International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 896 – 904

Early identification of Tomato Plant Leaf Diseases using Clustering and Neural Networks

Dr. S. Raju[1], Dr. R. P. Ram Kumar[2], Dr. P. Santhi[3] A. Gowri[4]

[1]Professor, Department of Information Technology,
Mahendra Engineering College (Autonomous), Namakkal, India.
[2]Professor, Department of Computer Science and Engineering,
Malla Reddy Engineering College (Autonomous), Secunderabad, India.
[3]Professor in Education, Kongunadu College of Education, Namakkal, India.
[4]Assistant Professor, Department of Information Technology,
Hindusthan College of Arts and Science, Coimbatore, India

Abstract

Agriculture aims at cultivating crops to gratify the human needs. A study claims that nearly 50% of crops are being affected with diseases in early stages. Even though there exists many traditional approaches to eradicate the early stage diseases, the implication of digital technology lays a foundation in a efficient manner. Early diagnosis not only reduces time and money, also avoids spread of diseases to other crops, in turn improving the overall production. By considering this fact in the mind (as the focal area), the paper proposes an approach to categorize diseases in the tomato plant leaf at an early stage(s). The proposed method adopts clustering technique to cluster the disease affected regions and extract the features from those regions subject to classification process. The techniques adopted for clustering, feature extraction and classification are K-Means, Discrete Wavelet Transform (DWT) followed by Gray Level Co-occurrence Matrix (GLCM) and Neural Networks (NN) respectively. The proposed approach is assessed on the standard dataset images from Kaggle repository. Among various classes of leaf categories, randomly 50 images are chosen from each category to evaluate the proposed method. Experimental findings unfold that the proposed disease identification plays a vital role in early stages that the existing state-of-the-algorithms.

Keywords: *K-Means, DWT, GLCM, Neural Networks, Tomato leaf disease, feature extraction*

1. Introduction

The planet earth is crammed with plant life irrespective of the location in the world. Plant life, one of the parts in eco-system, makes the world sustain. Unlike various land forms-hills, mountains, plains, plateaus and basins, the plants that grow in these land forms also show discrepancy. An apple tree that grows in temperate zone cannot be grown in tropical zone. The flora depends upon the land form of specific region. To develop a conservative and sustainable eco-system for future generation, humanity should preserve the flora of each and every landform.

The humanity is dependent on agriculture for its daily livelihood. The boom of computer vision in fields of communication, science and technology, transportation ended up with numerous contemporary techniques such as Biometrics, Artificial intelligence, drones and many more. Computer vision needs to be extended in the field of agriculture too.

Agriculture aims at cultivating quality crops to gratify the human needs. In turn, the crops are qualitative if and only if the plants are not diseased. 50% of overall plant life in agriculture is diseased. Certain traditional technique in eradication of the diseases is suitable for meager plant diseases. Identification of diseases in early stage avoids further spread of disease in plants. At this

juncture, Image Processing Technique contributes production of quality plants to achieve increased yield.

The diseases affected in plants might be due to biotic or abiotic factors. For instance, symptom like disease reflection in leaf of the plants (which is not the only symptom) are well analyzed by a farmer based on prior experience. Other symptoms that aid in identification include soil analysis and pH measurement, microscopic examination, and microscopic methods [4]. Early diagnosis of diseases, help the farmers to

- Identify disease causing agent.
- Minimize waste of time and money, and
- Prevent irreparable damage even before visualizing.

To achieve aforesaid advantages, scientific diagnosis techniques such as Multilayer Perceptron (MLP) Neural Network , Thresholding, Dual-segmented regression, Color-based, Quantification, Fuzzy-logic, Feature-based, Knowledge-based, Discriminant analysis, Region growing techniques, Third party image processing packages, Support Vector Machine and many more are used [5].

2. Existing Methods

The following section summarizes the salient features of existing (tomato) plant leaf classification disease

Hiteshawari Sabrol and Satish Kumar [6] studied Gray-Level Spatial Dependence-based approach for Feature Extraction to recognize the plant disease with Adaptive Neuro Fuzzy classifier to classify the diseased plants. When the method was tested on TDPS1.0 and BPDS 1.0 dataset recognition accuracy of 90.7% and 98% was achieved for tomato and brinjal/eggplant, respectively. Ning Fu, Chong Wang, and Xiaowen Ji [7] examined corn leaf spot disease through Fuzzy technology. The corn lesions were clearly identified from the sample images which were classified as Class A –maize leaf spot identification, Class B – corn leaf spot disease recognition, Class C – disease identification by patch superimposition and the recognition accuracy was 89.76, 87.43, and 85.6 respectively. The authors measured real time data samples including plant foliage via image processing thus enhancing the efficiency of the study.

Amrita Tulshan and Natasha Raul [8] identified plant leaf disease using Machine Learning – KNN classifier the type of plant disease were also identified. The plant leaf images underwent pre-processing, segmentation and feature extraction stages and the outputs of every stage processed in KNN classifier by the textural feature based algorithm classified the diseased leaf as well as its type. The results of the algorithm used for the study when compared with SVM classifier yielded 98.56% accuracy for 75 leaf images. Sammy Militante, Bobby Gerardo and Nanette DIonisio [9] utilized Deep learning to detect and recognize the diseases in plant leaves. 35,000 images of the Plant Village Repository were used for the testing. The images acquired were resized into 96X96 resolutions for processing. The multi layered convolution neural network classified the infected diseases with the accuracy rate of 96.5%. The model also yielded more than 99% accuracy result for plants like tomato, grape and apple plants.

Sherly Annabel, Annapoorani and Deepalakshmi [10] reviewed machine learning techniques for detecting the plant leaf disease followed by classification. The authors' detected disease in various plant leaves such as rice, grape, cotton, soybean, sugarcane via machine learning algorithms like SVM, Multi SVM, K-Means clustering, Fuzzy C-Means, CNN, ANN, Back propagation neural network, Decision tree, Multi layer perceptrons and many more. The comparative chart designed by the authors revealed the type of disease detected in the respective methods. Further, the comparison of classification algorithms not only disclosed the accuracy of algorithms but also the pros and cons of using each algorithm. Akshay Kumar and Vani [11] detected the diseases affected from tomato leaf images by means of Convolutional Neural Network (CNN). 14,803 images from Plant village dataset were studied. Among various architectures in CNN, LeNet, VGGNet, ResNet50 and Xception architectures were investigated. During investigation, VGGNet achieved an accuracy rate of 99.25%.

Rath and Meher [12] detected brown spot and blast diseases in rice plants by the method of computation. The dataset images include normal, brown spot and blast diseased rice plants obtained from the crop fields and other public domains among which 100 images were used as the training dataset. PCA preprocesses the images, Gabor Filtering technique extracted shape and texture features and finally to detect the diseases Radial Basis Function Neural Network classifier was used. The performance of the aforesaid steps when evaluated using metrics such as accuracy, recall and precision was 95%, 95% and 97%, respectively. Santhana Hari, Sivakumar, Renuga, Karthikeyan and Suriya [13] identified and detected plant diseases from the images through the technique of Convolutional Neural Network. The put forth Plant Disease Detection Neural Network (PDDNN) has 16 layers like input layer, convolutional, maxpooling, ReLu activation layer and fully connected layer. The dataset includes defected images of various plants and 400 images were treated as training dataset. Accuracy rate of 86% was achieved when the performance was evaluated.

Rishabh Yadav, Yogesh and Sushama [14] exploited Particle Swarm Optimization technique to identify, detect and classify diseases that occurs in plant leaves. The acquired 8750 plant leaf images from Plant Village dataset were pre-processed and resized to 250X250 pixel size. With the aid of Alexnet, the features were extracted from the resized images and passed into PSO to deduce and optimize the extracted features. Finally, the leaf diseases were classified by the classifiers like XGBoost, Random Forest, KNN and SVM classifiers among which SVM classifier classified the images with the cross validation accuracy of 99.07%. Mosin, Bhavesh and Krina [15] utilized drones based Precision Farming method in the field to identify disease affected area and minimized the usage of pesticide based on leaf category such as good, bad or average. To categorize the leaf images from drones, the inception model in CNN trained the dataset. The model achieved accuracy of 99% on 2600 images procured from internet and local farms.

3. Materials and Methods

This section unfolds the procedure for early disease identification of tomato plant leaves adopted in the proposed approach and incorporates the clustering based and feature selection methods for extracting the disease affected regions of the tomato plant leaf. The plant village dataset is collected from [1], which consists of ten class of tomato leaf classes namely, (1) class 1: leaf affected with target spot (2) class 2: leaf affected with mosaic virus (3) class 3: leaf affected with yellow leaf curl virus (4) class 4: leaf affected with bacterial spot (5) class 5: leaf affected with early blight (6) class 6: healthy leaf (7) class 7: leaf affected with late blight (8) class 8: leaf affected with mold (9) class 9: leaf affected with Septoria leaf spot and (10) class 10: leaf affected with two spotted spider mite disease. Table 1 shows the dataset specification details with respect to tomato leaf diseases. Various classes of tomato plant leaves from Class 1 to Class 10 is illustrated from (a) to (j), respectively. Figure 1 shows few tomato leaves affected various diseases collected from [1].

Table 1. Dataset Description

Sl . No	Class Names	Total Count	No. of Images chosen randomly
1	Target Spot	1404	50
2	Tomato mosaic virus	373	50
3	YellowLeaf__Curl_Virus	3209	50
4	Bacterial_spot	2127	50
5	Early_blight	1000	50
6	Healthy	1591	50
7	Late_blight	1909	50
8	Leaf_Mold	952	50
9	Septoria_leaf_spot	1771	50

10	Spider_mites_Two_spotted_spider_mite	1676	50
	Total	**16012**	**500**

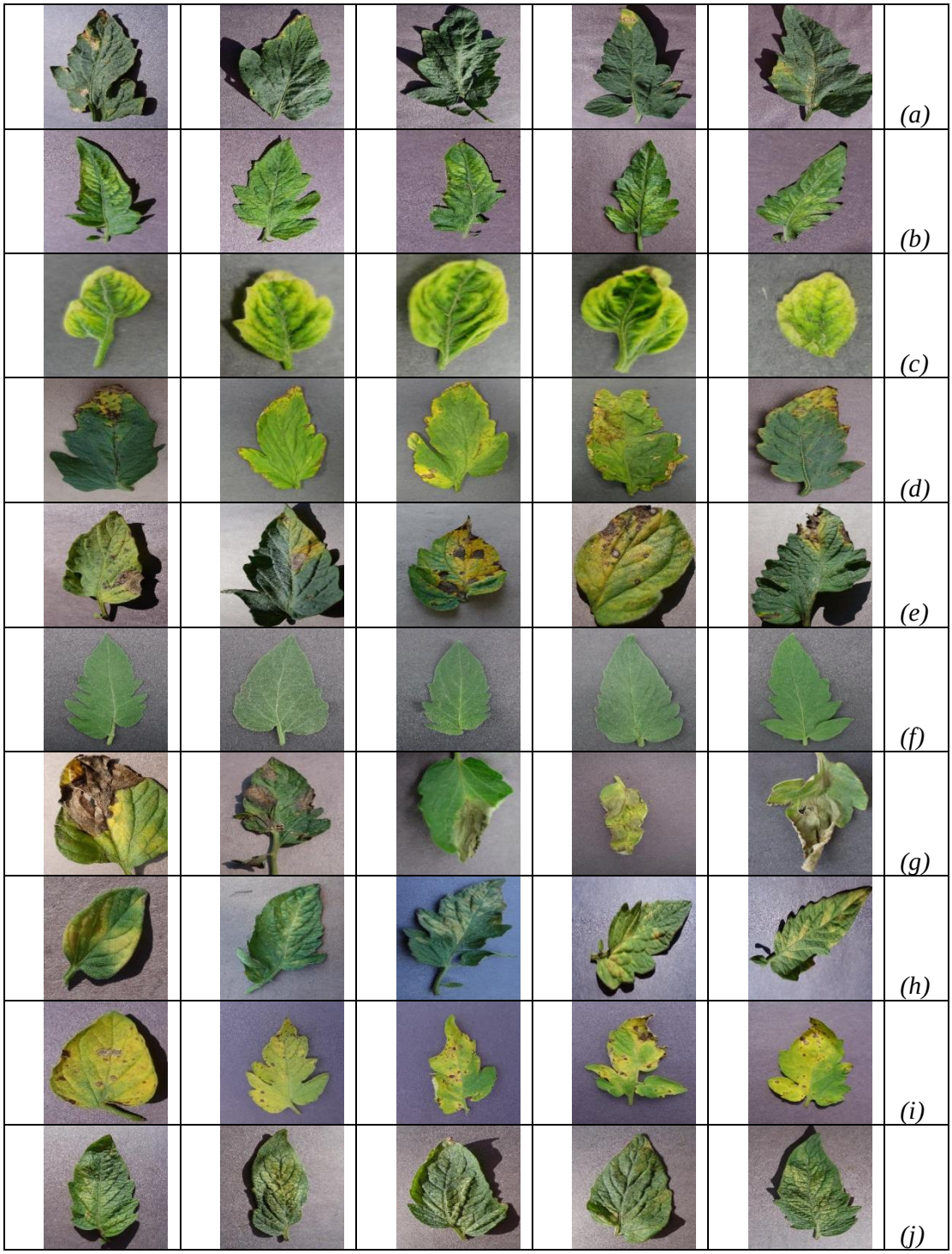

Figure 1. Samples tomato plant leaves affected with various diseases collected from [1]

3.1 Approach

The proposed method has three steps namely, (1) Clustering (2) Feature Extraction (FE) and (3) Classification. The techniques adopted in the proposed method are K-Means for Clustering, Discrete Wavelet Transform (DWT) and GLCM for FE and finally the Neural Networks (NN) for classification. Figure 2 shows the pseudo code of the proposed tomato leaf disease detection method.

Input: Leaf Image

Output: Disease Category

Read the Input Image.

2. Cluster the affected region using K-Means Clustering Technique.
3. From the resultant clustered image, choose the leaf affected cluster area and repeat the following:
   a. Extract the DCT features.
   b. Extract GLCM features from the resultant image.
   c. Apply the Neural Networks (with the specified features) on the extracted GLCM features.
   d. Maximum value in the corresponding row (index) determines the corresponding type of disease.

Figure 2. Pseudo Code for the

Proposed Tomato Leaf Disease Detection Method

The simplest clustering technique, K-Means Clustering, is adopted in the proposed method to cluster the disease affected regions in the tomato plant leaves. It includes the following steps summarized from [16]:

a) Initially, random points are selected (in this case, the product of rows and columns in the input image is considered) as cluster center.

b) Number of random centroid is chosen as ten.

c) Based on the distance between the chosen centroid and the other points, those points are clustered to the concern centroid(s) with respect to the minimum distance.

d) The cluster center(s) are revised in accordance with the mean of designated observations.

e) The above two steps (step (c) and step (d)) are repeated till the convergence occurs.

Finally, the outcome of K-Means Clustering approach is the clustered regions of disease affected and unaffected regions in the input tomato plant leaf. Figure 3 shows the resultant of clustering operations for a sample image from [1], where the Figure 3 (a), (b), (c), (d) and (e) represents the input image, clustering operations, Clustered Region 1 (CR1 – the clustering of leaf portion affected with disease), CR2 - unaffected leaf portion and CR3 - the clustered leaf, respectively.

Input Image (a) | **Clustering Operation (b)** | **CR1 (c)** | **CR2 (d)** | **CR3 (e)**

Figure 3. Resultant of Clustering operations for a sample image

In the FE method, the proposed method adopts DWT and GLCM approaches for extracting the predominant features of the input tomato leaf. In general, one dimensional (1D) DWT perform operations on vector, which is of length 2x, (where x ranges from 2, 3, and so on) and produces a transformed vector of same size. Initially, the input vector is filtered with Low Pass Filter (LPF) and High Pass Filter (HPF) of specific size and stored in the first half and second half of the transformed vector, respectively. Thus, ends up with one-level wavelet transform of input vector. Subsequently, these steps are repeated for three iterations, resulting in a three level 1-D DWT. With respect to two dimensional (2D) DWT, one level 1-D DWT is applied along the rows followed by columns subsequently. As a result, low-low, low-high, high-high and high-low bands of transformed images are generated. Subsequently, these steps are repeated for three iterations, resulting in a three level 2-D DWT [2]. Followed by DWT, GLCM and Haralick features are extracted. The features examined in the proposed method are Contrast, Correlation, Energy, Homogeneity, Mean, Variance, Entropy, Standard Deviation, Smoothness, Kurtosis, Skewness, Inverse Difference Moment (IDM) and Root Mean Square (RMS) Values and determined using the Equations from (1) to (13), respectively.

$$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \tag{1}$$

$$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right] \tag{2}$$

$$\sum_{i,j=0}^{N-1} P_{i,j}^2 \tag{3}$$

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{4}$$

$$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}),\ \mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j}) \tag{5}$$

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2,\ \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j-\mu_j)^2 \tag{6}$$

$$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j}) \tag{7}$$

$$\sigma_i = \sqrt{\sigma_i^2},\ \sigma_j = \sqrt{\sigma_j^2} \tag{8}$$

$$R = 1 - \frac{1}{1+\sigma^2} \tag{9}$$

$$K = \{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[\frac{p(i,j)-\mu}{\sigma}]^4\} - 3 \quad (10)$$

$$S = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[\frac{p(i,j)-\mu}{\sigma}]^3 \quad (11)$$

$$H = \sum_{i,j}\frac{P(i,j)}{1+|(i-j)|} \quad (12)$$

$$y = \sqrt{\frac{\sum_{i=1}^{M}|u_{ij}|^2}{M}} \quad (13)$$

where,

N and P_{ij} denotes the rows/columns and probability of points,

- σ, σ_i and σ_j denotes the standard deviation (SD) and SD of rows and columns,
- μ, μ_i and μ_j denotes the mean (M), and M of rows and columns,
- R, K and S denotes the contrast measure, kurtosis and skewness,
- H and y denotes the IDM and RMS

Table 2 illustrates the dimension variations of input image during the feature extraction phase.

Table 2. Image Dimensions after Features Extraction

Sl. No	Input image dimensions (in pixels)	Image dimensions after applying (in pixels)	
		DWT	GLCM
1	256 x 256	38 x 152	1 x 13

Followed by the feature extraction, classification of plant leaf diseases is accomplished through the Neural Network (NN) with following specifications:

- Number of sample selected for training = 80%
- Number of sample selected for testing = 10%
- Number of sample selected for validation = 10%
- Number of hidden layers (chosen) = 15
- Training function adopted in the pattern recognition network is Bayesian Regularization Back Propagation ('trainbr'). The significance of this network lies in updating the weight and subsequent bias values with respect to Levenberg-Marquardt Optimization technique. Moreover, it reduces the consolidation of weights and squared error [3] and then regulates the appropriate combination to improvise the network performance. This type of network is referred as Bayesian Regularization (BR).

4. Experimental Results

The proposed method of tomato plant leaf disease detection approach is evaluated on the sample leaf images collected from plant village dataset [1]. The dataset consists of ten classes of tomato leaf categories (nine affected classes and one healthy class) resulting in 16,012 number of images. Randomly, 50 images are selected from each category resulting to 500 images for evaluating the proposed method. As an initial step, K-Means Clustering is applied on the input image and clustered images are resulted for leaf regions and disease affected regions. Upon choosing the disease affected region cluster, the feature extraction process is carried out using 2D-DWT and GLCM approaches. Thus, the input image of uniform size (256 x 256) pixels are subjected to feature reductions of 38 x152 pixels and 1 x 13 pixels due to DWT and GLCM approaches, respectively. Later, resultant features are classified using NN of specified specifications. Table 3 unfolds the proposed method's performance with respect to classification accuracy and metrics used to evaluate are Sensitivity and Specificity.

Equations 14, 15 and 16 are used to determine the Sensitivity, Specificity and Classification Accuracy respectively.

$$\text{Sensitivity} = \frac{TP}{TP+FN} * 100 \tag{14}$$

$$\text{Specificity} = \frac{TN}{TN+FP} * 100 \tag{15}$$

$$\text{Classification Accuracy (CA)} = \frac{TP+TN}{TP+TN+FP+FN} * 100 \tag{16}$$

Here, the variables TP, TN, FP and FN denote the True Positive, True Negative, False Positive and False Negative respectively. Further, the best training performance is 5.029e-10 at 45 epochs. When compared to the existing tomato plant leaf disease classification approaches namely, [6], [7], [8] and [13], the proposed classification has significantly higher performance in terms of classification accuracy.

Table 3. Performance Measures

Number of images	TP	FN	TN	FP	Sensitivity	Specificity	Classification Accuracy
500	445	5	48	2	98.88	96.0	98.6

5. Conclusion

The proposed method of tomato plant leaf disease identification approach has three phases namely (1) Clustering of affected and healthy regions using K-Means approach (2) Feature Extraction through DWT and GLCM and (3) Classification using NN. The proposed work was evaluated on Plant Village Dataset with randomly selected 500 images comprising of ten classes. When compared with the performance of the existing approaches, the proposed method has significant classification accuracy. The proposed model can be enhanced with the real dataset and evaluated with different NN architectures (such as AlexNet and GoogLeNet). Further, the scope may be extended to imperatively suggest the remedial measures for the concern leaf diseases.

References

[1] "PlantVillage Dataset", https://www.kaggle.com/emmarex/plantdisease.

[2] "Introduction to the Discrete Wavelet Transform (DWT)", https://mil.ufl.edu/nechyba/www/eel6562 /course_materials/t5wavelets/intro_dwt.pdf. [Accessed: 02-Jul-2020].

[3] "Bayesian Regularization Backpropagation - MATLAB Trainbr - MathWorks India", https://in.math works.com/help/deeplearning /ref/trainbr.html

[4] "Plant Disease Diagnosis", https://www.slideshare.net/sobhysalama/plant-disease-diagnosis. [Accessed: 02-Jul-2020].

[5] J. G. Arnal Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases", SpringerPlus 2, 660, **(2013)**, pp. 1-12.

[6] Hiteshwari Sabrol and Satish Kumar, Plant leaf disease detection using Adaptive Neuro-Fuzzy Classification, K. Arai and S. Kapoor (Eds.), Proceedings of the CVC 2019, AISC, vol. 943, Springer International Publishing, **(2020)**, pp. 434-443.

[7] Ning Fu, Chong Wang, and Xiaowen Ji, "Study on visual detection device of plant leaf disease," Proceedings of the 2019 IEEE Int. Conf. Mechatronics Autom. ICMA 2019, **(2019)**, pp. 86-90.

[8] Amrita S. Tulshan and Nataasha Raul, "Plant leaf disease detection using Machine Learning," Proceedings of the 10th Int. Conf. Comput. Commun. Netw. Technol., ICCCNT 2019, **(2019)**, pp. 1-6.

[9] Sammy V. Militante, Bobby D. Gerardo, and Nanette V. DIonisio, "Plant leaf detection and disease recognition using Deep Learning," Proceedings of the 2019 IEEE Eurasia Conf. IOT, Commun. Eng. ECICE 2019, **(2019)**, pp. 579-582.

[10] L. Sherly Puspha Annabel, T. Annapoorani, and P. Deepalakshmi, "Machine Learning for plant leaf disease detection and classification - A Review," Proceedings of the 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019, **(2019)**, pp. 538-542.

[11] Akshay Kumar and M. Vani, "Image based tomato leaf disease detection," Proceedings of the 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019, **(2019)**, pp. 1-6.

[12] A. K. Rath and J. K. Meher, "Disease detection in infected plant leaf by computational method," Archives of Phytopathology and Plant Protection, vol. 52, no. 19–20, **(2019)**, pp. 1348-1358.

[13] S. Santhana Hari, M. Sivakumar, Dr. P. Renuga, S. Karthikeyan and S.Suriya, "Detection of Plant Disease by leaf image using Convolutional Neural Network", Proceedings of the 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), **(2019)**, pp. 1-5.

[14] R. Yadav, Yogesh Kumar Rana, Sushama Nagpal, "Plant Leaf Disease Detection and Classification Using Particle Swarm Optimization", In: É. Renault, P. Mühlethaler, S. Boumerdassi, (Eds), Proceedings of the Machine Learning for Networking (MLN) 2018, Lecture Notes in Computer Science, vol. 11407, Springer, Cham, **(2019)**, pp. 294-306.

[15] Mosin Hasan, Bhavesh Tanawala and Krina J. Patel, "Deep Learning Precision Farming: Tomato Leaf disease Detection by Transfer Learning", Proceedings of the 2nd International Conference on Advanced Computing and Software Engineering (ICACSE-2019), **(2019)**, pp. 171-175.

[16] "Step By Step To K-Means Clustering - Healthcare.ai," Healthcare.ai, https://healthcare.ai/ step-step-k-means-clustering/. [Accessed: 02-Jul-2020].