
Tomato leaf disease detection using image processing technique

Sivagami, S.^{1*} and Mohanapriya, S.²

¹Research Scholar, Periyar University, Salem, Tamil Nadu, India; ²Head, Department of BCA, Sona College of Arts and Science, Salem, Tamil Nadu, India.

Sivagami, S. and Mohanapriya, S. (2021). Tomato leaf disease detection using image processing technique. *International Journal of Agricultural Technology* 17(3):1135-1146.

Abstract Agriculture is the backbone of India. It contributes nearly 17% of the total GDP and it employs more than 60% population. Now a day's plants are severely affected by a different type of diseases. The agriculture industry in all over the world are affected by severe economic losses due to these diseases in plants. In early days manual inspection was proposed to identify the disease, it is a difficult and time-consuming process and this can be overcome by automated methods in recent days. In this paper, a new method is proposed to identify disease in tomato plant. The proposed method consists of four stages, namely pre-processing, segmentation, feature extraction, and classification. In the pre-processing step, acquired images are resized and the noise was removed using the Weiner filter technique. Segmentation one of the important steps, for that modified K-Means image segmentation algorithm, was proposed. After the segmentation process, important features are extracted from the segmented image using Grey Level Co-occurrence Matrix (GLCM) feature extraction method. Finally, diseased leaves are classified using classification algorithms like Support Vector Machine (SVM) and Adaptive Neuro Fuzzy Inference System (ANFIS). The experiments are performed from the tomato leaf images in plant village datasets. The proposed methodology is tested for five types of diseases in tomato plants namely Bacterial Spot, Early Blight, Leaf Mold, Late Blight, Septoria Leaf Spot, and normal tomato leaf. Results proved that the classification accuracy was improved by using modified K-Means and ANFIS classifier when compared to the K-means segmentation with classification algorithms.

Keywords: Disease, K-Means, Modified K-Means, Tomato

Introduction

Image processing is one of the emerging techniques to be used in most of the fields like medical, Remote sensing, video processing, Pattern recognition, Agriculture, transmission and encoding, etc., Among all agriculture is the most important one and in agriculture detection of disease and deficiency is a most important one because if we identify the disease and deficiency in the early stage (Santhosh Kumar and Raghavendra, 2019). It is easy for the farmer to take necessary action in order to avoid economic loss. Due to crop damage,

* **Corresponding Author:** S. Sivagami; **Email:** sivagamiacar@gmail.com

economic benefits are reduced this is called economic loss. crop damage may occur due to weather, pests, and diseases. Among these three conditions we cannot change the weather it is a natural one but we can control pests and diseases. To control the disease first we should know that it occurs for that image processing mainly used (Youwen *et al.*, 2008).

There are lots of tasks in image processing among them segmentation is one of the important tasks because after segmentation only the remaining task like feature extraction and classification occurs. There are two types of image segmentation techniques that are layer-based and block-based segmentation methods. Block-based segmentation methods are based on many things in the image that is color information of the image which is used to create histogram, pixels which are used to identify edges or boundaries of the image or texture information. The block-based segmentation method also divided into two category regions based and edge or boundary-based. Split the whole image into subgroups or clusters is the concept used in the region-based segmentation method. Clustering, split and merge, normalized cuts, region growing, and thresholding are some of the region-based segmentation methods. Among these methods in this paper, we used the clustering method. K-Means segmentation method comes under clustering, it splits the given image into K segments by adding points to the segment where the difference between the point and average of the segment is smallest (Nirmalakumari *et al.*, 2020).

More than 80 percent of diseases in plants are caused by fungi and remaining are caused by Bacteria and Viruses. Once the plant was affected by disease than its color, shape its ability to respond to pathogens that are changed. If the plant is affected by diseases then the quantity and quality of the product from plants are get reduced this will lead to economic loss (Shrivastava *et al.*, 2016).

In the agriculture domain, there are lots of challenges that exist such as complex background, the existence of noise, color-based similarity among various symptoms, there exist more symptoms in one image. These challenges are not easy to detect from an image. The challenges based by the researcher in this agriculture domain using soft computing techniques are a. Noise removal and enhancement in the pre-processing step. b. Region of interest is correctly segmented by an efficient segmentation algorithm. c. The sufficient number of features extracted from the segmented region for correct classification d. Efficient classification algorithm that correctly classifies the given image in less time (Khan *et al.*, 2020).

Image segmentation is the most important task in image processing and for image segmentation, lots of algorithms are already available among all the algorithms K-means algorithms are one of the simplest, most accurate, and

unsupervised algorithms that means it does not need any prior knowledge about the problem in hand. There are some drawbacks in K-Means that are we need to initialize the number of cluster and cluster center initially, it is very difficult to guess initially that how many numbers of clusters occur as well as what is initial cluster center (Nirmalakumari *et al.*, 2020). To overcome this problem, we proposed a new method in k-means it does not require any initialization. After segmentation important features are extracted by using the GLCM algorithm then classification was done by using some of the classification algorithms.

Materials and methods

The proposed method identifies the disease efficiently using a modified k-means algorithm for image segmentation and feature extraction by GLCM after extracting features using these algorithms classification was done by using two algorithms SVM and ANFIS. The step by step procedure of the proposed system is shown in Figure 1.

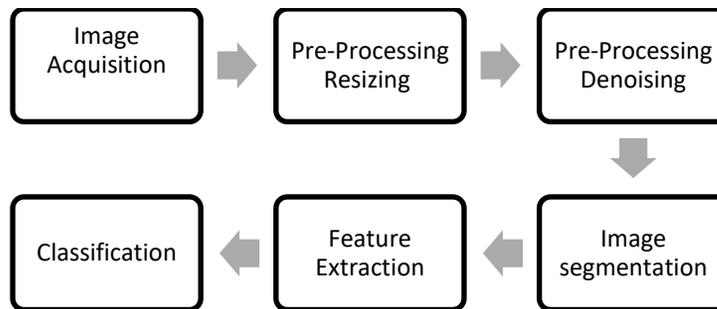


Figure 1. General Work flow of the proposed method

Image pre-processing

Images are collected from plant village dataset. The collected images are of different sizes and different color combinations and some images may contain noise. In the pre-processing step, all the images are converted into the same size. The main source of noise in a digital image are sensor may be affected by environmental conditions, light levels may not sufficient, dust particle may present on the scanner screen. There are different types of noise and that are classified as Gaussian noise, Impulse noise, Uniform noise, Multiplicative noise, and periodic noise. These noises were removed Using filtering algorithms, lots of filtering algorithms are available like a median filter, adaptive filter, Weiner filter, etc. In our proposed method we used

Weiner filter for removing noise. After denoising the image, we got an enhanced image that suppresses unwanted data in the image. The sample leaf images from each category are shown in Figure 2.

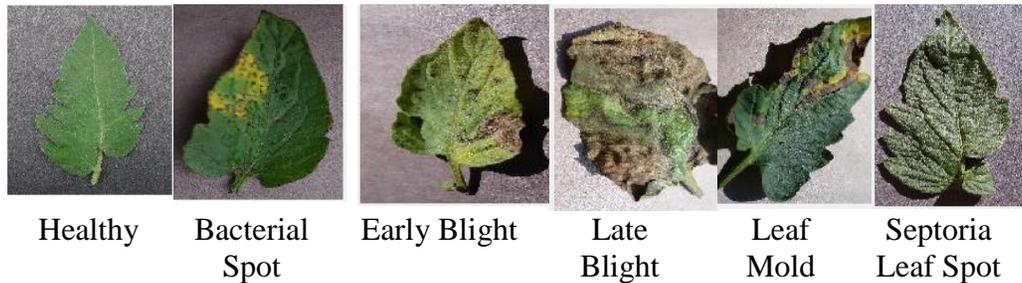


Figure 2. Sample leaf images

Image segmentation

After pre-processing the input image, segmentation was done. Segmentation is the process of splitting the image into different regions and focus on the region of interest. To segment the image there are lots of algorithms that are available among all the algorithms k-means is one of the simplest and easy to understand and unsupervised algorithms. Even though K-means has lots of advantages it has some disadvantages also, that is in the traditional K-means segmentation algorithm we have to initialize the number of clusters and cluster centers.

Algorithm 1: K-Means Segmentation algorithm

Step 1: initialize the k no of clusters, and k-means iteration Create an empty matrix for distance and labels

Step 2: Finding distance from cluster centers, cluster label, and minimum distance

Step 3: check cluster label is equal to k no of clusters, find the centers, and if the centers are nan value replace with random points. If not, nan keeps the center value as it is. Reshape the labeled index value

Step 4: make the clustered image using these centers.

It is very difficult to guess the number of clusters and cluster center initially. The segmentation results mostly based on these initial values so correct values must be predicted initially. To overcome this problem, we propose a modified k-means segmentation algorithm. In this modified k-means

segmentation algorithm it is not necessary to initialize any parameter, it automatically identifies the number of clusters and cluster center. Our proposed modified k-means is as follows.

Algorithm 2: Modified K-Means

Input: Colour image

Output: Segmented image

1. Pre-processing.
2. Make image pixels in a single column format.
3. Find the mean and the distance between mean value and the image, bandwidth for cluster center and check the values are in selected bandwidth or not and finally update the mean. Increment the counter for each iteration.
 $dist = (\text{sqrt}((\text{array}-\text{seed}). ^2));$ // Find distance between Seed and Gray Value.
 $disth = (\text{sqrt}(\text{sum}((\text{array}-\text{seed}). ^2)/\text{numel}(\text{array})));$ //Find bandwidth for Cluster Center.
 $qualified = dist \< disth;$ // Check values are in selected Bandwidth or not.
 $\text{newseed} = \text{mean}(\text{array}(qualified));$ //Update mean.
4. update the mean value
 check maximum number of clusters, if it is greater than 10 then reset the counter to 0.
5. Sort the centers
 Find the difference between two consecutive centers
 Find out minimum distance between two cluster centers
 Discard the cluster center lesser than the distance
6. Make the clustered image using these centers

Feature extraction

After segment the image the Region of Interest (ROI) was found, from this ROI, important characters are retrieved using the feature extraction method. In our proposed method we used the GLCM algorithm to extract useful features from ROI.

Grey level co-occurrence matrix

The following GLCM features are used for this proposed method Contrast, energy, entropy, Homogeneity, uniformity, etc.,

$$\begin{aligned}
\text{Contrast} &= \sum_i \sum_j (i-j)^2 m_{ij} \\
\text{Entropy} &= - \sum_i \sum_j m_{ij} \log m_{ij} \\
\text{Energy} &= \sum_i \sum_j m_{ij}^2 \\
\text{Homogeneity} &= \sum_i \sum_j \frac{c_{ij}}{1 + |i-j|}
\end{aligned}$$

Classification

Final step in the disease identification is classification. In our proposed work we did the classification based on two classifiers.

Support Vector Machine (SVM)

SVM was introduced by Cortes and it is mostly used for classification purposes. It is a supervised learning model along with learning algorithms that are used to analyze the data and mainly used to recognize the patterns for classification. Basic SVM is a binary linear classifier that takes a set of input and for each input, it predicts which of the two classes it belongs. In a linear classifier best hyperplane is the one that separates between two classes. If the best hyperplane exists for the given data set then it is called maximum margin hyperplane and the linear classifier is known as a maximum classifier. Svm minimizes experimental classification error as well as maximize the margin so it is called maximum margin classifier.

Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS integrates both artificial neural networks and fuzzy logic. It can acquire the benefits of both in one framework. It uses fuzzy rules and reasoning which was based on fuzzy set theory. Fuzzy set theory is different than normal set theory. In normal set theory, an element can either completely belongs to one set or not belongs to that set. If the element belongs to the set means value 1 otherwise 0 will be considered. In the fuzzy set theory, it will define how much the element belongs to the set that is it defines the value between 0 and 1. Fuzzy inference system (FIS) was created using the concepts of fuzzy set theory, fuzzy reasoning, and fuzzy if-then rules. A simple FIS system consists of two inputs and one output. The fuzzy rule for this input is as follows:

- Rule 1: if x is p1 and y is q1 then f1=p1x+q1y+r1
- Rule 2: if x is p2 and y is q2 then f2=p2x+q2y+r2

The adaptive network for the above fuzzy rules is illustrated in the following Figure 3. In this network framework, there are five layers each layer accepts inputs from the previous layer and produces output to the next layer. In each layer there may be many nodes, all the nodes in the same layer do the same function. Each node in a layer computes output based on the weighted sum of all input.

Every node in Layer 1 is the adaptive node which is doing the following function.

$$O_{1,i} = \mu_{P_i}(x), \text{ for } i = 1,2$$

$$O_{1,i} = \mu_{Q_i}(y), \text{ for } i = 3,4$$

Here x and y are the input to the node i, P and Q are the labels associated with that node.

In layer 2 every node is a fixed node with labelled π , whose output is the product of incoming signals.

$$O_{2,i} = w_i = \mu_{P_i}(x)\mu_{Q_i}(y), i = 1,2.$$

In layer 3 every node is fixed node labelled N. The ith node in this layer performs the following function.

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1,2.$$

Every node in layer 4 does the following function:

$$O_{4,i} = w_i f_i = w_i (a_i x + b_i y + c_i)$$

Here a_i, b_i and c_i are the parameter set of nodes of this layer 4.

In layer 5 there is only one node which computes overall output as summation of all incoming signals.

$$O_{5,i} = \sum_i W_i f_i(x, y) = \frac{\sum_i W_i f_i(x, y)}{\sum_i W_i}$$

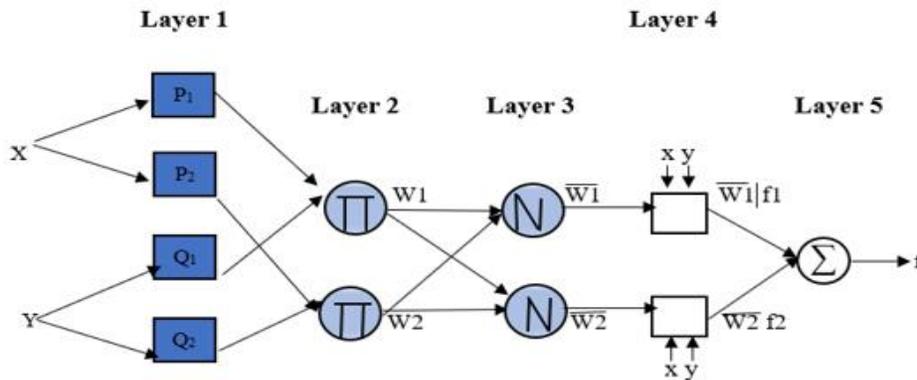


Figure 3. ANFIS architecture

Results

The proposed method considered five main diseases like Septoria leaf spot, Early Blight, Bacterial Spot, Late Blight, and Leaf Mold in the tomato plant. The input tomato leaf images were of different sizes, so the first image resize was done then the color image is converted into the gray image after that image denoising was done using the Gaussian filter which is followed by segmentation by using proposed modified K-means segmentation algorithm. After performing segmentation useful features are extracted from ROI using the GLCM feature extraction method. After getting useful features classification was done using SVM and ANFIS classifiers. Leaf and segmented images are shown in Figure 4. The dataset of images used for the proposed method is shown in Table 1.

Table 1. Tomato Leaf Dataset

| Type of Leaf | Number of Training Images | Number of Testing Images | Total Number of images |
|----------------------------|---------------------------|--------------------------|------------------------|
| Healthy images | 160 | 40 | 200 |
| Bacterial spot disease | 204 | 51 | 255 |
| Early blight diseases | 60 | 15 | 75 |
| Late blight disease | 124 | 31 | 155 |
| Leaf mold disease | 128 | 32 | 160 |
| Septoria leaf spot disease | 100 | 25 | 125 |

For this proposed procedure, the following GLCM features are used: contrast, energy, entropy, and homogeneity.

$$\begin{aligned}
 \text{Contrast} &= \sum_i \sum_j (i - j)^2 m_{ij} \\
 \text{Entropy} &= - \sum_i \sum_j m_{ij} \log m_{ij} \\
 \text{Energy} &= \sum_i \sum_j m_{ij}^2 \\
 \text{Homogeneity} &= \sum_i \sum_j \frac{c_{ij}}{1 + |i - j|}
 \end{aligned}$$

Where m_{ij} represents pixel value at point (i, j) of the given image. Leaf image and its segmented images are shown in Figure 4.

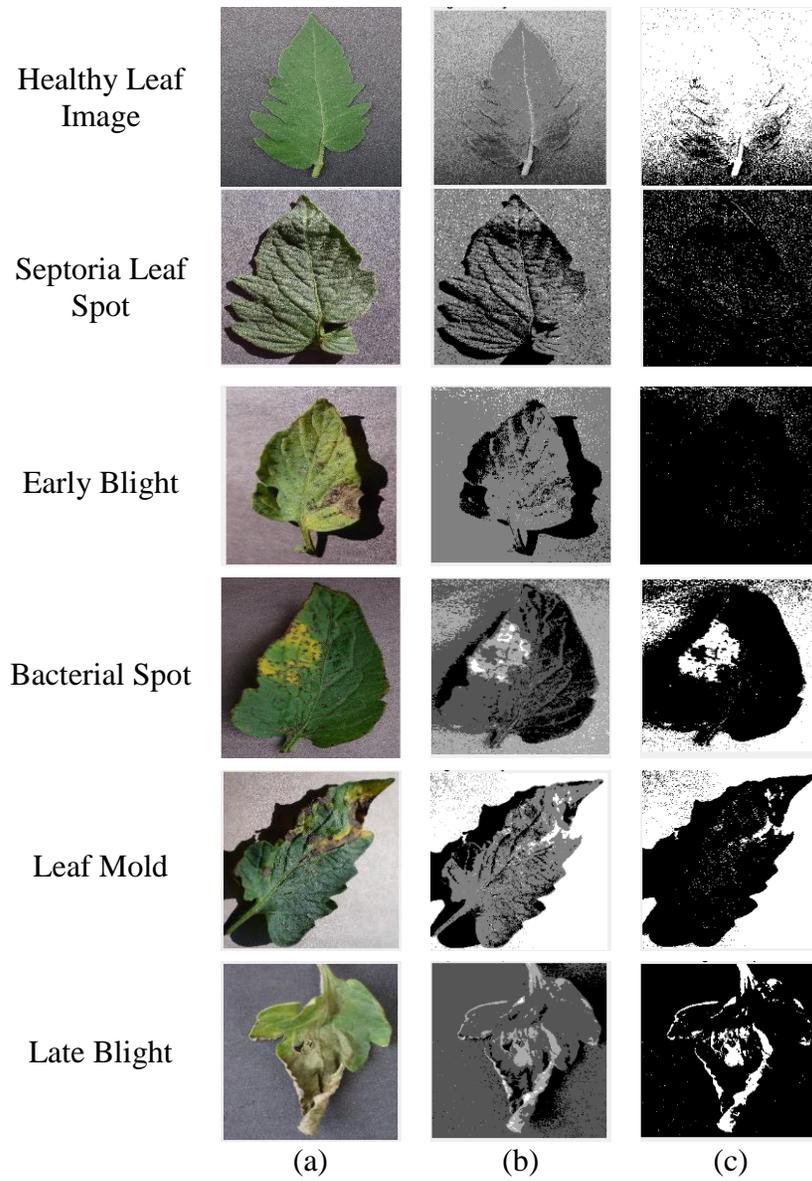


Figure 4. (a) Healthy leaf image (b) Segmented by K-means (c) Segmented by modified K-Means

Table 2. Feature extracted using GLCM feature extraction method

| Type of Leaf | Contrast | Energy | Entropy | Homogeneity |
|----------------------------|----------|--------|---------|-------------|
| Healthy images | 3.5794 | 0.7875 | 0.4832 | 0.9284 |
| Bacterial spot disease | 5.1853 | 0.6025 | 0.7873 | 0.8963 |
| Early blight diseases | 7.2169 | 0.4605 | 1.0057 | 0.8557 |
| Late blight disease | 6.9600 | 0.5121 | 0.9345 | 0.8608 |
| Leaf mold disease | 3.2125 | 0.6414 | 0.6972 | 0.9358 |
| Septoria leaf spot disease | 4.6759 | 0.8134 | 0.4133 | 0.9065 |

$$Accuracy = \frac{\text{Number of data correctly classified}}{\text{Total number of data}}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{FP + FN}$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

The precision of traditional K-means and SVM classification algorithms was 0.82. It was improved from 0.82 to 0.87 by using K-Means and ANFIS classification algorithm. The precision further improved from 0.87 to 0.92 by using modified K-Means and SVM Classification algorithm. The highest precision of 0.96 was attained using modified k-means and ANFIS classification algorithm.

The recall of traditional K-means and SVM classification algorithms was 0.84. It was improved from 0.84 to 0.86 by using K-Means and ANFIS classification algorithm. The recall is further improved from 0.86 to 0.90 by using modified K-Means and SVM Classification algorithm. The highest recall of 0.96 was attained using modified k-means and ANFIS classification algorithm.

The specificity of traditional K-means and SVM classification algorithms was 0.97. It was improved from 0.97 to 0.98 by using K-Means and ANFIS classification algorithm. The specificity is same by using modified K-Means and SVM Classification algorithm. The highest specificity of 0.99 was attained using modified k-means and ANFIS classification algorithm. The percentage of accuracy of traditional K-means and SVM classification algorithms was 85.1%. It was improved from 85.1% to 88.7% by using K-Means and ANFIS

classification algorithm. The accuracy is further improved from 88.7% to 92.8% by using modified K-Means and SVM Classification algorithm. The highest accuracy of 98.60% was attained using modified k-means and ANFIS classification algorithm. The results are clearly shown in Table 3.

Table 3. Performance Analysis of different classifiers

| Classification Type | Precision | Recall/ Sensitivity/ TPR | Specificity | Accuracy (%) |
|--|-----------|--------------------------------|-------------|--------------|
| SVM with K-Means segmentation | 0.82 | 0.84 | 0.97 | 85.1 |
| ANFIS with K-Means segmentation | 0.87 | 0.86 | 0.98 | 88.7 |
| SVM with Modified K-Means Segmentation | 0.92 | 0.90 | 0.98 | 92.8 |
| ANFIS with Modified K-Means Segmentation | 0.96 | 0.95 | 0.99 | 98.60 |

Discussion

The accuracy of detecting disease in tomato plant is 92.7% by Al Bashish *et al.* (2011) using K-Means segmentation and Neural network-based classification. The accuracy of detecting rice disease is 95% by Rath and Meher (2020) using Radial basis function neural network (RBFNN). The accuracy of detecting Apple disease 78.80% by Jiang *et al.* (2019) using INAR-SSD. The accuracy of detecting apple disease is 95.94% by Dubey and Jalal (2015) using K-means segmentation, colour, shape and texture feature extraction and M-SVM classification. The accuracy of detecting disease in cucumber is 98.08% by Khan *et al.* (2020) using Sharif saliency-based segmentation, VGG-19 & VGG-M feature. The accuracy of detecting diseases in banana is 97.5% by Athiraja and Vijayakumar (2020) using Threshold classification and ANFIS classifier. The highest accuracy of 98.60% was attained using our proposed Modified K-Means with ANFIS classification method.

This work mainly focused on the prediction of tomato leaf disease with the proposed modified K-means segmentation algorithm. In this modified K-Means segmentation algorithm it is not necessary to initialize cluster number and cluster center. Our work also compared the accuracy of different classifiers. Among the classifiers, ANFIS provides more accuracy than others, and our proposed Modified K-means and ANFIS classifier produced a better accuracy of 98.60%. The proposed method produces better result than result

produced by traditional K-Means by Tichkule and Gawali (2016) and Al Bashish *et al.* (2011). Traditional K-Means is a semi-automatic algorithm since users have to enter cluster number and cluster center, to overcome this fully automatic algorithm based on K-Means was proposed and achieve better result. Athiraja and Vijayakumar (2020) have used ANFIS classifier to detect banana leaf disease whereas our proposed method used improved ANFIS that is PSO based ANFIS to improve the accuracy from 97.5% to 98.60%.

Acknowledgements

The author would like to offer particular thanks to Dr. H. Vijayaraghavan, Professor (Crop Physiology) and Dr. S. Senthil Kumar, Assistant Professor (Plant Pathology) of Adhiyamaan college of Agriculture and Research, Hosur, Tamil Nadu, for their valuable support.

References

- Al Bashish, D., Braik, M. and Bani-Ahmad, S. (2011). Detection and classification Of Leaf Diseases using K-means based segmentation and Neural-networks-based Classification. *Information Technology Journal*, 10:267-275.
- Athiraja, A. and Vijayakumar, P. (2020). Banana disease diagnosis using computer vision and machine learning methods. *Journal of Ambient Intelligence and Humanized Computing*, DOI:10.1007/s12652-020-02273-8.
- Dubey, S. R. and Jalal, A. S. (2015). Apple disease classification using color, texture and shape features from images. *Signal, Image and Video Processing*, 10:819-826.
- Jiang, P., Chen, Y., Liu, B., He, D. and Liang, C. (2019). Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access*, pp.59069-80.
- Khan, M. A., Akram, T., Sharif, M., Javed, K., Raza, M. and Saba, T. (2020). An automated system for cucumber leaf diseased spot detection and clasification using improved saliency method and deep features selection. *Multimedia Tools and Applications*, 79:18627-18656.
- Nirmalakumari, K., Rajaguru, H. and Rajkumar, P. (2020). Plant leaf disease detection using modified segmentation process and classification. *Conference in Advances in Intelligent Systems and Computing*, pp.442-452.
- Rath, A. K. and Meher, J. K. (2020). Disease detection in infected plant leaf by computational method. *Archives of Phytopathology and Plant Protection*, 52:1348-1358
- Santhosh, Kumar, S. and Raghavendra, B. K. (2019). diseases detection of various plant leaf using image processing techniques: A review. In: *IEEE 5th International Conference on Advanced Computing & Communication Systems*, pp.313-316.
- Shrivastava, S., Singh, S. K. and Hooda, D. S. (2016). Soybean plant foliar disease detection using image retrieval approaches. *Multimedia Tools and Applications*, 76:26647-26674.
- Tichkule, S. K. and Gawali, D. H. (2016). Plant diseases detection using imag processingtechniques. *International Conference on Green Engineering and Technologie*, DOI: 10.1109/GET.2016.7916653.
- Youwen, T., Tianlai, L. and Yan, N. (2008). The recognition of cucumber disease based on image processing and support vector machine. *IEEE Congress on Signal and Image Processing*, pp.262-267.

(Received: 20 October 2020, accepted: 30 April 2021)