments with 1000 samples of rice and groundnut leaf images show promising performance.

**Keywords:** Clustering, Disease Severity, Plant leaf analyses, Quantification, Recognition

### Introduction

Plant Disease assessment is required for aiding in settlement of crop insurance claims. It is vital to choose the one that is most appropriate and will provide the most accurate and precise assessments to obtain the right level of disease intensity. This issue in hand received much immersion in the scientific and image processing applications with many researchers focusing on the automatic detection of plant diseases<sup>1-3</sup> combining the features and method to characterize the leaves from multi spectral images. Segmentations algorithms<sup>4</sup> are compared with the proposed models. Noises in the images are filtered using wiener filter by the process called denoising.<sup>5</sup> In order to improve the accuracy of the diagnosis, many researchers proposed different models such as extraction of Abnormalities using ROI segmentation<sup>6</sup>, Probabilistic neural network.<sup>7</sup> A Survey on image processing algorithms for leaf disease detection is presented in Barbedo, 2013.<sup>8</sup> Damaged leaf detection and Stress analysis in fruits are performed in Dubey *et al.*, 2013<sup>9</sup> and Fenyvesi *et al.*, 2013<sup>10</sup> using computer vision.

### Materials and Methods

The proposed rearing process is described in Fig. 1 in detail; the disease affected images are trained and tested proposed work. During training, the detection of feature key points and feature clusters are

identified to train the leaves through RM-SVM. During testing, the recognition using DGLCM and RSP is performed to extract the disease characteristics followed by quantification using k means clustering (KMC) algorithm to create the disease and non-disease clusters and classification process using fuzzy logic. RM-SVM is performed to regions of disease and non-disease areas. Images analyzed for diseases identification in this proposed work are: Afla Root, Alternaria leaf spot, Anthracnose disease, Bacterial Blight, Bacterial Leaf Streak, Bacterial sheath brown rot, Brown Spot of Rice, Bud Necrosis Virus, Collar rot or seedling blight or crown rot, Dry root rot or dry wilt, Early and Late leaf Spots, Flase Smut, Grassy stunt disease, Groundnut Rust, Groundnut Yellow Mold, Leaf Scald, Narrow Brown Spot, Peanut Clump, Root Rot, Ragged stunt virus, Ring Nematode, Root-knot nematodes, Root-Lesion Nematode, Sheath Blight, Sheath rot of Rice, Sheath blight of Rice, Stem rot or Sclerotium wilt, Tungro Virus, Yellow dwarf.

The detected disease area were analysed for feature patterns for performance metrics analysis. For proposed rearing process the images were taken from early stage to final stage and in training, the set of training vectors were represented as  $T_I^{m,n} = \{t_i, i = 1, 2, \dots, l\}$  and the set of output vectors are represented as  $O_I^{m,n} = \{o_i, i = 1, 2, \dots, l\}$  where T is training, O is output, I is input image, (m,n) are image pixels, i is the total number of input images and the ( $t_i$ ) and ( $o_i$ ) are represented in n-dimensional

\*Author for Correspondence  
E-mail: rajendra.14ph0416@gmail.com

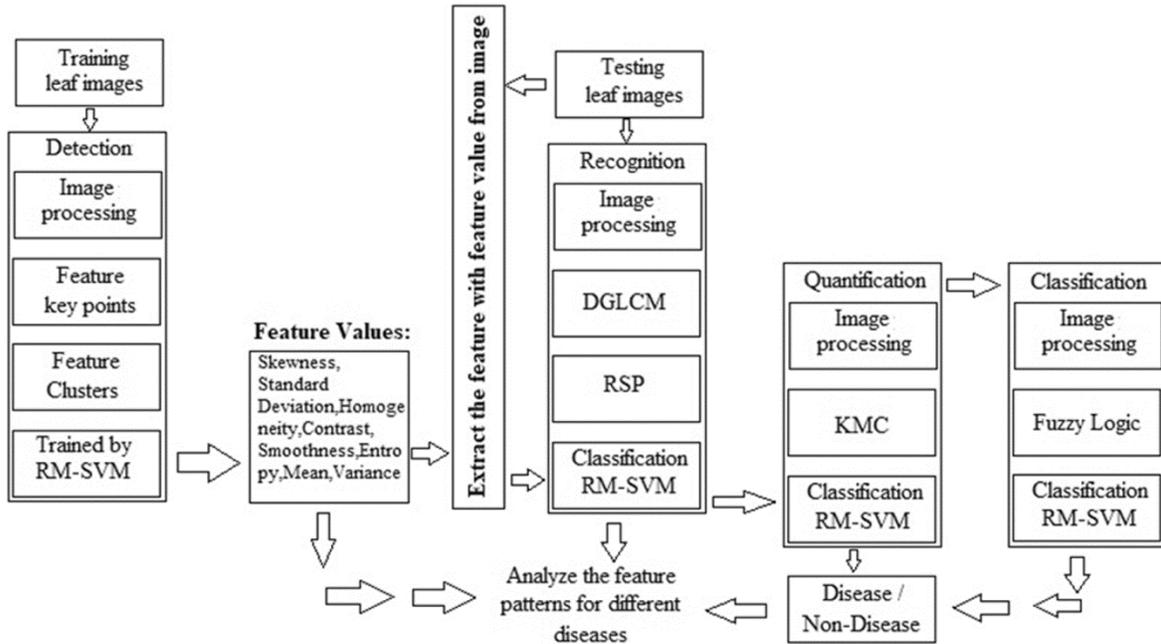


Fig.1 — Proposed System Framework

vectors. Given the training images, learning the image features is performed by mapping, to map input to the output through the mapping function, as shown in the Eq. (1)

$$\sum_i^l \|M(t_i) - (o_i)\| \quad \dots (1)$$

This mapping function is having an input features parameter  $f_i$ , through which the training is made on by choosing  $f_i$  feature parameter and output vectors are observed easily, resulting in to the Eq. (2)

$$M(f_i, t_i) \rightarrow o_i \quad \dots (2)$$

An accurate output image which is similar to input image which is trained results from Eq. (2). But this may produce more than one output image, may result in the mapping function. To make this mapping function accurate and result to only one output image, mutual function error is introduced, by computing the weight of input image where  $W(f_i, t_i) = f_i$  and to compute the output where  $W(f_i, t_i) = R_i$ . The mutual function error for this weight parameter is shown in the Eq. (3)

$$\sum_i^l \|M(f_i, t_i) \rightarrow R_i\| \quad \dots (3)$$

By using Eq. (3) only one output image from training was obtained. By further adjusting the

$W(f_i, t_i)$  value this process guarantees the precise performance of the implemented training work. In this work, the training process is designed to train 1000 images as training set and the mutual function error result in to correct output. The output image of this training process was further analysed for feature extraction.

**Detection:**

This process is performed in two stages:

- Stage 1 — Analysis and extraction of feature key point vectors through implemented SIFT based distinctive scale-invariant mapping and
- Stage 2 — Feature Clusters through implemented fuzzy maximization

**Feature key point vectors:**

*SIFT based distinctive scale-invariant mapping procedure* was implemented in this work to analyse and extract the feature key points from the features extracted image. This method extracts different unchanged features from the images and also having an important parameter in this method to extract the matching in different views of an object in an image with the trained image, with a robust mapping across image viewpoint, noise effected area and illumination varied pixel areas. The key feature of this method is, through a single feature with high probability having large number of images in the data base can extract the matched feature image as shown in Fig. 2.

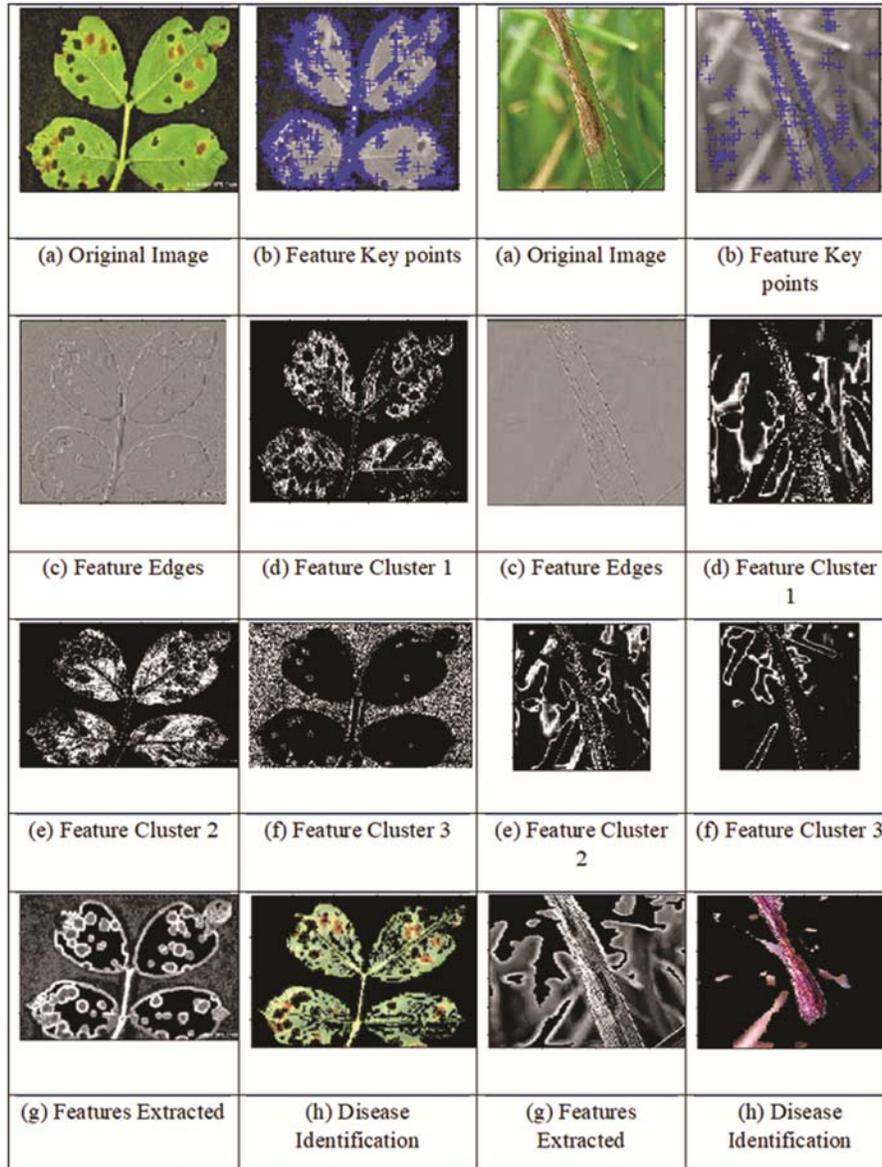


Fig.2 — Proposed Work results on Ground nut Leaves and Rice Leaves

This process is performed in five steps, as shown below:

Procedure 1: Procedure for Feature key points

- Step 1 Individual textures of known image objects are matched with the images in the database.
- Step 2 Scale the known image textures through identification, belonging to matched objects
- Step 3 Verify the matched objects with known objects for consistent unchanged features
- Step 4 Localize the objects at each image location to measure the textures parameter.
- Step 5 Image gradients are measured at the location around each texture and these are transformed

into different local shapes at significant changes to the next level of process.

**Feature Clusters:**

*Fuzzy maximization procedure* is implemented in this work to cluster the features key points extracted with correlation functions.

**Recognition:**

This process is performed in two stages  
 Stage 1 Image Texture Abnormalities: through implemented Discriminative Gray-level co-occurrence matrix (GLCM) Procedure

Stage 2 Disease Knowledge: through implemented Random Separation - Propagation (RSP)

**Image Texture Abnormalities:**

Discriminative gray-level co-occurrence matrix (DGLCM) procedure was implemented in this work, to find the disease features of the input leaf image and also provide the differentiation between the abnormalities of disease and non-disease features of the leaf in subject of interest to the unaffected leaf image. DGLCM procedure extracts the statistical features of the entire image to analyse and detect the disease features through classification based in texture features by discriminating the disease parameters with non-disease parameters. This process was performed in following steps, as shown below:

Procedure for discriminative gray level co-occurrence matrix (DGLCM)

- Step 1 Process: The clustered image is processed using filters, through which the noise and unwanted pixels are distinguished with the disease pixels. Weiner filter is used to perform this operation.
- Step 2 Identify: This step creates a volume of interest (VOI) region to identify the disease affected regions further from the clustered image. It is based on the technique that, there is a large volume of disease area and small volume of non-disease area, and differentiates between them by identifying their filter coefficient values respectively. This step is achieved through the following steps: First step, thresholding was applied on filtered clustered image and values greater than threshold was graded as disease and below as non-disease. Second step, the disease and non-disease areas were marked. Third step, the marked regions were identified with the original image areas. Fourth step, update the filtered cluster image by its original image, and giving the identified disease and non-disease areas in vector form.
- Step 3 Extract: As a function of gray level, the texture level information of identified disease area was analysed through spatial distribution, by computing the local features of the disease and non-disease regions, shows the gray level occurrence at each pixel located at a fixed geometric position relative to each other pixel.

- Step 4 Analyse: For the purpose of leaf disease analysis, the analysis of texture features between the two regions was done to find out the significant features among them, if found they were recognised as disease, else non-disease.

**Disease Knowledge:**

Random Separation Propagation (RSP) Procedure is implemented in this work, to separate the disease and non-disease regions of the extracted DGLCM process, by maximizing the minimum region space between the two regions, by resulting into only one disease depended region called separated vectors. These two regions operate on two mappings, (1). Linear and non-linear mapping of DGLCM image into a high-dimensional feature space, (2). Two regions were mapped based on the optimal separation.

Procedure: For Random Separation - Propagation (RSP)

- Step 1 Read the DGLCM processed image and read randomly 65536 pixels and keep them in a variable matrix.
- Step 2 Train the random pixels with RBF Kernel for disease vectors classification
- Step 3 Cross-validate the pixels, to best-fit the disease regions.
- Step 4 Perform forward and backward computations by disease layer by layer basis.
- Step 5 The local pixels of disease were propagated through number of computations and iterations for disease knowledge until the chosen stopping criteria is met.
- Step 6 For further knowledge analysis, the number of cross-validations can be increased to improve the disease analysis.

**Quantification:**

This process was performed through Kmeans clustering (KMC)

- Step 1 Initialise number of clusters and provide the initial centroid as random value.
- Step 2 Calculate the Euclidean distance from each RSP processed image for each cluster and each disease are assigned to a cluster.
- Step 3 Each disease cluster, a new centroid is computed and each centroid seed value is placed by their respective cluster centroid.
- Step 4 Euclidean distance is calculated from disease object area to each cluster and that object area

is allocate to the cluster with the smallest Euclidean distance.

- Step 5 Continue this process until disease regions are identified as clusters.
- Step 6 Extract the number of clusters as given in step 1.

**Classification: This process was performed in two stages:**

- Step 1 Disease Severity: Fuzzy Logic
- Step 2 Structuring: Redundant multi- class SVM

**Disease Severity:**

Fuzzy Logic is implemented in this work; to grade the disease severity for leaf image under process and the result disease detected grade with its percentage is shown in Table 1. The Procedure for Fuzzy Logic is shown below:

Procedure: For Fuzzy Logic

- Step 1 Read the KMC image and the total leaf area AT and the total Disease Area AD is calculated. The percentage infection (PI) is calculated by this relation:  
 $PI = (AD/AT)*100.$
- Step 2 Perform triangular membership function to grade the disease based on the amount of PI results. The result grade value with the percentage of affected area can be noted.

**Structuring: Redundant multi- class SVM (RM-SVM)**

Redundant multi- class SVM was implemented in this work to linear separate the KMC image into multiple classes, and a decision region is drawn to closely relate to the original image area. The Procedure for Redundant multi- class SVM is shown below:

Table 1 — Comparison of Design Metrics with Rice and Groundnut Leaf images

Parameters	Ground Nut Original Image	Rice Original Image
Affect	15.00	15.01
Mean	51.64	18.87
Standard Deviation	79.41	51.19
Entropy	3.17	1.67
RMS	51.64	18.87
Variance	5356.76	2163.7
Smoothness	1	1
Kurtosis	2.33	9.50
Skewness	1.03	2.74
IDM	255	255
Contrast	2.42	0.54
Correlation	0.77	0.84
Energy	0.40	0.70
Homogeneity	0.87	0.94

**Procedure : Redundant multi-class SVM**

**Input:**  $R \in \mathcal{C}^2$ , redundant features  
 $C^*$  is the classifier model,  
 $W = \{x(i) \in R, y(i) | i = 1, 2, \dots, N\}$  are the previous samples,  
 $x$  is the location of the disease,  
 $y$  is the disease level,  
 $k$  is the number of cluster vectors required,  
 $D$  is the distance between clusters,  
**Output:**  $C_{next}$ , is the number of clusters along  $R$  within the cluster.  
 $R \leftarrow (\alpha, \beta)$   
**for**  $i=1:N$  samples **do**  
     $R \leftarrow (\alpha, \beta)$   
     $D^* \leftarrow D$   
    **for**  $j=1:k$  **do**  
         $x \leftarrow argmax_{r \in R}$   
        Expected Cluster Centroid ( $r, u_y, P(y|x, D^*)$ )  
        Distance along Clusters ( $x, R$ ) to  $C$   
        Features  $P(y|x, D^*)$   
        Observations  $D^*$   
    **end**  
    Assign  $C_{(1)}$  to Disease Area  
**end**  
 $C_{next} = 1/N_{samples} \sum_i S_i$

**Advantages and Challenges:**

The percentage of the successful classification of the proposed model is best compared to other conventional models. Our proposed system automatically detects, classifies and quantifies the plant leaf with the best accuracy as shown in Fig. 3. RM-SVM's are a kind of discriminative classifier which shows great performance for classifying image features. They are accurate, powerful and quick to apply to test samples. Texture of leaf image is easily examined by using DGLCM. K means clustering produces tighter clusters than other clustering methods only thing is prediction of K value is of at most important

In spite of the high success rate of the proposed framework, there are different reasons that make it still very a long way from being a conventional tool that could be utilized in real conditions. The issue that should be noted and should be settled is that the testing dataset utilized for the evaluation of the models, was a part of a similar database that comprised the training set. This is basic practice in all models; however, the real estimation of the created framework, particularly for having the option to be

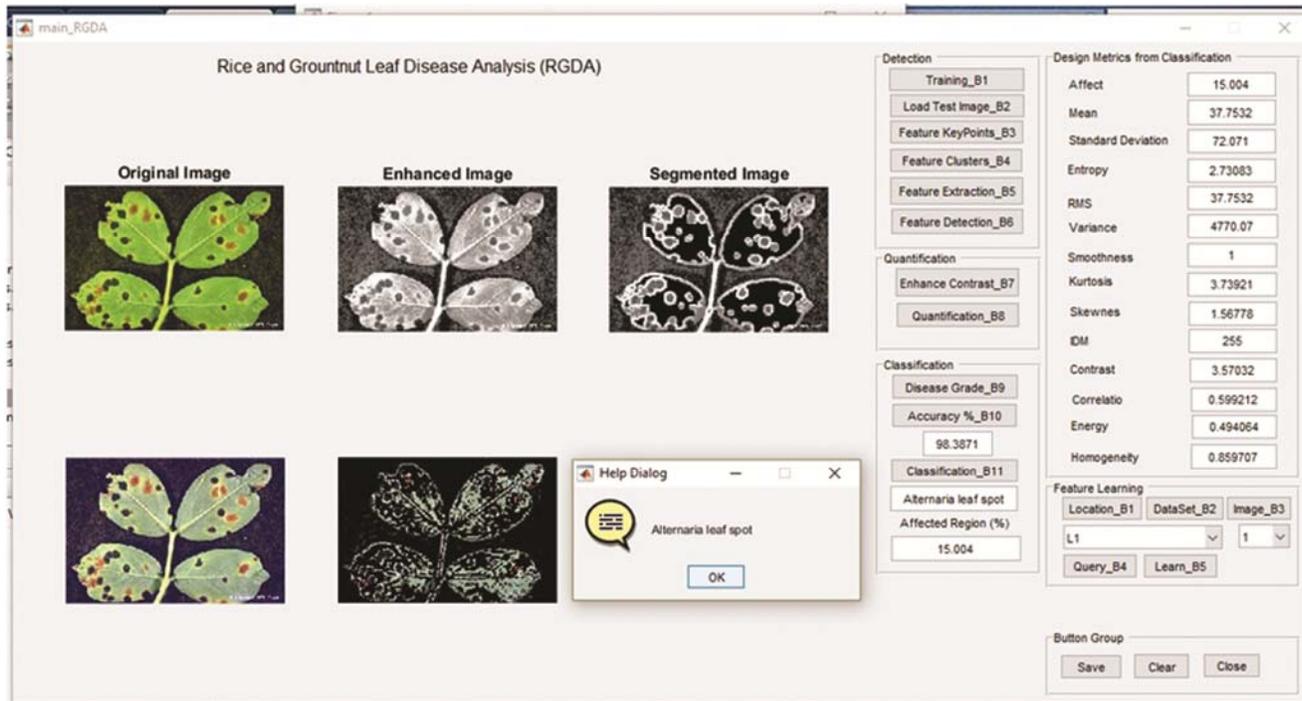


Fig. 3 — Rice and Groundnut Leaf Disease Analysis (RGDA) System GUI

Table 2 — Rice and groundnut leaf disease recognition accuracy and affected region

Between the same diseases	Among the different classes diseases	The total recognition rate
93.125 %	97.681 %	95.211 %
Image	Accuracy (%)	Affected region (%)
Rice Disease Image	98.2	17.59
Groundnut Disease Image	98.38	59.54

utilized in real circumstances, should be demonstrated in testing information that would come from various sources as well as databases.

**Results and Discussion**

Experimental environment is as follows: the operating system is windows 10, Intel Core I3 Processor, 4 GB internal storage RAM, and the software is matlabR2015a. Experimental Subjects: 1000 images are selected to verify the accuracy of the proposed algorithm. Every disease responds to 25 images, and there are 39 kinds of diseases. There are ten templates for each type of disease and 100 test samples. Visual comparisons with the existing methods are tabulated in Fig. 2. The comparison of design metrics with Rice and Groundnut images are shown in Table 1. The 1000 plant disease images are divided into 8 independent subsets, and each subset

contains 125 plant disease images. Each subset has 10 template images and 312 training samples. The training and testing process is repeated for every process, and finally the best test results are averaged.

From Table 2, recognition rates between the same diseases and recognition rate among the different classes’ diseases reach above 90%. The accuracy and affected region are also shown in Table 2.

Accuracy of the proposed work is compared with BLDC<sup>11</sup> (0.8528) tested on Banana leaf, SVM<sup>12</sup> (0.82) tested on Wheat, If-then classifier<sup>13</sup> (0.75) tested on Rice, Back-propagation Neural Network<sup>14</sup> (0.9) tested on Rice, Rule Generation<sup>15</sup> (0.9229) tested on Rice, which gives the best performance of proposed work.

**Conclusions and Future Scope**

Because of the inefficient rice and groundnut leaf disease processes, proposed work put forward a new rice and groundnut leaf disease recognition algorithm using the random separation propagation (RSP) procedure and redundant multi class SVM. These methods can improve the speed of rice and groundnut leaf disease recognition. With the experimental analysis, the average accuracy of RSP procedure for rice and groundnut leaf disease recognition is up to 98.2% and 98.38% respectively. Thus, RSP procedure is an efficient method for recognition of leaf diseases.

The proposed algorithm comparatively can recognize the rice and groundnut leaf disease, and it also has guiding significance for other disease recognition to an extent. In order to improve the accuracy, a wide variety of training data samples be collected, from various sources of different geographical areas, cultivation conditions, and image capturing modes and sets.

### References

- 1 Gouveia F & Reis V, Biometry: The characterization of chestnut tree leaves using computer vision, *IEEE Int Symp on Ind Electron*, **3** (1997) 757–760.
- 2 Bauer S, Korc F & Forstner W, The potential of automatic methods of classification to identify leaf diseases from multispectral images, *Precis Agric*, **12** (2011) 361–377.
- 3 Cui D, Zhang Q, Li M, Hartman G L & Zhao Y, Image processing methods for quantitatively detecting soybean rust from multispectral images, *Biosys Engg*, **107** (2010) 186–193.
- 4 Altuntas V, Altuntas S & Gok M, Automatic colony segmentation on agar surface by image processing, *J Sci Ind Res*, **77**(08) (2018) 451–456.
- 5 Xiaobo Z, Image denosing using dual-tree complex wavelet transform and wiener filter with modified thresholding, *J Sci Ind Res*, **75**(11) (2016) 687–690.
- 6 Vidhya K, Babu T R G & Devi S S, Extraction of abnormalities in MRI, CT, X-ray and Ultrasound images towards development of efficient compression algorithm, *J Sci Ind Res*, **77** (01) (2018) 29–43.
- 7 Wu S G, Bao F S, Xu E Y, Wang Y X, Chang Y F & Xiang Q L, A Leaf recognition algorithm for plant classification using probabilistic neural network, *Comp Sci Artif Intell*, (2017) 1–6.
- 8 Barbedo J G A, Digital image processing techniques for detecting, quantifying and classifying plant diseases, *Springer Plus*, **2**(1) (2013) 660–672.
- 9 Dubey S R, Dixit P, Singh N & Gupta J P, Infected Fruit Part Detection using K-Means Clustering Segmentation Technique, *Int J Interact Multimed Artif Intell*, **2**(2) (2013) 65–72.
- 10 Fenyvesi L, Csatar A & Fenyvesi D, Stress analysis in fruits, *Hindawi Publ Corp*, (2013) 1–6.
- 11 Amara J, Bouaziz B & Algergawy A, A deep learning-based approach for banana leaf diseases classification, *Bus Tech Web*, (2017) 79–88.
- 12 Santosh T & Manjunath C R, A comparative study on agricultural crop disease detection system, *Int J Trends Sci Res Dev*, **2**(4) (2018) 163–168.
- 13 Shah J, Prajapati H B & Dabhi V, A survey on detection and classification of rice plant diseases, *IEEE Int Confon Curr Trends Adv Comput*, 10–11 March 2016.
- 14 Liu L & Zhou G, Extraction of the rice leaf disease image based on bp neural network, *Int Conf on Comput Intell Softw Engg*, 11–13 December 2009.
- 15 Phadikar S, Sil J & Das A, Rice diseases classification using feature selection and rule generation techniques, *Comp Electron in Agric*, **90** (2013) 76–85.