JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

ADVANCED ENSEMBLE CLASSIFIER APPROACH FOR TOMATO PLANT DISEASE CLASSIFICATION

Padmaja Goli[#1], S Venkateswarlu [*2]

[#1,2] Computer Science and Engineering, KLEF, Vijayawada, India

[#1]Research Scholar, KLEF, Vijayawada

[1]golipadmajaap@gmail.com

Professor, KLEF, Vijayawada, India

Abstract: Advances in various technologies which have been applied on various sectors implemented in Agriculture sector too. Application of latest technologies includes Major areas like Medicine, Space, Banking, Security Systems along with Agriculture which help to yield quality and quantity enhanced products. Agriculture application of technology helps in early disease detection, classification and prevention that help farmers to protect the crops and enhance quality of the food products being produced. Early detection of diseases that may occur in plants stands as a primary solution to prevent losses in the harvest and amount of agricultural products been grown. The important objective of the proposed work is to develop a method to identify plant diseases and classify the disease using advanced Ensemble classifier. The mixture of multiple features helps to identify and classify diseases in less time reducing Computational time and improve accuracy. The working methodology followed is Segmentation, Feature Extraction, Feature Selection and Classification. The performance of the proposed work has been measured by precision and accuracy.

Keywords: *Ensemble Classifier, Image Segmentation, Feature Extraction, Image Preprocessing, Plant Disease, Segmentation.*

I. Introduction

In numerous developing nations, still agriculture stands as backbone for several people as a livelihood. In India especially Agriculture plays an important part in the profitable growth. Majority of Indian economy lies on agriculture sector. Using of nontraditional procedures and usage of heavy chemicals lead to huge loss in the production of crops and the entire profitability is being affected. One of the sensitive part of the plants mostly affected are leaves. The leaves being vital part of a plant and getting affected with various diseases directly impact the productivity of the plant. These diseased leaves pose several symptoms that are seen in later stages. Any crop's productivity directly depends on the growth of the plants. Hence, it is veritably crucial to identify the any disease affected to the plant at an earlier stage. Once the disease has been detected it is important to take care that it should not be spread to other parts of the plant. In general, farmers keep monitoring the color and shape of the plant's leaves if they are affected with any disease. Most experience and hard work is needed for a farmer to identify the disease occurred for a plant by seeing it's leaves and it is difficult for identifying in larger farms. To identify various infections occurred within the different regions of the plant can be done by signs, spots and color variations in the plant. For enhancement of

Submitted: 21/10/2022 **Accepted**: 30/11/2022

crop production rate automated diagnosis is very essential for the agriculture sector. Various advanced processing tools are being used I now-a-days to deal with the problems arise in the agriculture sector applications. They include the disease detection [1] in leaves, stem and produced fruits. Image processing techniques are being used in identifying and evaluating the diseases detected in the plants. Figure.1 shows the stages that are employed using image processing in detection of disease in the leaf of a plant.

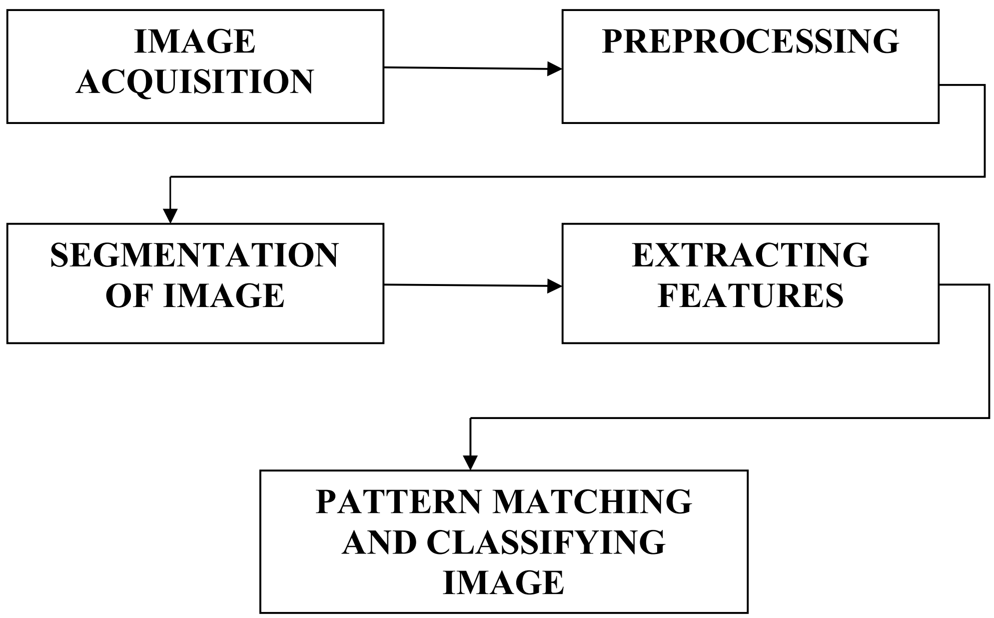

Fig. 1.1 Image Processing Stages for Plant Disease Detection

Functioning of above stages is as described below:

1. **Image Acquisition**: General meaning of Image Acquisition is restoring an image from an existing original image. This process involves generation of a graphical image represented as a physical view or internal structure of an image being used. For restoration of image a hardware source is employed. The processing of source image is done along with another processes and emerging of later processes are needed at a later stage. The image obtained will be a natural image which follows Image Acquisition work flow sequence which is a necessary process in Image Processing.
2. **Image Preprocessing**: Every image acquired from any source will have redundant or noise data which is not required for the system to process the image. Removing the redundancy and noise data helps the image acquire good quality for further processing. Noise filtration, Changing image size dynamically, translation of image and image enhancement are the various methods used in Image Preprocessing stage.
3. **Segmentation**: Division of a digital image into several segments is termed as Image Segmentation. Segmentation main objective is to extract the exact needed object in the preprocessed image. It helps in simplifying the process of inspecting the portion of the image. Image boundaries and objects can be easily located by using Segmentation.

4. **Extracting Features**: This stage main goal is to extract the essential features of an image that obtained region of interest by the earlier stages. Several set of values are extracted from an image using the feature extraction method which are called as features of an image.
5. **Pattern Matching and Classification**: In this Pattern Matching method extracted features from the above layer are matched with existing patterns which have similar features to that of an extracted image. The databases which contain patterns of existing images of diseased plants are compared with the new image taken for pattern matching. Thus matched patterns now undergo classification procedure which is used to identify the pattern and categorize it based on the data obtained after pattern matching. Here supervised and unsupervised classification methods can be used to classify the given patterns.

II. LITERATURE REVIEW

V. Singh, et.al (2015) has introduced a renewed image segmentation algorithm for segmentation of plant images with a main objective of using the algorithm to identify and classify the disorder occurring in the leaves [7]. Several types of classification algorithms were reviewed in the work. All these different algorithms could be applied to detect several types of diseases that exist within the plant leaf. For Image segmentation, Genetic algorithm was utilized in their work which played an important role for detection of diseases within plant leaf.

A. Devaraj, et.al (2019) states farming is not just a technique as it was the main source of food continuous growing population [8]. Around 70% of the whole population of Asian countries depended on agriculture for livelihood. Due to various types of diseases reduced the quality of crop being produced. Effective disease detection helps prevention of losses to farming. The work aimed in developing a software approach for the automated classification and disease detection that includes different phases. Using plant's leafs infections are detected.

R. M. Prakash, et.al (2017) implemented image processing algorithms to detect leaf disorders [9]. Applying image analysis and classification algorithms to detect and classify leaf diseases was the main purpose of this work. There were four stages included in the suggested system. This work made use of K-means approach for segmenting leaf image to identify infectious regions. This work performed the task of texture feature extraction using GLCM features. In this work, the classification task had been performed by SVM classifier.

D. M. Sharath, et.al (2019) stated that plant diseases in farming had become a serious issue as these diseases caused losses in the production. This also affected the quality of the farming produce [10]. It was an extremely complex job to monitor the health of plant and identify different pant infections in manual manner. Expertise in the plant infection detection was required for this purpose. Moreover, this process was very time consuming. Therefore, plant infections were identified using image processing. There was various infection detection stages included in this process. Disease affecting plant was monitored on the basis of output achieved using these stages. In this work, the images of infectious plants were used to discuss the technique implemented for plant infection recognition.

G. K. Sandhu, et.al (2019) reviewed and summarized different plant infection methodologies with the help of image processing. In the last few years, these techniques attracted many researchers for detecting plant diseases [11]. These algorithms classified plant leaves as normal or contaminated. However, this procedure raised different issues. These issues included the computerization of the recognition system by complex pictures clicked in outside lighting and extreme environ circumstances. In this work, it was concluded that these infection recognition methodologies showed competence and accurateness. These techniques could run the system designed to detect plant leaf infections in spite of some limits. Thus, in future, more research work would be carried out in this area to improve the earlier studies.

D. Sachin, et.al (2015) analyzed that the losses in the harvest and quantity of the agricultural goods could be prevented significantly by identifying plant infections [12]. The study of the plant infections represented the study of visibly recognizable patterns within the plant. For sustainable cultivation, it was extremely important to monitor health of the plant and detect disease. However, manual monitoring of plant infection was an extremely complex task. This process was tedious and required proficiency in the plant diseases. Also, this process was very time consuming. Therefore, the plant infections were detected using image processing. These were different stages involved in plant disorder recognition. A discussion was made here for detecting leaf disorders with the help of leaf pictures. Different sorts of algorithmic approaches employed to detect plant disease were also talked about in this work.

R Anand, et.al (2016) introduced a renewed algorithmic approach of plant leaf infection detection. This algorithm carefully detected diseases [13]. In this work, image processing and artificial neural algorithms were used for detecting brinjal leaf infection. Leaf disorders were the major concern in this work. This reduced the produce of brinjal significantly. In this work, instead of whole brinjal plant, focus was on just brinjal plant leaf. Almost all disorder arose on the leaves of eggplant. K-means approach was used for segmenting while NN was used for classification in this work. The leaf disorders were detected efficiently using suggested approach on ANNs.

C. G. Dhaware, et.al (2017) stated that diseases within plant leaves were detected and identified by implementing image processing methodologies [14]. It was advantageous to detect plant leaf infections using some automated technique as it reduced a lot of hard work to observe disease in big agricultural fields. The disease symptoms were detected at early stage using this technique. Different steps were involved in the plant infection recognition and classification. In this work, a discussion was made on different methods used for image pre-processing and image segmentation algorithm for identifying plant disorders in automatic manner. This work was carried out on different plant leaf infection classification techniques applicable to classify leave diseases.

U. M.Korkut, et.al (2018) made use of image processing and machine learning approaches for identifying leaf infections in automatic manner [5]. Early and exactdetection of leaf disorders contributed significantly for crop quality and production. The cost of plant diseases and usage of unnecessary pesticides could be reduced by timely detection and interference. Leaf images of different plant species were collected here. This work implemented TF (Transfer learning) technique for taking out important attributes of the pictures. Accuracy rate of 94 % was obtained by proposed model using different machine learning techniques.

P. B. Padol, et.al (2016) analyzed that infections within plant leaves were detected and classified using a popular approach called image Processing [6]. This work made use of SVM classifier for detecting and classifying disorder in the leaves of grape. Initially, this work implemented K-means approach for the segmentation to detect the infectious region. Later, the extraction of two particular attributes was performed. At last, different types of leaf infections were identified by classification algorithm. Accuracy rate of 88.89% had been attained by the introduced scheme in the classification and detection of diagnosed diseases.

R. Pawar, et.al (2017) stated that image processing was extremely important in plant infection detection or in grading of the quality of the fruit [17]. Losses in the production and the quantity of farming goods were reduced to a large extent by detecting leaf infections. The study of the plant infections represented the study of visibly recognizable patterns on the plant. For sustainable cultivation, it was extremely important to monitor health of the plant and detect disease. Manual monitoring of plant infections was an extremely complex task. In this paper, different Pomegranate plant disease detection techniques were applied using plant leaf pictures.

III. METHODOLOGY

The primary objective of this work is detecting plant diseases from the leaves. The authors proposed an advanced ensemble model based on Random Forest and KNN for detecting plant diseases using its leaves. The test work has been carried out on benchmark dataset of Plant Disease Images[5]. The respective dataset source is kaggle. URL of the dataset source is: https://www.kaggle.com/emmarex/plantdisease. The test dataset consists of one thousand images of Healthy Leaves and three plant diseases i.e. Bacterial spot, Mosaic Virus and Yellow Curl. Figure 2 represents the images of the plant disease extracted from dataset along with a Normal leaf without any disease infected to it.

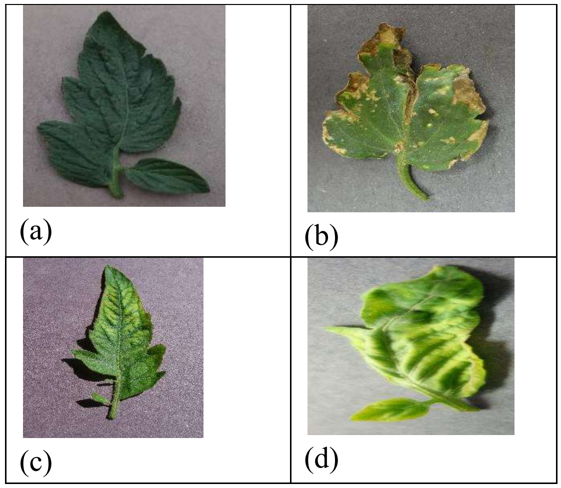

Fig. 2: (a) Normal leaf (b) Bacterial Spot (c) Mosaic Virus (d) Yellow Curl.

Various diseases identified for proposed work:
Bacterial Spots: This disease occurs due to bacterial species Xanthomonas occurring world-wide where tomato plants are producted. It causes leaf and fruit spots leading to defoliation, sun-burnt fruit, and loss in yield. Due to variety of bacterial spot pathogens, Bacterial Spot disease can be found in wide range of temperatures and it is a major drawback to tomato production globally. Fig 2(b) shows a tomato leaf infected with Bacterial Spot disease.

Mosaic Virus: There are several variety of ways by which a tomato plant can be infected with mosaic virus. Majorly, this infection can be occurred due to the remains of virally infected plants which still exist in the same soil where tomato plants are grown. This type of virus can live for a longer period of over 50 years in dead and dried remains of previously infected plants. Tomato leaves with mosaic virus seen in Light and dark green spotted regions on the leaf. Figure 2(c) shows the tomato leaf that is infected with mosaic virus.

Yellow Curl: Tomato Yellow Leaf Curl Virus (TYLCV), which is a type of DNA virus from Geminiviridae family of viruses. This stands as one of the most damaging diseases of tomato plant. Typically it appears in the regions of subtropical and tropical geographical regions. This virus is transmitted through insect belonging to Aleyrodidae family. Figure 2(d) shows a tomato leaf infected with Yellow Curl disease.

The proposed methodology has been presented in the figure 3, and it works in four phases: Preprocessing, Segmentation, Feature Extraction, and Classification by Ensemble Model.

3.1 **Pre-Processing**: In this initial stage the primary aim is to remove noise from the image. This will convert the input leaf picture into the gray scale image. Then the noise from the image will be removed by using filtering method. For removing noise, the Gaussian filter is applied to reduce the noise and technique of the histogram equalizer used to contrast of the input image.

3.2 **Segmentation**: It is the approach which is used to segment the image into several segments. The region-based segmentation method called k-mean clustering is applied for this purpose here.

3.3 **Feature Extraction**: GLCM, a method of textural feature analysis is applied in the current research work. This algorithm will extract around 13 features of the given input images. The extracted features from the source image are: Correlation, Energy, Contrast, Mean, Standard Deviation, Homogeneity, Entropy, RMS, Variance, Smoothness, Kurtosis, IDM and Skewness.

3.4 **Classification**: In this phase an ensemble classification method applied for the disease detection of the plant leaf. The ensemble classification method is derived using integration of the two classification methods i.e. RF (Random Forest) and KNN method. Random Forest technique applied for optimal feature selection, and later features have been classified using K-Nearest Neighbour method. The extracted feature set has been classified by Support Vector Machine in order to analyse performance of the proposed method.

IV. PROPOSED APPROACH AND RESULTS

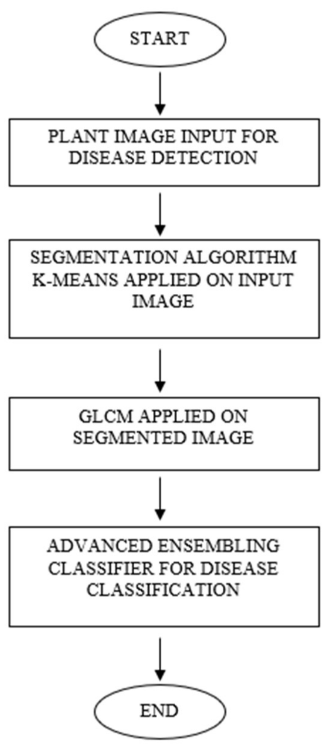

Fig. 3: Proposed steps for classifying the Disease from the tomato leaf

Proposed Algorithm: Steps involved for Disease Detection.
Step 1: Input image of the leaf procured from the dataset.
Step 2: Divide the image Mmask into smaller regions Ktiles (square tiles).
Step 3: Apply GLCM algorithm for Feature Extraction on the Segmented Image
Step 4: Finally, apply advanced Ensemble Classifier on the feature extracted part to detect disease.
Step 5: Stop

The proposed work is mainly focused on detection of plant disease type. The advanced ensemble classification approach is proposed to detect and classify the disease of the tomato leaf. The advanced ensemble classifier combines Random Forest and K-Nearest Neighbour machine learning methodology. The proposed approach performance has been evaluated with a traditional support vector machine by measuring performance using three measures accuracy, recall and precision.

Accuracy: For any work accuracy refers to the measure of exactness of the work. Here in classification, accuracy can be defined as the number of correctly classified patterns compared to the total given number of samples. Accuracy measurement is carried out using the formula stated below:
X = (T/N) * 100

Recall: The total number of true positive patterns retrieved from the total number of positive identified patterns is called as recall. The carried out formula for Recall is:

Recall = TruePositives / (TruePositives + FalseNegatives)

Precision: It is called as the number of positive samples taken out from the total number of samples acknowledged positive by the classification model and is calculated as follows:

Precision = TruePositives / (TruePositives + FalsePositives)

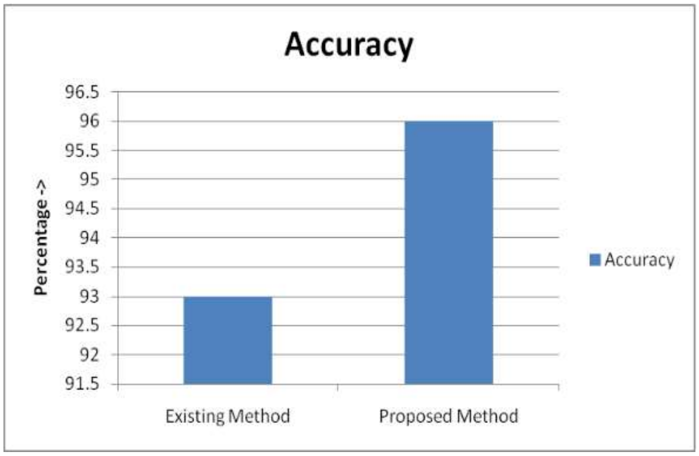

Fig 4 Comparison of the Accuracy

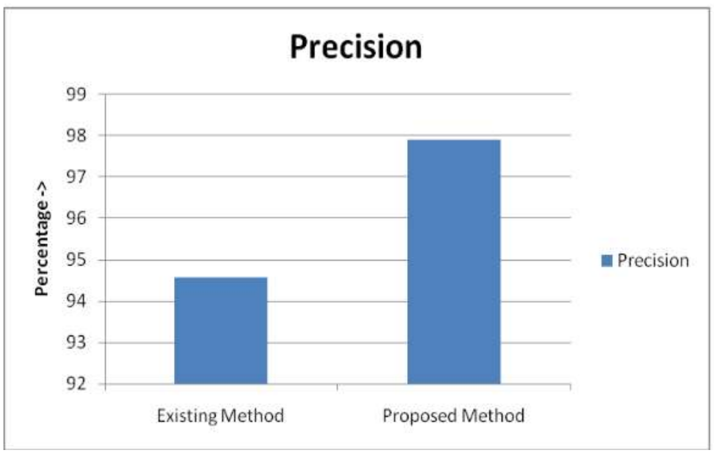

Fig 5 Comparison of the Precision

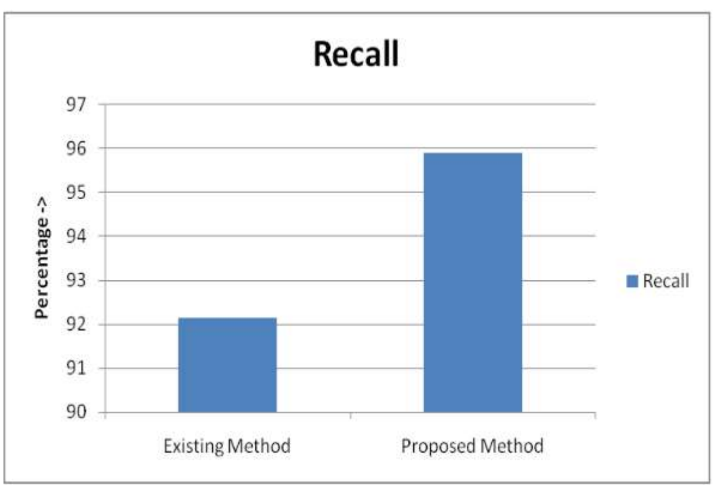

Fig 6 Comparison of the Recall

Fig. 4 depicts the accuracy comparison between the proposed advanced ensemble classifier and Support Vector Machine (an existing method). The advanced ensemble classifier has achieved the accuracy of 96% whereas SVM accuracy is at 93%. Fig. 5 shows the precision values that are achieved by the traditional SVM method and the proposed advanced ensemble classifier. And the figure 6 presents the recall values acquired by SVM and the proposed ensemble classifier respectively. Thus the experimental results revealed that the proposed approach outperformed SVM.

V. CONCLUSION AND FUTURE SCOPE

Plant disease detection stands as a major challenge in image processing and machine learning techniques. The proposed work presents an advanced ensemble classifier approach for detecting the tomato plant diseases. This work applied K-mean algorithm for the region-based segmentation. Later, GLCM method is applied for feature extraction from the segmented images. The advanced ensemble classification which is a combination of Random Forest and K-Nearest Neighbor for the classification of the detected disease. This proposed method of research is shown to be effective for disease detection in tomato leaf with good accuracy, precision and recall. The future of this work lies in developing a mobile based app which can be used by agricultural field workers or farmers with which they can capture images of the diseased plant and themselves they can see the results in the app which shows the type of disease that a plant is suffering with.

REFERENCES

[1]. Naresh K. Trivedi et.al., Early Detection and Classification of Tomato Leaf Disease Using High-Performance Deep Neural Network", Sensors 2021.

[2].RekhaBalodi, SunainaBisht, AbhijeetGhatak and K.H. Rao, "Plant disease diagnosis: technological advancements and challenges", Indian Phytopath. 70 (3): 275-281 (2017).

[3]. MonishankerHalder, Ananya Sarkar, HabibullahBahar, "PLANT DISEASE DETECTION BY IMAGE PROCESSING: A LITERATURE REVIEW", 2019, SDRP Journal of Food Science & Technology

[4]. Surender Kumar, Rupinder Kaur, "Plant Disease Detection using Image Processing: A Review", International Journal of Computer Applications (0975 – 8887) Volume 124 – No.16, August 2015

[5]. Sabah Bashir, Navdeep Sharma, "Remote Area Plant Disease Detection Using Image Processing", IOSR Journal of Electronics and Communication Engineering (IOSRJECE), Volume 2, Issue 6 (Sep-Oct 2012), PP 31-34.

[6]. Rajneet Kaur, Miss. Manjeet Kaur, "A Brief Review on Plant Disease Detection using in Image Processing", International Journal of Computer Science and Mobile Computing, Vol. 6, Issue. 2, February 2017, pg.101 – 106.

[7]. Sushil R. Kamlapurkar, "Detection of Plant Leaf Disease Using Image Processing Approach", International Journal of Scientific and Research Publications, Volume 6, Issue 2,

February 2016
[8]. Vijai Singh, Varsha, A K Misra, “Detection of unhealthy region of plant leaves using image processing and genetic algorithm”, International Conference on Advances in Computer Engineering and Applications, Year: 2015 | Conference Paper | Publisher: IEEE.
[9]. AbiramiDevaraj, KarunyaRathan, SarvepalliJaahnavi, K Indira, “Identification of Plant Disease using Image Processing Technique”, International Conference on Communication and Signal Processing (ICCSP), Year: 2019 |Conference Paper | Publisher: IEEE
[10]. R. Meena Prakash, G.P. Saraswathy, G. Ramalakshmi, K.H. Mangaleswari, T. Kaviya, “Detection of leaf diseases and classification using digital image processing”, International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Year: 2017 | Conference Paper | Publisher: IEEE.
[11]. Sharath D M, Akhilesh, S Arun Kumar, Rohan M G, Prathap C, “Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight”, International Conference on Communication and Signal Processing (ICCSP), Year: 2019 | Conference Paper | Publisher: IEEE.
[12]. Gurleen Kaur Sandhu, Rajbir Kaur, “Plant Disease Detection Techniques: A Review”, International Conference on Automation, Computational and Technology Management (ICACTM), Year: 2019 | Conference Paper | Publisher: IEEE
[13]. Sachin D. Khirade, A.B. Patil, “Plant Disease Detection Using Image Processing”, International Conference on Computing Communication Control and Automation, Year: 2015 | Conference Paper | Publisher: IEEE.
[14]. R Anand, S Veni, J Aravinth, “An application of image processing techniques for detection of diseases on brinjal leaves using k-means clustering method”, International Conference on Recent Trends in Information Technology (ICRTIT), Year: 2016 | Conference Paper | Publisher: IEEE.