Leaf Disease Detection Based on Local Gabor Binary Pattern Histogram Sequence and Neural Network

Nagamani H S, Saroja Devi H.

Abstract: ***Agriculture forms the main source of food in India, especially in the southern area. The economy of India directly depends on agriculture plants. But due to some major diseases such as blast, brown spot, and bacterial blight, there is a reduction in plant growth which greatly affects agricultural productivity. The farmers add irrelevant pesticides with their limited knowledge which will degrade the quality of the crop but also degrade the soil quality. In the proposed method Machine Vision techniques based on neural networks are used to detect plant health or diseases indicated by leaf anomaly. Image processing algorithms such as K means clustering is used to segment affected areas. From the segmented images of the plant leaf, features are extracted using Color Coherence Vector (CCV) and Local Gabor Binary Pattern Histogram Sequence (LGBPHS). The extracted features are fed as input to a backpropagation neural network to classify the unhealthy leaf.***

Keywords**:** ***Plant health, Color Coherence Vector (CCV), Local Gabor Binary Pattern Histogram Sequence (LGBPHS).***

I. INTRODUCTION

Plants and fruits are essential as they are the primary source of energy for humans and animals. These are useful in many ways due to their medicinal values. Almost 50% population of countries such as Asia and Africa depends on agriculture production. Each year 30 to 40% of plants are lost due to the production chain. Crop diseases create a drop in economic impact. Crop failure due to different diseases causes starvation. Hence it is required to evaluate the plant quality and its health.

Plant disease can be identified potentially early so that it helps in establishing enhancement towards economical, biological, sociological and ecological losses. Plant diseases can be broadly classified as either infectious or non-infectious. Superior level of safety and quality in agricultural products is of high concern. The foundation of quality assessment is basically dependent upon features of leaves such as its texture, cracks, surface, and appearance. Fig.1 demonstrates a general method of categorization of plant diseases.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author

Nagamani HS*, Department of Computer Science, SMVHD Central Institute of Home Science, Bengaluru, India. Email: nagamani.hs@gmail.com

Dr.Saroja Devi H., Department of Computer Science & Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India. Email: sarojadevi.n@nmit.ac.in

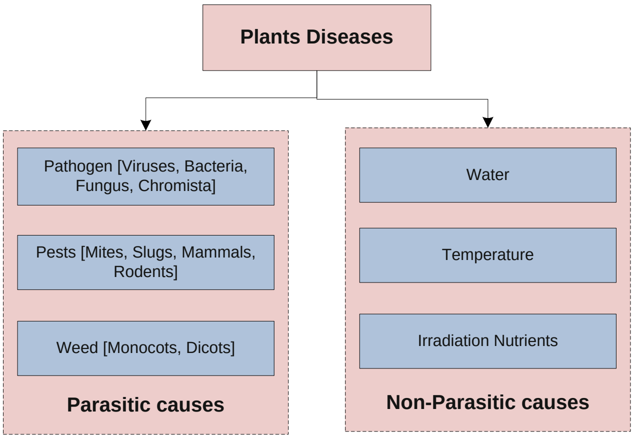

Fig.1. Method of Categorizing Plant Diseases

II. LITERATURE SURVEY

Gittaly Dhingra et.al [1] proposed a comprehensive discussion on plant disease detection and recognition using Machine Vision. This discussion highlights the methods and concepts used by various researchers to identify and classify diseases and challenging issues. The techniques in this are partitioned into detection and classification. The primary aim here is to limit the impact of the diseases on agricultural-related production using machine learning and vision concepts. The survey extends to further research as an automatic estimation.

A. D. Nidhis et.al [2] considered rice as the plant to study since it is the main food in India, especially in southern parts. The diseases that occur in rice plant are brown spot, rice blast and bacterial blight. To avoid these diseases, farmers use some irrelevant chemicals and pesticides, which are not appropriate to cure the plant diseases due to their limited information on plant disease. Further, this pesticide makes the crop quality poor and also soil quality degradation. It is possible to detect the type of disease affected to a plant and to calculate the severity of disease using a Machine Vision technique. The infected region is segmented and its features are extracted. The extracted features are taken as input to the classification algorithm to classify three different diseases.

S. Kalaivani et.al [3] performed an assessment of the significance of plant disease diagnosis at an earlier stage and the use of the technical implementation in the agriculture field. Machine Vision helps to analyze and predict disease at its initial stage based on leaf image. The pre-processing step in Machine Vision analyzes each image pixel and successfully detects the affected regions of the diseased leaf by segmentation technique.

The affected region is segmented by computing each pixel to the maximum histogram values and measuring the similarity of the affected pixel region using dice similarity metrics. Segmentation based on histogram intensity and similarity metrics gives 98.79% accuracy compared to the existing technique.

Nazish et.al. [4] proposed a Machine Vision based algorithm to detect the region of interest (ROI) in plant leaf to distinctly recognize the botanical disease. The curvelet transform is adopted to compute the similarity in leaf images and support vector machine (SVM) classifier to classify the disease with better results. K-means algorithm segments the leaf disease and curvelet transform extract the features. The accuracy obtained using curvelet and SVM is 98.5%. The future scope is to consider three types of diseases for a specific plant leaf disease.

Gittaly Dhingra et.al. [5] introduced a fuzzy set neutrosophic logic segmentation algorithm to detect and find the region of interest. The segmented neutrosophic image is further classified by using three elements such as intermediate, true and false regions. Features of the segmented regions such as color, texture, and histogram-based features are extracted to evaluate the affected region or healthy region of the leaf. Experimental demonstration is validated by considering 400 cases of which 200 are healthy leaf and 200 with the disease. The classification accuracy obtained with the Random Forest (RF) method is 98.4%.

Peng Jiang et.al [6] proposed a deep learning technique, which is based on improved Convolutional Neural Networks (CNNs) for detecting tomato leaf diseases in real-time applications. The complex laboratory tomato dataset is constructed by annotation and augmentation. A new CNN model named INAR-SSD is developed by introducing the GoogLeNet inception combining with a method termed Rainbow concatenation to detect multi-scale diseases and to enhance small disease detection performances. CNN implemented in the Caffe framework on GPU to train with a dataset of 26,377 unhealthy leaves' images. The values of detection performance and speed reaches are 78.80% and 23.13 FPS respectively.

Tanmoy Bera et.al [7] analyzed a variety of rice plant leaf diseases using various Machine Vision and data mining techniques. The Machine Vision concept helps to detect and analyze the disease region in rice plant leaf and to collect similarly affected region related features. Data mining algorithms, on the other hand, extract relevant defected information, which is useful for detecting disease. A comparative study based on feature selection, segmentation, and classification is done. These algorithms can be improved further to provide increased accuracy in rice plant disease detection.

The study in [8] carries out an implementation survey of several leaf disease detection algorithms and compares the performance. For image segmentation, a genetic algorithm is used in this work.

III. METHODOLOGY

This section provides details of the architecture, the process adopted, and the algorithms involved. The overall system architecture is depicted in fig. 2 where the working of the system is shown divided into two phases: 1) training and 2) testing phase. In the training phase, all features belonging to the disease region from the segmented image are extracted and stored in the knowledge base. Further, these extracted features are trained using Back Propagation Neural Network (BPNN). In the testing phase firstly the input leaf images are fed to the pre-processing technique to obtain an enhanced image. The enhanced image is segmented using K-means clustering to obtain a segmented image. The segmented image thus obtained consists of either a healthy region or disease-affected region of the leaf. Features of the segmented image are extracted using descriptors such as color coherence vector and Local Gabor Binary Pattern Histogram Sequence (LGBPHS) [11]. The resultant feature vectors are classified using BPNN to recognize the leaf disease.

A. Pre-Processing

In the pre-processing step, the quality appearance of the input image is enhanced. Generally, the image captured from the camera is subjected to noise due to climatic condition which leads to low contrast. To analyze and segment the diseased region in leaf image good quality of the image is required without any noise. Hence the input leaf image is first enhanced by improving or by adjusting the contrast of the image.

B. K-means Clustering

After increasing the quality of the image, the K-means clustering algorithm is adopted to generate clusters. Using the k-means algorithm, the image is segmented into k number of clusters. The clustering technique divides the set of data points into a specific group. It classifies the data points into k number of disjoint clusters. K-means works in two phases. In the first phase, k centroids are computed. In the second phase, each data point is taken to the cluster which is having the nearest centroid from the respective data point. To find the nearest centroid from the data point distance calculation methods such as Euclidean distance method is used. Once clustering is done a new centroid is recomputed. Using this new centroid a new Euclidean distance is generated between the cluster center and data point. The data points are assigned to the cluster having a minimum Euclidean distance, each centroid cluster represents data point where the sums of distances from all data points in that cluster are minimized. Thus K-means iterative algorithm minimizes the sum of distances from each data point to its cluster centroid. Consider for an image resolution $\mathrm{x\,x\,y}$, suppose image is required to cluster into k number. Let $\mathrm{p(x,y)}$ input pixels or data points required to cluster and $\mathrm{c_k}$defines cluster centers [9]. The steps involved in the k-means algorithm are given below.

Initialize number of cluster k and center $\mathrm{c_k}$.

2. Find Euclidean distance from the cluster center to each pixel of an image using the equation,

$$\mathrm{d} = \|\mathrm{p(x,y)} - \mathrm{c_k}\|$$

3. Assign pixels to the cluster which is having the nearest center using distance $'\mathrm{d}'$.

4. Once all pixels are assigned to a cluster, a new center is recalculated using the equation,

$$c_k = \frac{1}{k}\sum_{y\in c_k}\sum_{x\in c_k} p\,(x,y)$$

5. Continue the above steps until the algorithm satisfies the error value.

6. Reshape pixels value of each cluster into the image.

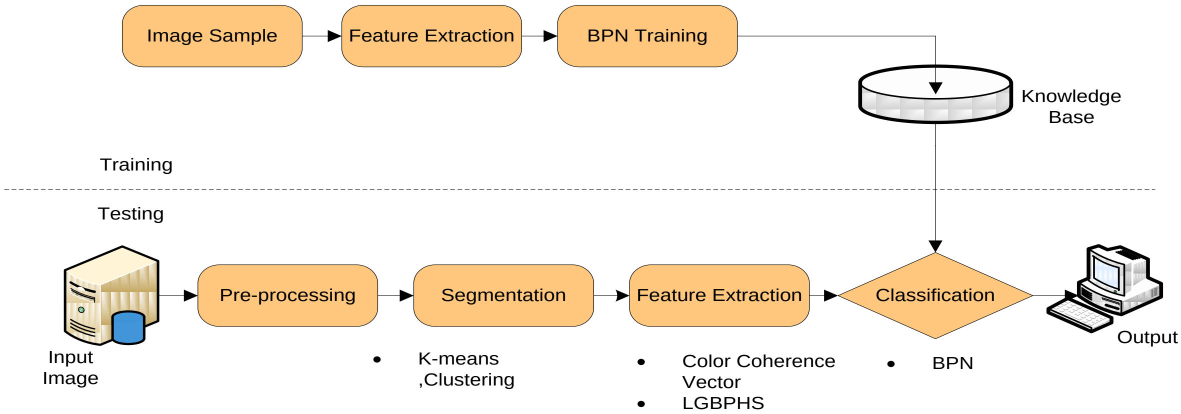

Fig. 2. Overall System Architecture

C. Color Coherence Vector

The Color coherence vector (CCV) method [10] is the extension and enhanced method of color histogram method. The CCV method considers the spatial information of image pixels having the same color coherence region. The CCV effectiveness can be improved based on the number of colored regions sized which helps to discriminate images. A CCV scheme is divided each pixel bin into coherent and non-coherent pixels. A pixel in a bin is said to be coherent if it is part of a large similarity colored region. By default 8-neighborhood connected component is used to extract connected regions of the same color for two-dimensional images. For three dimensional images, the connectivity of components is analyzed using 26 neighborhoods. Pixels regions whose size is more than a threshold (1% of image size) are counted as coherent pixels, and those are less are counted as non-coherent regions. Color histograms and color moments have lack information about the spatial distribution of colors. So CCV is proposed to incorporate spatial information into color histogram representations.

D. Local Gabor Binary Pattern Histogram Sequence (LGBPHS)

LGBPHS [11] is robust to various illuminations in image conditions. It representations a multi-resolution spatial histogram, which combines local intensity distribution with the spatial information. Thus LGBPHS is robust to noise and local image transformations such as illumination, pose, and occlusion. To determine the spatial histogram information, in this method multi-scale and multi-orientation Gabor filters are used, which is followed by the Local Binary Patterns (LBP) operator. This combination enhances the spatial histogram sequence representation sufficiently. To develop LGBPHS training stage is not necessary. Histogram intersection exploits the different LGBPHSes and the nearest neighborhood used for final classification.

Gabor Filters

Gabor filters are also called Gabor wavelets or kernels for feature extraction and recognition [12]. Gabor filter extracts multi-resolution spatial local features. 2D Gabor filters in the spatial domain are defined in Eq. (1), (2) and (3) respectively.

$$\psi_{u,v}(x,y)= \frac{f^2{}_u}{\pi\kappa\eta} e^{-\left(\left(\frac{f^2{}_u}{\kappa^2}\right)x'^2+\left(\frac{f^2{}_u}{\eta^2}\right)y'^2\right)} e^{j2\pi f_u x'} \ \dots (1)$$

$$x' = x\cos\theta_v + y\sin\theta_v (2)$$

$$y' = -x\sin\theta_v + y\cos\theta_v (3)$$

where $\theta_v = v\pi/8$ and $f_u = \frac{f_{max}}{2^{(\frac{u}{2})}}$

Gabor filer represents Gaussian kernel function by modulating with center frequency f_u and orientation θ_v respectively. The ratio between the center frequency and Gaussian envelope size is determined by κ, η. Gabor filters construct a filter bank featuring filters with five scales and eight orientations. Gabor filter with its real and imaginary parts is depicted in Fig. 3. The real parts of the Gabor filter bank are used for feature extraction which consists of 40 filters. The input segmented image is convolved with this 40 Gabor filters to generate Gabor features. Since imaginary or phase information is time-varying only its magnitude information is used to extract features. Hence at each pixel, one Gabor magnitude value is computed for each Gabor filter resulting in 40 Gabor Magnitude Pictures (GMP).

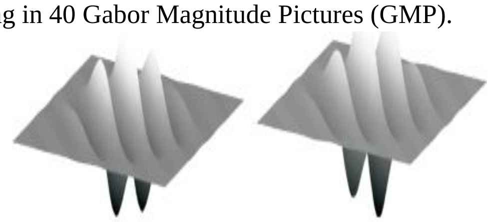

Fig. 3. (a) Real Part (b) Imaginary Part Gabor Filter

Local Gabor Binary Pattern (LGBP)

While Gabor filters can be used for image texture feature extraction and thereby adopted for classification [12], the GMP information can also be further enhanced by encoding the magnitude values with the LBP descriptor. The LBP operator labels the pixels of an image by thresholding the 3x3 neighborhood of each pixel f_P with center value f_c and considering the result as a binary number. The LGBP descriptor denotes the LBP operates on GMP.

$$S(f_P - f_c) = \begin{cases} 1 & f_P \geq f_C \\ 0 & f_P < f_C \end{cases} \quad (4)$$

LGBP Histogram Sequence

Local histogram features are used to represent the regions in terms of the LGBP patterns.

Each LGBP Map is spatially divided into multiple non-overlapping regions. Then, Histogram from each region is obtained. Finally, all the histograms features are computed from the regions of all the LGBP Maps and then concatenated into a single histogram sequence to represent the given image. Flowchart of the LGBP based feature extraction is shown in fig. 4. It produces an LGBP Histogram sequence (LGBPHS).

E. Back Propagation Neural Network

The extracted LGBP histogram sequence-based features are fed to a pre-trained neural network for classifying the leaf diseases automatically. The neural network model used here is a radial basis model. Since its precise classification results in many real-time applications, a neural network is used as a classification method. The processes of preparation and testing are essential steps in implementing a particular model. The database needs two-stage training and testing processes.

1. Training Phase: The features used for training the NN model called training features.

2. Testing Phase: The features used to classify the test data using the trained NN model.

Before the database is applied for ANN training, the network should be designed properly, such as the network type and training method. The ANN configuration is shown in fig.5. This set up was done in the network training phase, where the network is trained using the feed-forward backpropagation network (BPNN). In the training phase, weights are updated until they reached the defined iteration number or acceptable tolerance error. Hence, the capability of the ANN model to respond accurately is assured by the Mean Square Error (MSE) criterion to emphasize the model validity between the target and the network output. The ANN training performance is evaluated and is as shown in fig.6.

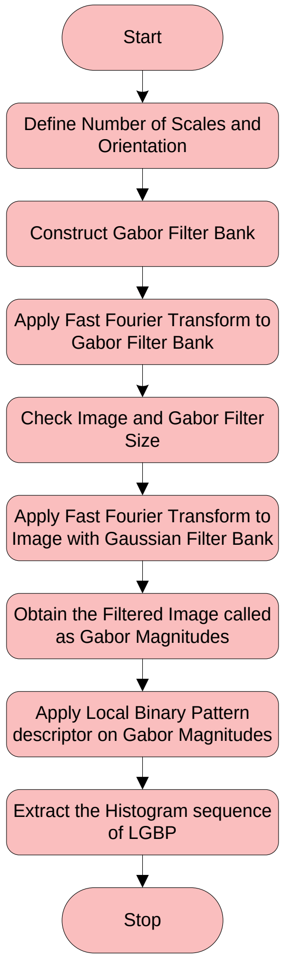

Fig. 4. LGBPHS: Flowchart

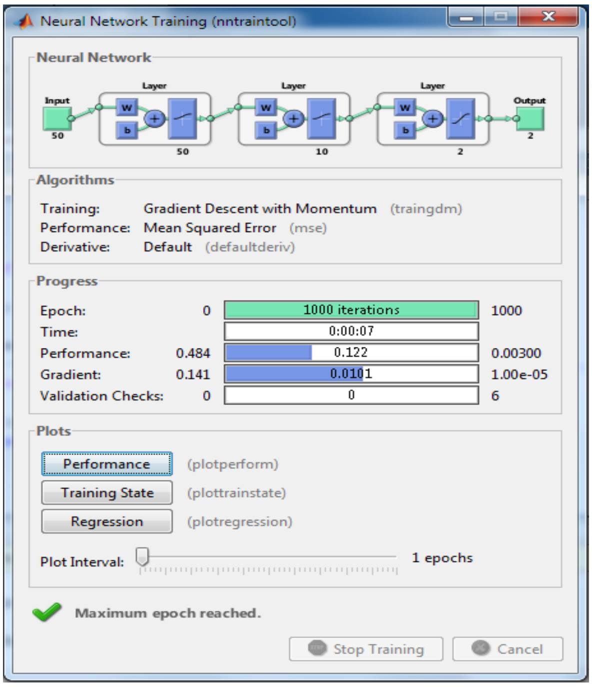

Fig. 5. ANN Design implementation with configuration setup

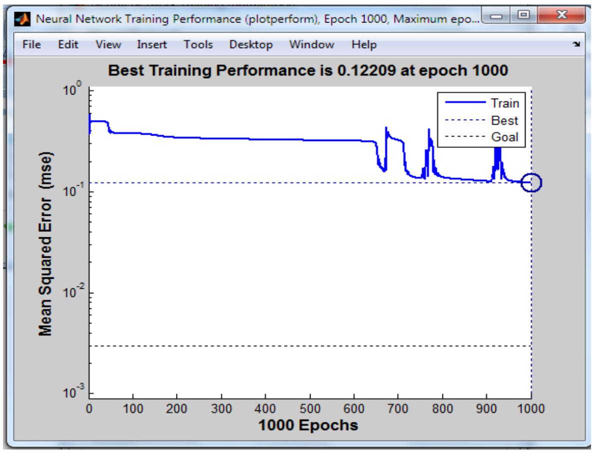

Fig. 6. ANN training Performance

IV. EXPERIMENTAL RESULTS

All the experiments are performed in MATLAB.

For input data, samples of plant leaves like tomato leaves are taken to recognize whether the plant is healthy or unhealthy. The total number of images used for the testing phase is 200, out of which 100 images are healthy and another 100 are unhealthy. Experiments include both training and testing. The ANN training performance is shown in fig. 6.

Two sample input images are shown in Fig. 7 (a) and (d) are taken for which the analysis is carried out. Fig.7 (b) and 7(e) show the output segmented images using K-means clustering. The extracted features of the segmented image are then trained and classified into the healthy or unhealthy leaf by ANN classifier shown in fig. 7 (c) and (f).

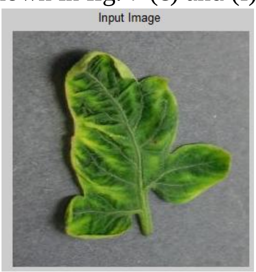

Fig. 7(a). Sample Input Image 1

Fig. 7(b). K Means Clustering Result

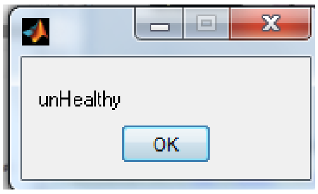

Fig. 7(c). ANN Output

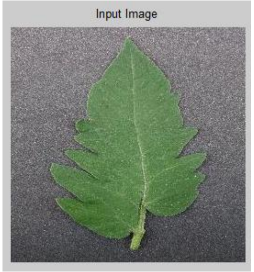

Fig. 8(d). Sample Input Image 2

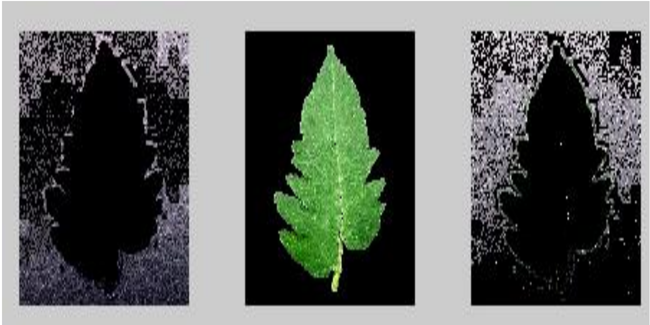

Fig. 7(e). K Means Clustering Result

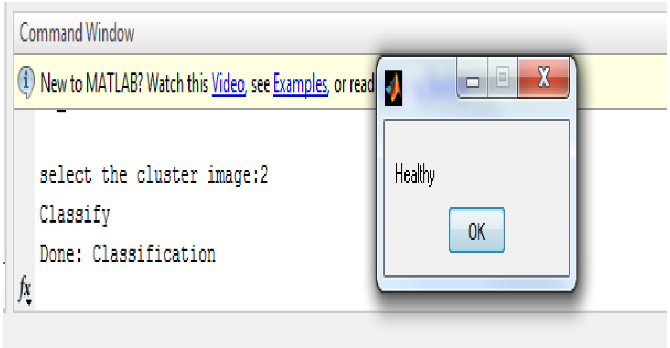

Fig. 7(f). ANN Output

The confusion matrix based evaluation of the proposed system is as shown in table I.

Table- I: Confusion Matrix

	UnHealthy	Healthy
UnHealthy	89 (TP)	11 (FP)
Healthy	8 (FN)	92 (TN)

The confusion matrix parameters are shown in table I include, TP representing True Positive, FP representing False positive, FN representing False-negative and TN representing True Negative. Various formulae for calculating the evaluation metrics are as below.

Accuracy = (TP+TN) / (Total Population) .... (5)
Calculated Accuracy is 90.5%.
Sensitivity = (TP) / (TP + FN) .... (6)
Calculated Sensitivity is 91.75%.
Specificity = (TN) / (TN + FP) .... (7)
Calculated Sensitivity is 89.32%.
Sensitivity = (TP) / (TP + FP) .... (8)
Calculated Sensitivity is 89%.

V. CONCLUSION

In the proposed work, Machine Vision algorithms and Artificial Neural network-based approaches have been applied for plant leaf disease detection. LGBP method is used for extracting a set of color and texture features of leaf diseases. The extracted set of features has been used as an input to train a feed-forward back propagation neural network and detection of leaf diseases. Based on the proposed approach, an efficient, automatic, and reliable system can be developed for the detection and classification of plant diseases. The achieved accuracy is 90.5%. The future work includes a detailed study concerned with color and texture features of leaves, employing other segmentation and classification techniques with an objective to enhance the efficacy of the proposed method.

ACKNOWLEDGMENT

The present research work is carried out as part of the Ph.D. work under the University of Mysore, Karnataka in the research center of Nitte Meenakshi Institute of Technology (NMIT), Bengaluru. The authors are thankful to the Director and Principal of NMIT, and also to the CSE department staff, for the support they have extended. The authors are obliged to the University of Mysore to have provided an opportunity to conduct this research. The authors are grateful to the doctoral committee members for their valuable suggestions and encouragement.

REFERENCES

1. Gittaly Dhingra, Vinay Kumar, and Hem Dutt Joshi, "Study of Digital Image Processing Techniques for Leaf Disease Detection and Classification", Multimedia Tools and Applications, Vol. 7, No. 15, 2018, pp. 19951-20000.
2. A. D. Nidhis, Chandrapati Naga, Venkata Pardhu, K. Charishma Reddy and K. Deepa, "Cluster-Based Paddy Leaf Disease Detection, Classification and Diagnosis in Crop Health Monitoring Unit", Computer-Aided Intervention and Diagnostics in Clinical and Medical Images, Springer, Cham, 2019, pp. 281-291.
3. S.Kalaivani, S. P. Shantharajah, and T. Padma, "Agricultural Leaf Blight Disease Segmentation using Indices Based Histogram Intensity Segmentation Approach", Multimedia Tools and Applications, 2019, pp. 1-15.
4. Nazish T., Abdul Latif Memon, Faheem Yar Khuhawar, and Ghulam Mustafa Abro, "Detection of Infected Leaves and Botanical Diseases using Curvelet Transform", International Journal of Advanced Computer Science and Applications, Vol. 10, No. 1, 2019, pp. 516-520.
5. Gittaly Dhingra, Vinay Kumar, and Hem Dutt Joshi, "A Novel Computer Vision-Based Neutrosophic Approach for Leaf Disease Identification and Classification", Measurement, Vol. 135, 2019, pp. 782-794.
6. Peng Jiang, Yuehan Chen, Bin Liu, Dongjian He, and Chunquan Liang, "Real-Time Detection of Tomato Leaf Diseases using Deep Learning Approach Based On Improved Convolutional Neural Networks", IEEE Access, Vol. 7, 2019, pp. 59069-59080.
7. Tanmoy Bera, Ankur Das, Jaya Sil, and Asit K. Das, "A Survey on Rice Plant Disease Identification using Image Processing and Data Mining Techniques", In Emerging Technologies in Data Mining and Information Security, Springer, Singapore, 2019, pp. 365-376.
8. Vijai Singh and A.K. Misra, "Detection of Plant Leaf Diseases using Image Segmentation and Soft Computing Techniques", Information Processing in Agriculture, Elsevier, Vol. 4, No. 2017 pp. 141-49.
9. Nameirakpam Dhanachandra, Khumanthem Manglem, and Yambem Jina Chanu, "Image Segmentation using K -means Clustering Algorithm and Subtractive Clustering Algorithm", Procedia Computer Science, Elsevier, Vol. 54, 2015.
10. Alaa Al-Hamami and Hisham Al-Rashdan, "Improving the Effectiveness of the Color Coherence Vector", The International Arab Journal of Information Technology, Vol. 7, No. 3, July 2010, pp. 324-332.
11. Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model For Face Representation and Recognition", IEEE International Conference On Computer Vision, Vol. 1, 2005, pp. 786-791.
12. Vishakha Metre and Jayshree Ghorpade, "An Overview of the Research on Texture Based Plant Leaf Classification", International Journal of Computer Science and Network, Vol 2, Issue 3, 2013.

AUTHORS PROFILE

Nagamani. H. S. received an MS in the computer system from the University of Mysore. M Phil in computer science and Engineering from Annamalai University, TN. Presently she is working as an assistant professor in the Department of Computer science, Smt. VHD Cultural Institute of Home science –Autonomous Bengaluru. Her academic experience is 16 years. She has published her research work in many journals and attended Seminars and workshops.

Dr. Saroja Devi H. is a Professor at Nitte Meenakshi Institute of Technology, Bengaluru. Her recent research interests include Machine Vision and Image processing. She has authored more than a hundred research papers in international journals and conferences. She is a recipient of a few national and international awards. She is a Ph.D. graduate from the Indian Institute of Science, Bengaluru.