Enhancing Betel Vine Leaf Disease Diagnosis Through Machine Learning Techniques

Rajkumar G[1], Gayathri Devi T[2]*, Karthikeyan S[3], Srinivasan A[4]

[1,3] Department of Electronics and Communication Engineering, School of Electrical and Electronics Engineering,
Sastra Deemed to be University, Thanjavur, India
[2,4] Department of Electronics and Communication Engineering, Srinivasa Ramanujan centre,
Sastra Deemed to be University, Kumbakonam, India

Abstract

Betel vine, known for its economic and cultural significance, is prone to various diseases that can severely impact its yield and quality. Traditional disease detection and diagnosis methods in betel vine crops often rely on expert visual inspection, which can be time-consuming and subjective. The study suggests machine learning based automated method for examining diseases in betel vines. The proposed work focuses on detecting and quantifying the impact of a disease on betel vine leaves using a Machine Learning algorithm. By leveraging a vision-based strategy, the proposed method effectively detects and analyzes external signs of illness. Machine learning techniques are used to locate the disease-affected portions of the leaves. Subsequently, the affected area is measured and extracted based on the collected data on plant features. Integrating modern advancements in machine learning has significantly enhanced the performance and accuracy of disease detection in betel vine leaves. This research aims to develop a cost-effective and efficient approach for studying diseases in betel vine leaves, catering to the needs of farmers and agricultural researchers. The results indicate a classification accuracy of 98.73 per cent for disease categorization.

Keywords: Betel vine, K-Means Clustering, Median Filter, Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine.

Introduction

India is renowned for its agriculture, with farmers being its backbone. They rely on factors such as soil and climate to ensure successful and healthy cultivation, making disease detection in underground farming crucial. The outward appearance of agricultural products plays a significant role in their sale value and consumer behaviour. Therefore, quality inspection and grading methods are essential in agriculture to cultivate robust and healthy plants. Plant diseases can lead to substantial losses in output and the economy within the agriculture industry, making disease management a challenging endeavour. These diseases, often characterized by coloured spots on leaves or stems, are predominantly caused by fungi, bacteria, and viruses. Detecting these symptoms manually is the usual practice, and farmers often resort to pesticides without realizing the specific disease affecting their crops. If left undetected, these illnesses can significantly reduce yield and productivity, posing difficulties for farmers.

The process of detecting and identifying plant diseases is time-consuming. Agriculture is crucial in human society, particularly in India, where it forms the backbone of the economy and employs over 70% of the population. Identifying various plant diseases is vital to avoid yield reduction, but it requires substantial labour and a long duration. To address this, image processing and machine learning (ML) models are employed to identify plant illnesses. Many farms struggle with diagnosing plant diseases, leading to agricultural product losses. By utilizing photos and videos of crops, agricultural scientists can provide more accurate diagnoses. Since many diseases have initially invisible symptoms that can result in social and economic losses, the use of various algorithms enables automated disease detection. Machine learning plays a crucial role in diagnosing plant diseases as it reduces human labour. It can be applied in agriculture for various purposes, such as detecting damaged fruit, leaves, or stems, measuring the affected area, and determining colour variations. Machine learning operates autonomously based on predefined instructions, leveraging extensive training data to predict accurate outcomes. Classification of diseases involves analyzing leaf colour, damaged leaf area, and textural features.

In this paper, different image features were explored to achieve the highest level of accuracy in identifying various plant leaf diseases. Previously, plant disease identification relied on visual leaf inspection by professionals or chemical techniques. However, these approaches require large expert teams and continuous plant monitoring, which can be costly for large farms. In such cases, the proposed approach demonstrates its capability to aid in monitoring vast crop fields.

Diseases primarily hinder the cultivation of betel vine, resulting in decreased output. Consequently, developing a system capable of identifying betel vine leaf infections is crucial. Machine learning algorithms can be employed for disease detection, involving image acquisition, processing, segmentation, feature extraction, and classification. These algorithms, including median filtering, k-means clustering, grey-level co-occurrence matrix, and Support Vector Machine, enable swift and accurate identification of diseases in a short timeframe.

Related Works

Sharath D. M. et al. (2019) developed a Bacterial Blight detection system for Pomegranate plants. They used various features such as colour, mean, homogeneity, standard deviation, variance, correlation, entropy, and edges. The system employed grab-cut segmentation for segmenting the region of interest and a Canny edge detector to extract edges. The developed system successfully predicted the infection level in the fruit. PoojaKulinavar, in 2017, presented a leaf classification method that involved converting the image to greyscale and performing segmentation using the K-Means clustering algorithm. Colour and texture features were extracted, and the classification was done using SVM classifier techniques. The disease detection and classification achieved an accuracy rate of 92% (PoojaKulinavar, 2017).

Shrutiet al. (2014) focused on examining tomato crop diseases, specifically "Cercospora leaf spot" or "Cercospora cruciferous," which causes the leaves to turn grey, white, and eventually black. They proposed a novel technique to visually monitor plant growth from the stem and determine the type of fungus infecting it. The classification of Grape Leaf Diseases using Neural Networks was conducted by Sanjeev S Sannakki in 2013. In this approach, texture analysis was developed using the spatial grey method, and classification was performed using feed-forward propagation neural networks. The diseases affecting grape leaves, such as powdery mildew, downy mildew, and grey mould, were accurately classified using hue features with a 100% accuracy rate (Sanjeev S Sannakki, 2013).

In 2013, Prof. Sanjay B. Dhaygude developed a plant leaf disease detection method using image processing. The Hue Saturation Value (HSV) transformation was applied to the image, followed by masking and removing green pixels. Features were extracted using a colour co-occurrence method (Prof. Sanjay B. Dhaygude, 2013). Smita Naikwadi et al. (2013) presented an experiment on classifying and detecting plant diseases. They emphasized the limitations of insecticides and proposed a technique that involved identifying green-coloured pixels, masking them using Otsu's method, and removing infected clusters based on red-green and yellow colours. Their proposed technique showed superior performance in plant detection.

Proposed Methodology

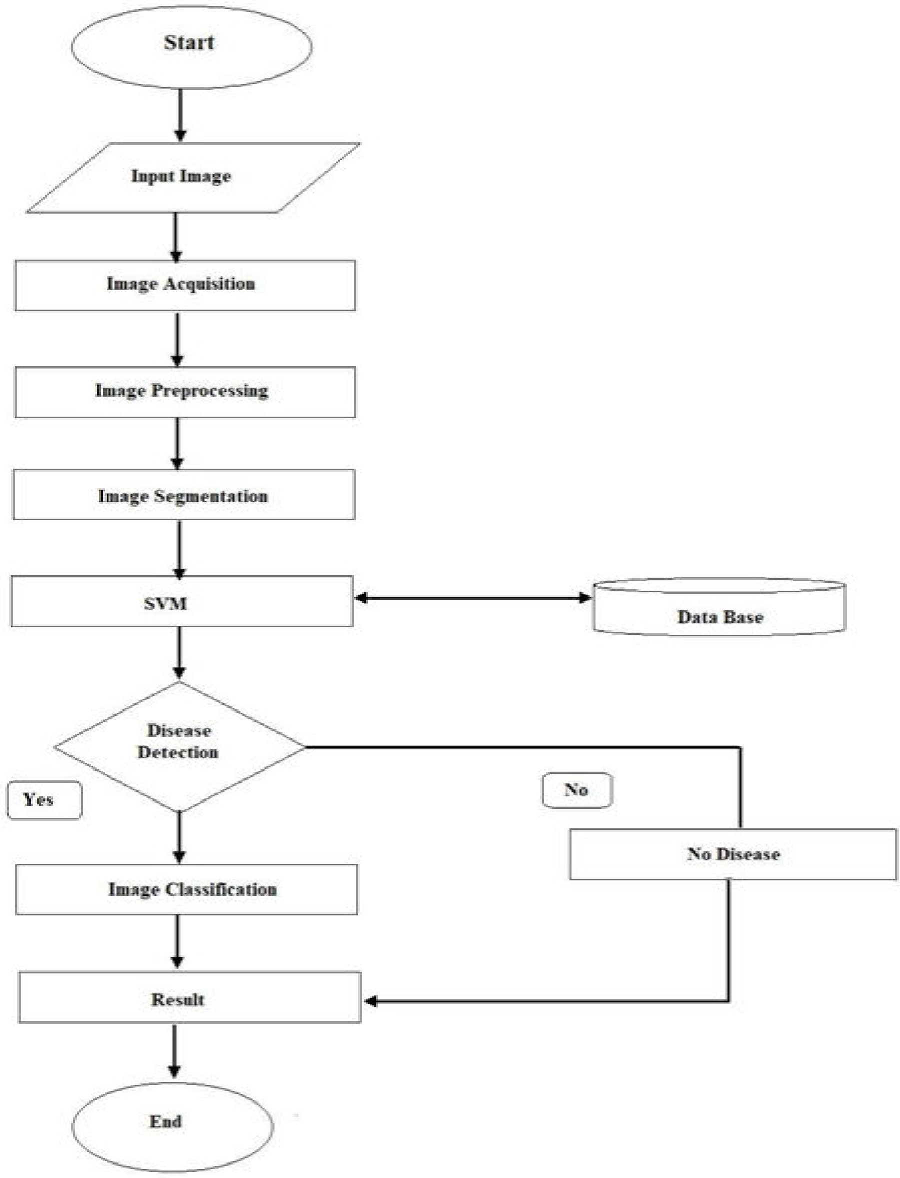


Figure 1 Proposed Methodology

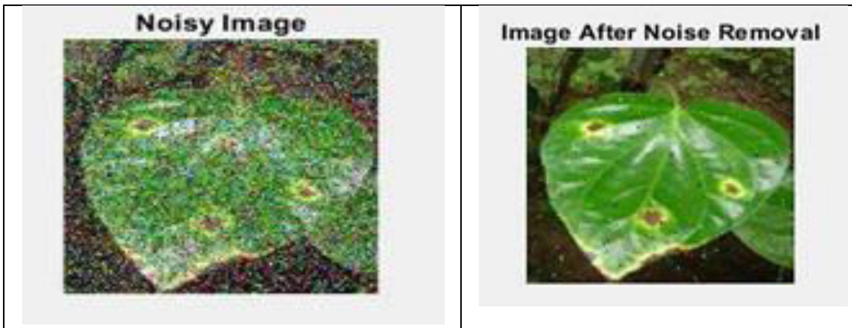
The proposed methodology consists of several stages. Firstly, a median filter is applied to the betel vine images to reduce noise and enhance the image quality. Next, K-means clustering-based segmentation is employed to separate the diseased regions from the healthy ones, effectively isolating the affected areas of the plant. This segmentation step enables targeted analysis and diagnosis of the disease. After segmentation, texture features are extracted from the segmented regions using the Gray-Level Co-occurrence Matrix (GLCM). GLCM provides statistical information about the spatial relationships of pixel intensities, capturing important textural patterns associated with diseases. These features are inputs to a Support Vector Machine (SVM) classifier, a powerful machine learning algorithm widely used in image classification tasks.

Image Preprocessing

Preprocessing is done to improve the quality of the image without reducing thesharpness of the image. We can remove unwanted disturbances and boost some properties important for the application we're working on by processing. These functionalities may differ depending on the application. Many image pre-processing techniques are applied to eliminate noise from an image or exclude other objects. Because the pixel size of the original image is large and the overall operation takes longer, image scaling is used to convertthe original image into thumbnails. After converting the image into thumbnails, the pixel sizewill drop, and the entire procedure will take less time.

Median Filter

The median filter is a non-linear filtering application for reducing image noise. Noise reduction is a typical pre-processing procedure used to improve the results of later processing. As it preserves edges while lowering noise under specific situations, median filtering is extensively used in digital image processing and has applications insignal processing.



Input Image	Median Filtered Image

Figure 2 Preprocessed Image

Image Segmentation

The main purpose of image segmentation is to extract important information from an image and facilitate further analysis, understanding, or manipulation of its contents. It is crucial in numerous applications, including medical imaging, autonomous driving, object recognition, video surveillance, and image editing. There are several image segmentation approaches, each with advantages and limitations.

Image segmentation is a challenging problem due to the complexity and variability of images. The choice of the segmentation technique depends on the specific application requirements, image characteristics, and available computational resources. Researchers continue to develop new algorithms and techniques to improve the accuracy and efficiency of image segmentation, driving advancements in various fields relying on computer vision.

K-Means Clustering

Here we are using the K-mean clustering technique for segmentation. The K-means clustering algorithm is a popular unsupervised machine-learning technique for grouping similar data points into clusters. It is an iterative algorithm that aims to partition a given dataset into K distinct clusters, where K is a user-defined parameter.

K-means algorithm works in the following way:

- Randomly select K data points from the dataset as the initial cluster centres (also known as centroids).
- Assign each data point to the nearest centroid based on a distance metric, commonly the Euclidean distance. This step forms K clusters.
- Recalculate the centroids of the clusters by computing the mean of all data points assigned to each cluster.
- Repeat steps 2 and 3 until convergence criteria are met. The convergence criteria can be a fixed number of iterations, a threshold on the change in centroid positions, or when the assignment of data points to clusters no longer changes.
- The algorithm terminates when convergence is achieved and the final cluster assignments are obtained.

The K-means algorithm aims to minimize the sum of squared distances between data points and their respective centroid within a cluster. It converges to a local optimum, meaning the final clustering may depend on the initial random centroid selection. To mitigate this, the algorithm is often run multiple times with different initializations, and the best clustering (i.e., the one with the lowest sum of squared distances) is chosen.

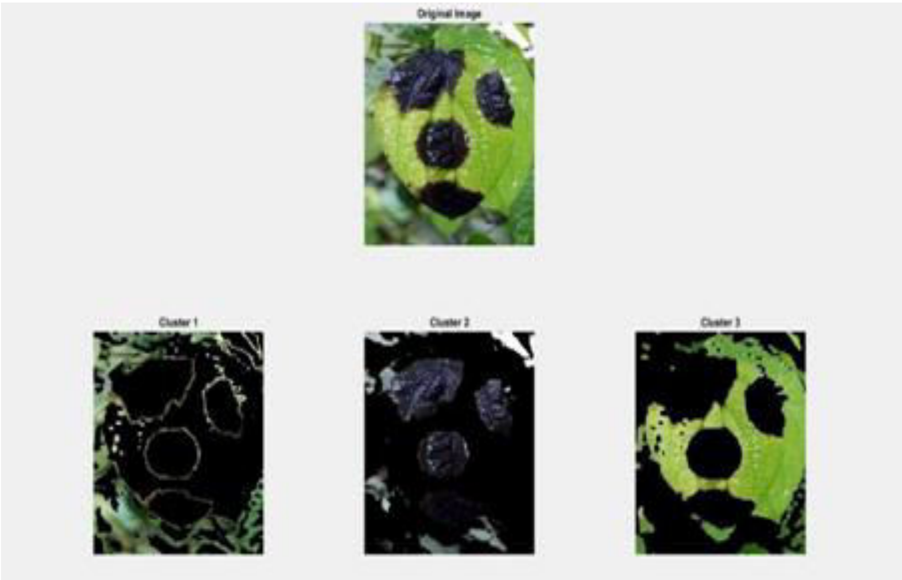


Figure 3 Segmentation Output

Feature Extraction

Disease detection relies heavily on feature extraction. This is one of the most important aspects of the detection of disease. Feature extraction is one of the processes of transforming the original visual input into numerical features. As a result, it reduces the no. of resources needed to represent the data. It allows it to be processed while preserving the raw data set'sinformation, resulting in better results. The image components, such as colour, intensity, and so on, are modified to highlight the disease's hidden features. Compared to a test database of leaf disease, these extracted attributes will aid in detecting disease rapidly and efficiently.

GLCM (Grey Level Co-occurrence Matrix)

GLCM, which stands for Gray-Level Co-occurrence Matrix, is a popular method for texture analysis and feature extraction in image processing and computer vision. It provides a statistical description of the spatial relationships between pairs of pixels in an image. GLCM is a versatile technique and has been widely applied in various applications, including object recognition, texture classification, medical image analysis, and remote sensing, to name a few. It provides valuable insights into an image's texture properties and helps extract meaningful information for subsequent analysis.

The basic idea behind GLCM is to calculate the frequency of occurrence of different pixel intensity pairs at various spatial relationships in an image. By examining these patterns, GLCM can capture information about the texture or spatial structure of the image. First, the input image is usually converted to a grayscale image if it's in colour, as GLCM operates on grey-level intensities. Then the GLCM considers pairs of pixels at specific spatial relationships or distances in the image. These spatial relationships can be defined by pixel offsets, such as horizontal, vertical, diagonal, or a combination of these directions. For each pair of pixels at a specific spatial relationship, GLCM calculates each possible grey-level pair's occurrence frequency or probability. The co-occurrence is usually represented as a matrix, where each element (i, j) represents the frequency or probability of pixel pair (i, j) occurring in the image. Once the GLCM is computed, various statistical measures can be derived to characterize the image's texture. Commonly used features include:

- Contrast: Measures the local intensity variations between adjacent pixels in the image.
- Homogeneity: Describes the similarity or closeness of the grey-level pairs to the diagonal elements of the GLCM.
- Energy: Represents the sum of squared elements in the GLCM, indicating the image's uniformity or complexity.
- Entropy: Measures the randomness or uncertainty of the texture in the image.
- Correlation: Captures the linear dependency between grey-level pairs in the GLCM.

These features can be used as inputs for machine learning algorithms or for further analysis and classification tasks.

Classification

During the classification phase, the features extracted from the leaves are compared with the stored feature values of leaves in the dataset. The SVM classifier is used for image classification. The Support Vector Machine (SVM) classifier creates a set of hyperplanes in a high-dimensional space to achieve effective separation. However, the SVM classifier is inherently designed for binary classification and does not directly support multi-class classification. The problem is divided into multiple binary classification tasks to handle multiclass classification, applying the same two-class principle.

The SVM classifier uses training samples. SVM is used to solve two types of issues in a common format. A multiclass SVM with K classes can be generated from binary problems. Where K is more than two. The two options are one-against-one and one-against-all. After thetraining phase, the features were retrieved and used to categorize the database using SVM classification.

Results and Discussions

The input Betel leaf images are collected from Plant Village Database. This experimental work uses 2200 images with four diseases and one healthy category. Out of the total image samples, 1175 images are used to train the Classifier and 1025 images are used to test the Classifier. Then these images are classified using an SVM classifier, and test results show that an accuracy of 96 percent is achieved out of 200 iterations.

The Betel diseases used in this work are Anthracnose, Leaf Rot, Bacterial Leaf Spot, Powdery Mildew Three approaches are used to extract the features of these input images. Then Classifier takes these extracted features to detect the disease category. MATLAB is used to perform all the experiments.

The figure shows the input test images, preprocessed images, segmented images and classification outputs of the Leaf Rot diseases. Here, the input images are first converted to grayscale and then preprocessed using a median filter. Then the image is segmented using K-Means Clustering, and from the segmented images, the features are extracted using Gray Level Co-occurrence Matrix.

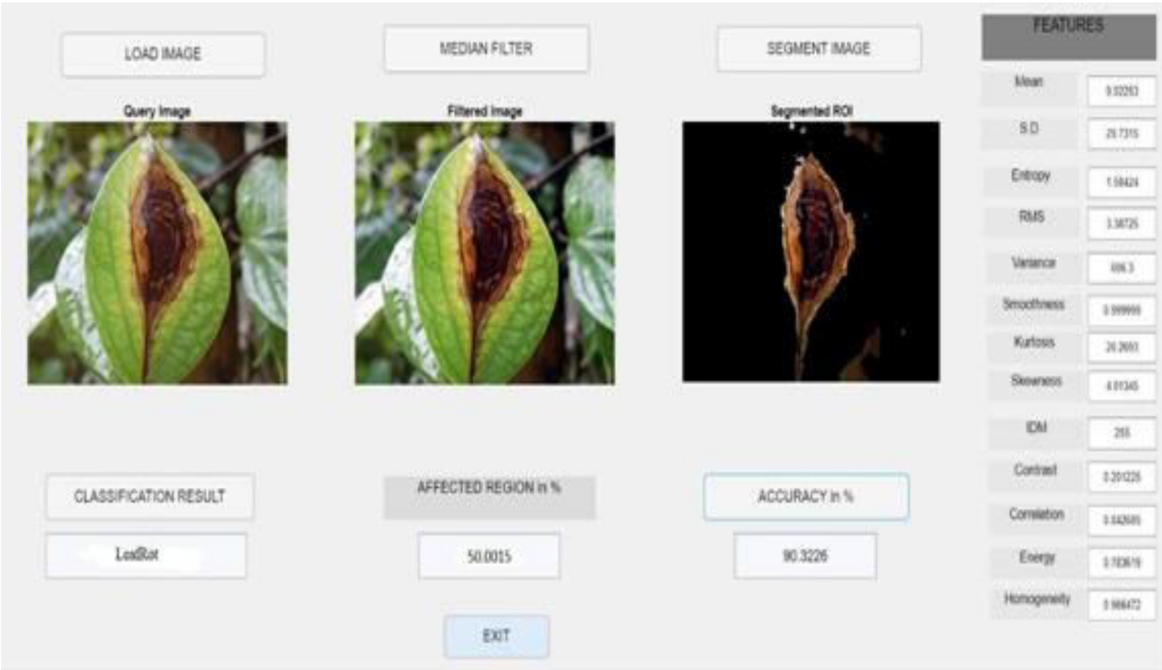


Figure 4 Results of various output images of Betel Anthracnose

Performance Measures

In addition to the accuracy, the performance of the proposed system uses the performance measures of precision, Recall, FPR and F-Measure.

Table 1 illustrates the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values has been acquired by the SVM classifier for each category in the betel leaves dataset. The SVM classifier has misclassified 11 images in healthy,7 images in Anthracnose, 10 in Leaf Rot, and 12 in the Powdery Mildew Category.

Table 1.TP, TN, FP, FN values of Betel Leaf Dataset

Category	TP	TN	FP	FN
Healthy	230	772	11`	12
Anthracnose	430	582	7	6
Leaf Rot	126	878	10	11
Powdery Mildew	201	802	10	12

Table 2 shows the performance Parameters of (Accur.acy, Precision, Recall, FPR & F-Measure) for the Betel Leaves Dataset obtained by the SVM Classifier

Table 2 Performance Parameters of (Accuracy, Precision, Recall, FPR & F-Measure) for Betel Leaves Dataset

Category	Testing Accuracy	Precision	Recall	FPR	F-Measure
Healthy	97.76	95.44	0.23	0.01	0.46
Anthracnose	98.73	98.40	0.42	0.01	0.84
Leaf Rot	97.95	92.65	0.13	0.01	0.26
Powdery Mildew	97.85	95.26	0.20	0.01	0.40

The SVM classifier achieved the highest accuracy value of 98.73 % in the Anthracnose, whereas 97.76% of accuracy was achieved in the healthy category, which was the lowest accuracy than other categories. It is observed from the experimental methods that the proposed method can identify and classify the leaf diseases in their initial stages with less effort.

Conclusion and Future Scope

This study proposes an automated approach for the analysis of diseases in betel vine crops using machine learning techniques. By leveraging a vision-based strategy, the proposed method effectively detects and analyzes external signs of illness on betel vine leaves. Machine learning algorithms are employed to identify disease-affected portions of the leaves, and the affected area is measured and extracted based on collected data on plant features. Integrating modern advancements in machine learning has significantly enhanced the performance and accuracy of disease detection in betel vine leaves.

The results of the study indicate a classification accuracy ranging from 98.73. per cent for disease categorization. This demonstrates the potential of machine learning techniques in accurately identifying and quantifying diseases in betel vine crops, providing a valuable tool for farmers and agricultural researchers. By automating the disease detection process, the proposed approach can save time and reduce subjectivity compared to traditional visual inspection methods.

The future scope of this work includes several potential avenues for further improvement and application. Firstly, expanding the dataset used for training the machine learning models can help improve the accuracy and robustness of disease detection. Additionally, incorporating more advanced machine learning algorithms or exploring deep learning techniques may enhance the performance of the system further.

Furthermore, integrating this automated disease detection approach with other agricultural technologies, such as remote sensing or Internet of Things (IoT) devices, could provide real-time monitoring and early detection of diseases in betel vine crops. This would enable prompt intervention and disease management strategies, ultimately improving crop yield and quality.

Moreover, the proposed method can be extended to develop a comprehensive decision support system for betel vine disease management. By incorporating historical data, weather information, and expert knowledge, the system could provide personalized recommendations for disease prevention and control, aiding farmers in making informed decisions.

Overall, this research presents a cost-effective and efficient approach for analyzing diseases in betel vine leaves, with significant potential for practical implementation and future enhancements in the field of agricultural disease management.

References

1. Md Zahid Hasan, Nahid Zeba, Md. Abdul Malek and Sanjida Sultana Reya(2021). A Leaf Disease Classification Model in Betel Vine Using Machine Learning Techniques, 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). 12-23.

2. J. Hang, D. Zhang, P. Chen, J. Zhang, and B. Wang (2019). Classification of plant leaf diseases based on improved convolutional neural network, Sensors, 19(19),4161.

3. N. Ganatra and A. Patel, (2018). Survey on diseases detection and classification of agriculture products using image processing and machine learning,Int. J. Comput. Applications. 180(3), 7–12.

4. S.K.Jayanthi and C. Lalitha (2017). Betel Leaf Disease Detecttion using Histogram of Oriented Gradients and Mukticlass SVM. International Journal of Innovative Research in computer and Communication Engineering. 5, 13994-14001

5. Suman T and DhruvakumarT (2015). Classification of Paddy leaf diseases using shape and color features. IJEE, 7(1),239250.

6. Shruti and Nidhi Seth (2014). Fungus/Disease Analysis in Tomato Crop using Image Processing Technique. International Journal of Computer Trends and Technology (IJCTT). 13(2).

7. Smita Naikwadi, Niket Amanda (2013). Advances in Image processing for detection of plant disease. International Journal of Application or Innovation in Engineering & Management, 2(11).

8. Yusnita R, Fariza Norbaya, and Norazwinawati Basharuddi(2012). Intelligent Parking Space Detection System Based on Image Processing. International Journal of Innovation, Management, and Technology. 3(3).

9. Al-Bashish, D., M. Break, and S. Bani-Ahmad (2011). Detection and classification of leaf diseases using K-Means based segmentation and neural networks based classification, Information Technology Journal, 10(2).

10. Bauer S.D., F. Korc., W. Forstner (2011). The potential of automatic methods of classification to identify leaf diseases from multi spectral images. Precision Agriculture, 12.

11. H.Al-Hairy, M.Reyalat, M.Braik and Z.Al Rahamneh (2011). Fast and Accurate Detection and Classification of Plant Diseases, International Journal of Computer. Applications . 17(1),1-48.

12. Sabine D. Bauer, Filip Korc, Wolfgang Forstner (2011), The Potential of Automatic Methods of Classification to identify Leaf diseases from Multispectral images,, Precision Agriculture. 12:361–377

13. Dheeb Al Bashish, Malik Break, (2010) A Framework for Detection and Classification of Plant Leaf and Stem Diseases, International Conference on Signal and Image Processing. 113-118.

14. Dae Gwan Kim, Thomas F. Burks, Jianwei Qin, Duke M. Bulanon (2009), classification of

grapefruit pee diseases using color texture feature analysis, International Journal Agriculture and Biological Engineering.2(3)..

15. Wang Jue, Wang shitong (2008). Image Thresholding Using Parzen Window Estimation.Journal of applied sciences 8(5),772-779.

16. Krystian Mikolajczyk and Cordelia Schmid (2005). A performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10).1615 – 1630.

17. Argenti, F., L. Alparone, and G.Benelli (1990).Fast algorithms for texture analysis using co-occurrence matrices, IEEE proceedings, 137(6).