Plant Leaf Disease Detection using Machine Learning

*

1st Amrita S.Tulshan
dept. of Computer Engineerng
Sardar Patel Institute of Technology
Mumbai, India
amrita.tulshan@spit.ac.in

2nd Nataasha Raul
dept. of Computer Engineering
Sardar Patel Institute of Technology
Mumbai, India
nataasharaul@spit.ac.in

***Abstract*—Every other field has got some benefit from new technologies as compared to the agricultural field. According to past studies, 42% of agricultural production is in loss and that too only because of the increasing rate of loss due to plant leaf diseases. To overcome this major issue, this plant leaf disease detection technique can be applied to detect a disease from the input images. This process involved steps like image pre-processing, image segmentation, feature extraction. Furthur K Nearest Neighbor (KNN) classification is applied on the outcome of these three stages. Proposed implementation has shown 98.56% of accuracy in predicting plant leaf diseases. It also presents other information regarding a plant leaf disease that is Affected Area, Disease Name, Total Accuracy, Sensitivity and Elapsed Time**

***Index Terms*—Image Segmentation, Machine Learning, Plant Leaf Disease Detection**

I. Introduction

With the advancement of new advances, the field of agriculture becomes more prominent as it not only used as food feeding to major population but also used in many applications. Plants are very essential in our life as they provide source of energy and overcome the issue of global warming. Plants nowadays are affected by many diseases such as they cause devastating economic, social and ecological losses and many more. Hence, it is most important to identify plants disease in an accurate and timely way. Plant diseases can be extensively grouped by the idea of their essential causal operator, either irresistible or non infectious.

Digital image processing tools are employed by the used method to obtain the desired output. It is not possible for a human eye to identify the disease extent accurately, as the resultants are subjective in nature. The observations done by the naked eye are usually used to decide diseases severity in the area of production. The significant development has done by the image processing in the field of agriculture. For the identification of the fungi disease, several neural network techniques have been utilized such as Back Propagation, Principal Component Analysis (PCA).To detect plant leaf disease by improving required rate in classification technique. Till now linear SVM is used which is a multi-class classification that only classify the data into two classes which is very inefficient and reduce accuracy of classification.

The main objective that were focused in this paper is to study different types of diseases which are found in plant leaf and also to study and analyze different techniques for plant leaf disease detection using image processing technique and main goal was to propose improvement in existing classification techniques for plant leaf disease detection using machine learning.

In our work, other classifier such as KNN will be used to classify data more than two classes. Total seven diseases were detected by our system. Dataset was of 75 images. More additionally this system detects disease name as well as accuracy percentage of affected area, sensitivity and elapsed time. Diseases that are discussed in this paper are Down Mildew, Early Blight, Mosaic Virus, Leaf Miner, White Fly.

Arrangement of section is as as follows: Section II is literature Survey in which all past study and work is described and white spaces on which the further work can be done. Section III is Methodology which explain the work flow of our approach. Section IV Implementation in which all the phases were explained in detail. Section V Experimentation is detailing about the experiment setup and its result are in Section VI and last section VII which conclude whole paper is conclusion.

II. Literature Survey

This paper present study on various order methods that can be utilized for plant leaf illness characterization. There are such huge numbers of grouping systems, for example, k-Nearest Neighbor Classifier, Probabilistic Neural Network, Genetic Algorithm, Support Vector Machine, and Principal Component Analysis, Artificial neural system, Fuzzy rationale. This paper gives a review of various characterization systems utilized for plant leaf malady order[1]. The created preparing plan comprises of four principle steps, initial a shading change structure for the info RGB picture is made, this RGB is changed over to HSI on the grounds that RGB is for shading

Identify applicable funding agency here. If none, delete this.

age and his for shading descriptor, at long last the surface insights is figured from SGDM lattices.[2] This paper tends to this issue with the target of creating picture preparing calculations that can perceive issues in yields from pictures, in light of shading, surface and shape to naturally recognize sicknesses or different conditions that may influence harvests and give the quick and exact answers for the rancher with the assistance of SMS. [3]. In this paper, they address an exhaustive report on diseases acknowledgment and grouping of plant leafs utilizing picture preparing techniques. In this manner, picture getting ready has been associated for the affirmation of plant illnesses. The paper has been confined into two standard classes viz. recognizable proof and gathering of leafs[4].

The present work proposes a procedure for recognizing plant sicknesses early and precisely, utilizing different picture handling strategies and fake neural system (ANN). The present work is planned to build up a basic infection recognition framework for plant illnesses. The work starts with catching the pictures. Separated and divided utilizing Gabor channel. Trial results demonstrated that order execution by ANN taking list of capabilities is better with a precision of 91%.[5] While the components of the component data were not diminished by using principal component analysis(PCA), the perfect affirmation results for grape ailments were gotten as the fitting precision and the desire accuracy were both 100%, and that for wheat diseases were procured as the fitting accuracy and the figure precision were both 100%. While the components of the component data were lessened by using PCA, the perfect affirmation result for grape ailments was gained as the fitting accuracy was 100% and the desire precision was 97.14%, and that for wheat sicknesses was gotten as the fitting exactness and the conjecture accuracy were both 100%.[6] Test results demonstrate that the course of action execution by SVM is better than that of neural frameworks. Affirmation right rate of cucumber disease reliant on SVM of shape and surface segment is better than that of simply using the shape highlight.[7] It is to demonstrate a brilliant system for stopping region distinguishing proof subject to picture dealing with. This system works day and night. There are various troubles in the midst of day and night time like lighting conditions, shadow impacts, between thing obstacle, etc.[8]

At first, an overall ulcer sore descriptor is used to distinguish citrus unfortunate sore from leaf-establishment. Thirdly, a two-level different leveled area structure is made to recognize the contamination damage and AdaBoost is grasped in feature assurance and classifier learning[9]. The guideline focus of this investigation is to develop a model system for diagnosing paddy contaminations, which are Blast Disease (BD), Brown-Spot Disease (BSD), and Narrow Brown-Spot Disease (NBSD). The strategy incorporates picture acquisition, changing over the RGB pictures into a twofold picture using modified thresholding reliant on adjacent entropy farthest point and Otsu method. Along these lines, by using age rule technique, the paddy diseases are seen about 94.7 percent of accuracy rates[10]. The illnesses considered are Powdery Mildew, Downey Mildew which can make overpowering setback Grape natural item. For recognizing verification of contamination features of leaf, for instance, genuine turn, minor center point, etc are expelled from leaf and given to classifier for arrangement.[11] The model using 14 picked HSI surface features achieved the best gathering precision (96.7%), which prescribed that it is perfect to use a diminished shade, submersion and power surface rundown of abilities to[12] separate citrus strip afflictions. Typical plan precision and standard deviation were 96.0% and 2.3%, independently, for a robustness preliminary of the gathering model, demonstrating that the model is vivacious for requesting new natural item tests as shown by their strip conditions[13].

The identification of the fungi is done from their morphology, due to the effects done on the reproductive structure directly[14]. Nowadays, there are many technologies that are utilized to detect plant disease at the remote locations. K-means division calculation and SVM Classifier are utilized for the integration purpose of disease detection technique in plants. The precautions measures are proposed by the Agri-Guide v1.0, it also provides treatment to plants with the response time 1ms. 83% of accuracy precision is provided by the Agri- manage v1.0 and 73% accuracy to the unknown patterns of monocot and dicot plant diseases. [15] Some where it has technique that is utilized to control the quality of the fruits and vegetables and specifying their botanical tissues. This proposed technique is Electrical Impedance Spectroscopy that is used to study the structure of botanical tissues and their electrical properties.Proposed in this paper that [16].

Some important factors come out at the time of survey the need of simple plant leaves disease detection system that would facilitate advancements in agribusiness. There are various techniques that are utilized to detect plant disease such as BPNN, SVM, K-implies bunching, and SGDM. Proposed in this paper that [17], every original image are a RGB image as it includes the mixture of primary colours such as red, green and blue. Due to enhancement in the image, enhanced image consists of high quality and clarity as compare to original image. The term disease is generally utilized just for pulverization of live plants. Paper demonstrates the important method to develop for detecting the plant disease [18]. The methods studies are for expanding throughput and accurate treatment of plant disease than human expectation. In Color model CIELAB color model is accurately identified disease and results are not affected by background, type of leaf, type of disease spot and camera flash.

[19] SVM classifier that has been used for the identification of plant sicknesses. As in the consequences of one picture ailing locale is 5.56Background and Black pixels are both portioned in starting advance. This examination contains a novel work that is it will ascertain of tainted zone of plants. [20] Next increasingly significant point is avoidance of the plant maladies as it lead to affordable development of the general country, as irresistible sickness corrupts the nature of the nourishment and yield misfortunes in the agrarian field. It expands the amount and the nature of the rural item. In this

paper, leaf pictures are utilized so as to decide the irresistible malady in the plants. ANN strategy has been used so as to characterize the infection in the plants, for example, self-dealing with highlight map, support vector machine, SVMs, etc. Proposed in this paper[21] for the discovery of the rust in the plants an early identification and acknowledgment strategies are examined in this paper. Fluffy c-infers Clustering calculation and infection discovery, acknowledgment of its sort and ID calculation has been created for the extraction of implanted highlights in the wheat leaves. This is done under the light of artificial neural networks (ANN).

III. Methodology

A. Pre-Processing Phase

This procedure is a readiness procedure to get plant leaf pictures. The RGB shading pictures of plant leaf are captured utilizing a digital camera, with pixel goals 568x1020. There have been gathered 75 information tests. It comprises five types of diseases which affected the plants. The model uses matlab picture preparing library plant leaf image is as an input which is furthur converted to gray scale. Block diagram of phases used in proposed work has been shown in Fig 1.

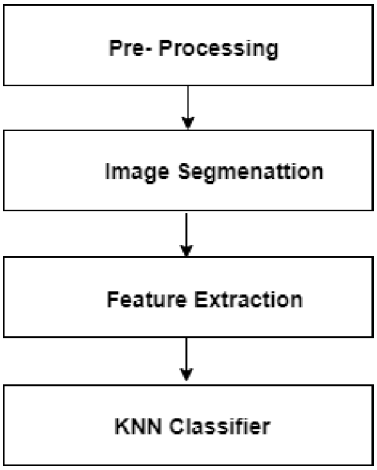

Fig. 1. Block Diagarm of all the steps

B. Image Segmentation

The Image segmentation technique is applied which will segment the image on the basis of their properties. The image segmentation techniques are generally categorized into two types that i.e region based segmentation and threshold based segmentation. The next stage is to extract the valuable sections. Not all portions contain noteworthy amount of data. So the part which contain more than 50% of the data are considered for the further examination.Fig. 1. General Block Diagram to describe different phases used in methodology. In this work, the region based k-mean segmentation technique is applied for the image segmentation because more noise immune, works well in homogeneous region. In image segmentation input will be data set given by us and its output is in form of cluster data.

C. Feature Extraction

Diverse features are picked to depict the distinctive properties of the leaves. A few leaves are with a particular shape, some have distinctive surface examples, and some are described by a mixture of these properties.

1) GLCM Algorithm: This strategy was first proposed by Haralick in 1973 and still is a standout amongst the most well known methods for surface examination [11]. The key idea of this technique is producing highlights dependent on gray level co-occurrence matrices (GLCM). Co-occurrence features are obtained from a gray-level co-occurrence matrix. Only 13 features were used in whole project because at a time GLCM algorithm takes only 13 features. Following is the step by step formula used.

- Count all the number of pixels in the matrix in which the data is saved.
- Store the counted pixels in matrix P[I,j].
- Check similarity between pixels in the matrix by applying histogram technique.
- The elements of g need to be normalized by dividing the pixels.

D. Classifier

In the classification stage, the co-event highlights for the leaves are extracated and contrasted and then compared it similarly to put away in the feature library. There are such so many numbers of classifier however in this paper two classifiers has been talked about that is existing one Support vector machines (SVMs) classifier and proposed is KNN classifier. SVM classifier is a lot of related supervised learning strategies utilized for classifying and reversion. All the more formally, a support vector machine develops a hyperplane or set of hyperplanes in a high-or vast dimensional space, which can be utilized for grouping, reversion, or different undertakings. The K-NN is additionally the classifier of the classification of supervised learning calculation. In supervised learning the objectives are known to us yet the pathway to target isnt known.

To grasp AI closest neighbors structures is the ideal model. Give us a chance to think about that there are numerous groups of named tests. The idea of things of the equivalent recognized bunches or gatherings are of homogeneous nature. Presently if an unlabelled thing should be named under one of the named labels. Presently to order it K-closest neighbors is simple and best calculation that have record of every single accessible class can impeccably put the new thing into the class based on biggest number of vote in favor of k neighbors. Along these lines KNN is one of the substitute to arrange an unlabelled thing into recognized class The precision and proficiency of k-NN calculation essentially assessed by the K esteem decided. Also, its favorable position is that KNN is a fair-minded calculation and have no presumption of the information under thought.

IV. Proposed Work

Fig 2 gives full idea about the proposed work. Proposed work has main three factors on whcih whole work is to be done.

A. Training Dataset

The image is given as info and it gets into the pre-processing phase in which leaf image quality gets improve and it additionally uses to expel unaffected area which is appeared in figure 3 indicates defected leaf. And after that GLCM calculation is connected. It is a strategy, in which the texture of an image is considered and converted into grayscale, and after that, it presents exceptional features. The algorithm of k-mean is connected which is the textural feature based segmentation algorithm. In the k-mean clustering algorithm, the focused point is determined from which the Euclidian separation is determined and the information is clustered by their likeness. Here, everytime partition performed by clusters in that In any case, K-means clustering divides the leaf image into four parts/cluster from which one or more than one parts may contain the disease which indicates that the leaf is infected by multiple diseases. In the wake of entering the group number, we will get bunch having unhealthy part by utilizing KNN classifier. Through KNN classifier we get the disease name as well as its affected area.

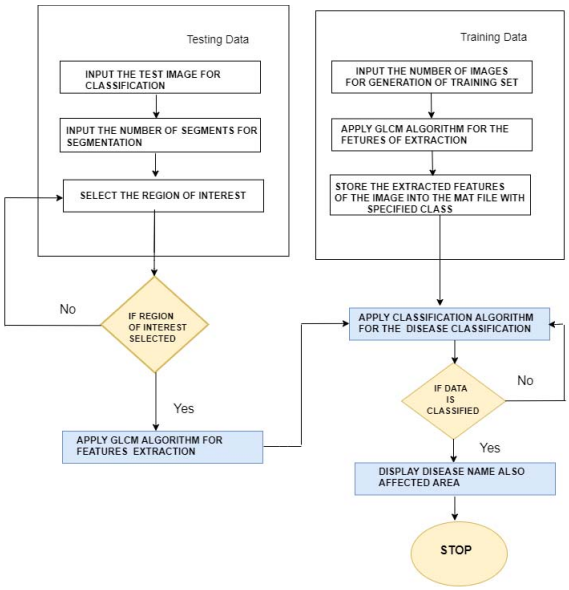

Fig. 2. Proposed Flow Chart

B. Testing Daatset

Here data might be prepared or un- trained. In the event that information is prepared, at that point, after element Extraction, it will legitimately go into the classification. And if it is untrained data then we need to input the number of segments for segmentation. At that point calculation of k-mean clustering which is the textural based calculation. In the k-mean clustering, the focused point is determined from which the Euclidian separation is determined and the information is clustered by their similar properties. We will enter the quantity of the group in which we have ROI (Region of Interest). After that entering cluster number, we will think of definite defected disease name through KNN classifier.

V. Experimentation

All the experiment are performed in MATLAB. For input data disease, sample of plant leaves were total 75 from which 55 data were trained dada and 20 were used for testing data.We tried to run various types of clusters on this leaf out of these number of clusters, cluster 4 was the one which we found the most accurate for detecting the disease of the leaf. For example, In figure 3, a leaf get affected by four different diseases, but final disease from which leaf is suffering is from Mosaic Virus and its affected area is 18.58% which is highest among all other four affected areas. So, from four different affected areas one area percentage was highest and therefore that percentage disease will be the final disease from which our disease get suffered as shown in Fig 3 to Fig 8. Fig 3 shows disease leaf then Fig 4 is a cluster 1 in that percentage affected area of leaf is 15.21%. and disease name is Leaf Miner in the same way Fig 5 is a cluster 2 in that percentage affected area of leaf is 18.48% and disease name is White Flylike ways remaining two clusters Fig 6 is a cluster 3 in that percentage affected area of leaf is 18.58%. And disease name is Mosaic Virus Fig 7 is a cluster 4 in that percentage affected area of leaf is 15.21%. and disease name is Early Blight. Fig 8 shows final disease name which was finalized on the basis of highest percentage area that is Mosaic Virus with 18.58%.

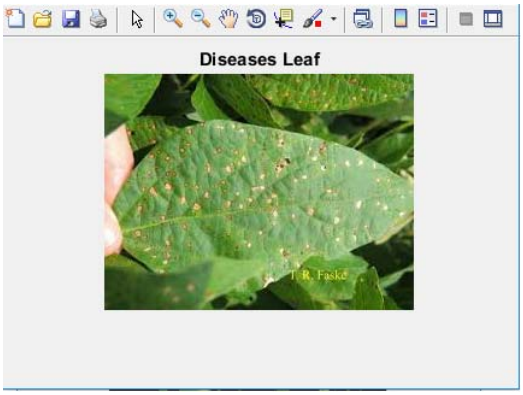

Fig. 3. Input plant disease leaf Image

VI. Results

Diseases that are taken into considerations are Mosaic virus, Powdery Mildew, Downy Mildew, Leaf Miner, White Fly shows images which gives false output and not able to predict correct diseases name. After Doing testing through KNN Classifier it shows exact disease name for any number of leaf diseases. While on the other hand Linear SVM shows error when it has given more than two diseases. It means that Linear SVM is better only for detecting two diseases and not for more

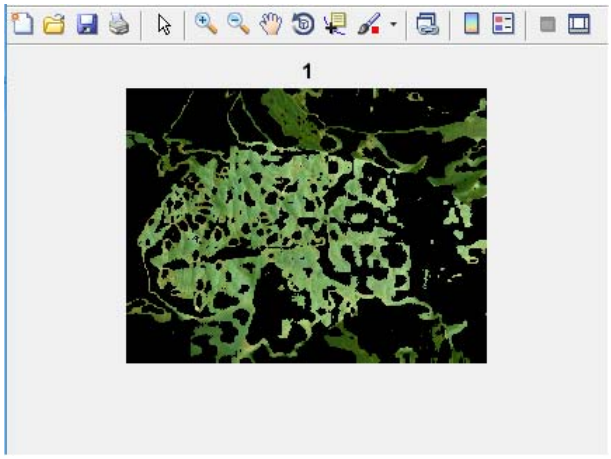

Fig. 4. Cluster 1
Percentage of affected Area: 15.21%
Disease Name Leaf Miner

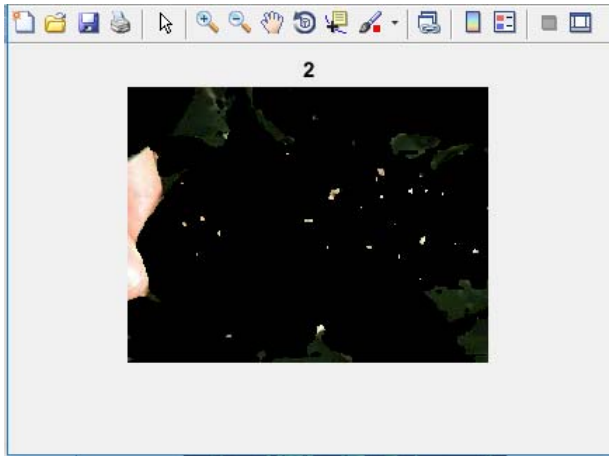

Fig. 5. Cluster 2Percentage of affected Area:18.48Disease Name White Fly

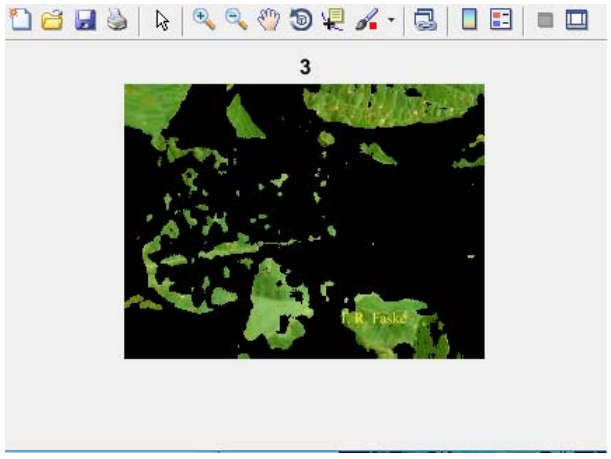

Fig. 6. Cluster 3
Percentage of affected Area:18.58%
Disease Name Mosaic Virus

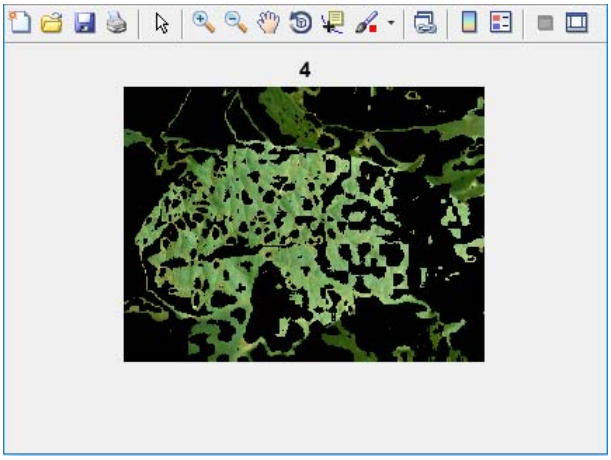

Fig. 7. Cluster 4
Percentage of affected Area:15.21%
Disease Name Early Blight

than that. And KNN classifier can detect number of diseases. Even it shows how much area of leaf is defected by particular disease. From table 1 we can see that the average accuracy of existing technique is 97.6% and that of the proposed technique is 98.56%.

TABLE I
COMPARISON BETWEEN EXISTING ALGORITHM AND PROPOSED ALGORITHM

Algorithm Used	Number of images used for dataset	Name of the Diseases	Detection Accuracy
Linear SVM Classifier(Existing Algorithm)	150	Sun Burn Disease, Fungal Disease, Early Scorch, Bacterial Leaf Spot, Frog eye leaf Spot	97.6%
KNN Classifier(Proposed methodology)	75	Mosaic virus, Powdery Mildew, Downy Mildew, Leaf Miner, White Fly	98.56%

VII. CONCLUSION AND FUTURE WORK

- In this paper, our main goal was to propose improvement in existing classification techniques for plant leaf disease detection using machine learning and that is proved using KNN classifier over SVM classifier.
- Another goal was to study on different leaf diseases which are not yet been studied so the proposed Algorithm was tested on five different diseases which influence on the plants they are: Early Blight, Mosaic Virus, Down Mildew, White Fly, Leaf Miner.
- The experimental results indicate that proposed algorithm accuracy is 98.56% while existing system has 97.6% of accuracy.
- Future work can be done on improving. Future work can be done on more dataset and can work on more accuracy.

REFERENCES

[1] Sanjay N. Ghaiwat, ParulArora, Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review, International Journal of Recent Advances in Engineering and Technology, ISSN(online) :2347-2812, Volume 2 Issues 3 2014

[2] Prof. Sanjay B. Dhaygude and Mr.NitinP.Kumbhar, Agricultural plant Leaf Disease Detection Using Image Processing International Journal of Advanced Research in Electrical,Electronics and Instrumentation Engineering Vol. 2, Issue , January 2013

[3] Mr. Pramod S. landge et.al , Automatic Detection and Classification of Plant Disease through Image Processing, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013 ISSN: 2277 128X

[4] Gittaly Dhingra Vinay Kumar Hem Dutt Joshi Study of digital image processing techniques for leaf disease detection and classification Springer Science+Business Media, LLC, part of Springer Nature 2017

[5] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li, Image Recognition of Plant Diseases Based on Backpropagation Networks, 5th International Congress on Image and Signal Processing (CISP 2012) 2012

[6] PiyushChaudharyAnand K. Chaudhari Dr. A. N. Cheeranand- ShardaGodara Color Transform Based Approach for Disease Spot Detection on Plant Leaf,International Journal of Computer Science and Telecommunications [Volume 3, Issue 6, June 2012

[7] Tian Touwen, Li Tianlai, Niu Yan The Recognition of Cucumber Disease Based on Image Processing and Support Vector Machine, 2008 IEEE

[8] [8] R. Yusnita, FarizaNorbaya, and NorazwinawatiBasharuddin, Intelligent Parking Space Detection System Based on Image Processing,International Journal of Innovation, Management and Technology, Vol. 3, No. 3, June 2012

[9] [9] Min Zhang, Qinggang Meng, Citrus canker detection based on leaf images analysis, 2010 IEEE.

[10] Nunik Noviana Kurniawati, Siti Norul Huda Sheikh Abdullah, Salwani Abdullah, Saad Abdullah Investigation on Image Processing Techniques for Diagnosing Paddy Diseases, 2009 International Conference of Soft Computing and Pattern Recognition.

[11] ABDOLVAHAB EHSANIRAD and SHARATH KUMAR Y. H. Leaf recognition for plant classification using GLCM and PCA methods, Oriental Journal of Computer Science Technology,Vol. 3(1), 31-36 June 04,(2010)

[12] Dae Gwan Kim, Thomas F. Burks, Jianwei Qin, Duke M. Bulanon, Classification of grapefruit peel diseases using color texture feature analysis, Vol. 2 No.3

[13] Akira Hirai, Satoshi Yamane, Masaki Miyazawa, and Kenji Ohshima Application of Multi-layered Fuzzy Inference Based on Back Propagation Method to the Robotic Welding, IEEE 2005

[14] Prof. Sanjay B. Dhaygude, Mr. Nitin P. Kumbhar, Agricultural plant Leaf Disease Detection Using Image Processing, 2013, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 1

[15] Anjali Chandavale, SurajPatil and Ashok Sapkal, Agri-Guide: An Integrated Approach for Plant Disease Precaution, Detection, and Treatment, 2016, Springer Science+Business Media Singapore, 78-981- 10-1678-3

[16] E. Borges, A. P. Matos, J. M. Cardoso, and C. Correia, Early Detection and Monitoring of Plant Diseases by Bioelectric Impedance Spectroscopy, 2012, springer, volume 5, issue 8

[17] Ms. Kiran R. Gavhale, Prof. UjwallaGawande, An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques, 2014, IOSR Journal of Computer Engineering (IOSR-JCE), 2278-8727Volume 16, Issue 1

[18] Deepak J. Dange and Prof. M. A. Sayyad, Computer Vision image Enhancement and Plant Leaves Disease Detection, International Journal of Modern Trends in Engineering and Research (IJMTER), Volume 02, Issue 06

[19] RajleenKaur, Dr. Sandeep Singh Kang, An Enhancement in Classifier Support Vector Machine to Improve Plant Disease Detection, 2015, IEEE

[20] D. Khirade, A. B. Patil, Plant Disease Detection Using Image Processing, 2015, IEEE

[21] DipteshMajumdar, Dipak Kumar Kole, ArunaChakraborty, An Integrated Digital Image Analysis System for Detection, Recognition and Diagnosis of Disease in Wheat Leaves, 2015