Machine Learning for Plant Leaf Disease Detection and Classification – A Review

L. Sherly Puspha Annabel, *Member, IEEE,* T. Annapoorani and P. Deepalakshmi

***Abstract*—Plants are considered to be important as they are the source of energy supply to mankind. Plant diseases can affect the leaf any time between sowing and harvesting which leads to huge loss on the production of crop and economical value of market. Therefore, leaf disease detection plays a vital role in agricultural field. However, it requires huge manpower, more processing time and extensive knowledge about plant diseases. Hence, machine learning is applied to detect diseases in plant leaves as it analyzes the data from different aspects, and classifies it into one of the predefined set of classes. The morphological features and properties like color, intensity and dimensions of the plant leaves are taken into consideration for classification. This paper presents an overview on various types of plant diseases and different classification techniques in machine learning that are used for identifying diseases in different plant leaves.**

***Index Terms*—Artificial Neural Network, Classification, Disease Detection, Support Vector Machine,Machine Learning.**

I. Introduction

INDIA is a fast developing country and agriculture is the back bone for the country's development in its early stage. However, agricultural field faces lots of hurdles including huge loss in the crop production. Plant leaf diseases are one of the important reasons for the loss in the production and plant leaf disease identification is also very difficult in agriculture field. Naked eye method is a traditional method of identifying the diseases which involves huge man power, inaccurate, time consuming and not applicable for larger fields. In addition, it is very expensive as it requires continuous monitoring by the experts. Hence, machine learning; a reliable prediction methodology is used for detecting various diseases of plant leaves caused by fungus, bacteria and virus. However, disease prediction using classification algorithms appears to be a difficult task as the accuracy varies for different input data. In this paper, several research contributions related to various plant leaf diseases detection using different classification algorithms are reviewed and compared. The classification of various plant diseases, classification algorithms in Section II & III and conclusion are discussed in section IV respectively.

L. Sherly Puspha Annabel, T. Annapoorani and P. Deepalakshmi are with the Department of Information Technology, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India (Corresponding Author: e-mail: sherlyannabel@gmail.com)

II. Classification of Plant Diseases

The plant leaves are affected by bacterial, fungal and viral diseases which include leaf rust, powdery mildew, bacterial blight, Downey mildew, brown spot etc. Fig. 1 illustrates the classification of the bacterial, fungal and viral diseases. J. D. Pujari, R. Yakkundimath, and A. S. Byadgi applied Artificial Neural Network, Probabilistic Neural Network, and Support Vector Machine for vegetable crops, commercial crops and cereal crops respectively for disease detection [1-4]. Balasubramanian Vijayalakshmi and Vasudev Mohan applied Fuzzy-Relevance Vector Machine classifiers in which the inputs like training features and the labels are used for leaf disease detection [5-7]. X. Wang, M. Zhang, J. Zhu and S. Geng foretold Phytophthora infestans disease diagnosis on tomatoes by using Artificial Neural Networks [8-10]. Dong Pixia and Wang Xiangdong proposed an approach called Minimum Distance Classifier for recognizing cucumber leaf disease [11-15]. S. Arivazhagan, R. Newlin Shebiah, S. Ananthi and S. Vishnu Varthini proposed an algorithm for classifying diseases of plants including jackfruit, tomato, etc. by using Support Vector Machine classifier [16-18]. Table I demonstrates the comparative study of disease detection in different plant leaves.

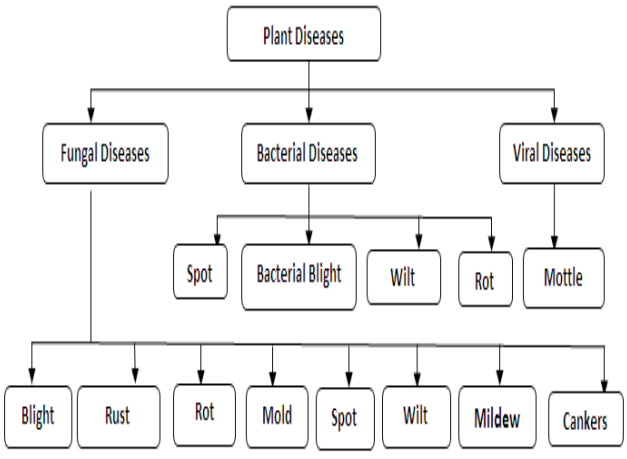

Fig. 1. Classification of Plant Diseases

978-1-5386-7595-3/19/$31.00 ©2019 IEEE

TABLE I
COMPARATIVE STUDY OF DISEASE DETECTION IN VARIOUS PLANT LEAVES

First Author, Year	Types of leaves	Algorithm	Types of disease detected
Harshal Waghmare, 2016 [6]	Grape leaf	Multiclass Support Vector Machine	Black rot, Downey mildew
T. H. Jaware, 2012 [7]	Cotton leaf	K-means Clustering	Early scorch, Tiny whiteness
Dheeb Al Bashish, 2010 [8]	Al-ghor area leaves	Neural Network Classifier	Cottony mold, Late scorch
A. A. Joshi, 2016 [9]	Rice leaf	Minimum Distance Classifier	Rice blast, Rice Sheath Rot
Bauer, 2009 [10]	Sugar beet leaf	Maximum Posterior Classification	Leaf rust, Powdery mildew
K. R. Gavhale, 2014 [11]	Citrus leaf	Support Vector Machine	Citrus canker, Anthracnose, Citrus greening
S. B. Jadhav, 2015 [12]	Soybean leaf	K-means Clustering	Bean Pod Mottle
K. Renugambal, 2015 [13]	Sugarcane leaf	Support Vector Machine	Red rot, Leaf spot, Sugarcane mosaic virus, Yellow spot, Brown spot
D. Luna, 2017 [14]	Herbal plants	Artificial Neural Network	Diseases in herbal leaves like akapulko, amplaya etc.
Sandika Biswas, 2014 [15]	Potato leaf	Fuzzy C-means Clustering Neural Network	Potato leaf blight
Bin Liu, 2018 [16]	Apple leaf	Deep Convolution Neural Network	Mosaic Rust, Brown spot, Alternaria leaf spot
Habibollah Agh Atabay, 2017 [17]	Tomato leaf	Convolution Neural Network	Early Blight, Late Blight, Septoria leaf spot
S. Gaikwad, 2016 [18]	Pomegranate leaf	Support Vector Machine	Bacterial Blight, Fruit Spot
Wan Mohd Fadzil, 2014 [19]	Orchid leaf	Pattern Classification	Black leaf spot, solar scorch
Jobin Francis, 2016 [20]	Pepper leaf	Back Propagation Neural Network	Berry spot, Quick wilt
Qin, 2016 [21]	Alfalfa leaf	Support Vector Machine	Leaf spot, Leaf Rust
Pooja Pawar, [22]	Cucumber Leaf	Artificial Neural Network	Powdery mildew, Downey Mildew
Kaur, 2016 [23]	-	Support Vector Machine	Bacterial Blight, Anthracnose
H. Al-Hiary, 2011 [24]	Plant Leaves (cotton)	K-Means Clustering, Neural Network	Tiny Whiteness, Cottony mold
Santanu Phadikar, 2008 [25]	Rice Leaf	Neural Network	Leaf blight, Leaf blast
Balasubramanian VijayaLakshmi, 2016	Norway Maple, White willow	Fuzzy Relevance Vector Machine	Leaf Type classification
Sun, 2017 [26]	Chinese Buckeye, Tulip Tree	Deep Learning	Large Scale classification of plants
John William Orillo, 2013 [27]	Rice Leaf	Artificial Neural Network	Level of Nitrogen content in healthy leaves
Mukherjee, 2017 [28]	Tulsi Leaf	Back Propagation Multi-Layer Perceptron Neural Network	Maturity of tulsi leaves
M. C. Ramya [29]	Mulberry Leaf	Decision Tree	Prediction of mulberry leaves for different soil types
Siddharth Singh Chouhan, 2018 [30]	Apple Leaf	Radial Basis Function Neural Network	Apple Rust, Early Blight
S. Arivazhagan, 2013	Tomato Leaf	Support Vector Machine	Yellow Spot, Late Scorch
X. E. Pantazi, 2019 [31]	Vine Leaf	One Class Classifier	Powdery Mildew

III. CLASSIFICATION ALGORITHMS

This section explains the classification algorithms in machine learning that are used for classifying diseases[19-24] in plant leaves. Its accuracy depends on the number of samples taken and varies according to the classification algorithms used. The classification algorithms are divided into supervised and unsupervised classification algorithms[25-27] Fig. 2 illustrates the various types of classification algorithms for plant leaf disease detection.

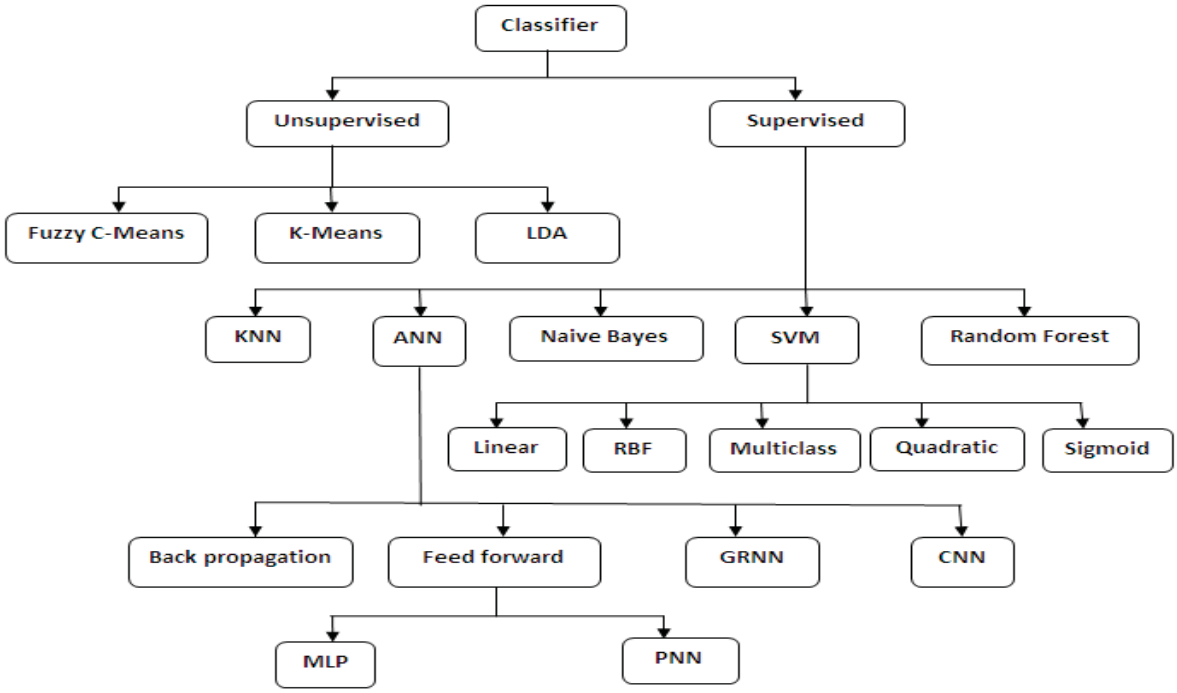

Fig. 2. Types of Classification Algorithms

A. Unsupervised Classification Algorithms

Fuzzy C-means is an iterative algorithm which helps to find the cluster centers that minimize a dissimilarity function and to handle the overlapped data efficiently. It gives better results in cases where data is incomplete or uncertain, but computation time is longer and it is sensitivity to noise. Fuzzy C-means clustering Neural Network[28] consists of unsupervised fuzzy clustering and supervised artificial neural networks which help in achieving more optimal results with relatively few data sets. K-means is an iterative learning helps in finding the cluster centers for each group and has no guarantee for optimum solution. It is easy to implement and computationally faster. But the number of cluster prediction is difficult. Principal Component Analysis [29] is an unsupervised technique helps in finding the most accurate data representation and maximizes the variance. Linear Discriminant Analysis finds the projection to a line and maximizes the component axes for class separation.

B. Supervised Classification Algorithms

K Nearest Neighbor is a used for statistical estimation and pattern recognition. It is easy, simple, flexible and robust to noisy training data but computation cost is higher. Artificial Neural Network uses forward propagation which is the heart of a neural network. Probabilistic Neural Network is a feed forward algorithm which is very faster and more accurate than multilayer perceptron network. Generalized Regression Neural Network is a supervised algorithm used for classification. Convolution Neural Network is a class of deep, feed-forward Artificial Neural Network which consists of input, output as well as multiple hidden layers, convolutional layers, pooling layers, fully connected layers and normalization layers. Pooling reduces the dimensionality of the features map by condensing [30] the output of small regions of neurons into a single output. Fuzzy-Relevance Vector Machine is effective in dealing with unbalanced data and reducing the effects of noise or outliers. Relevance Vector Machine is a machine learning technique that uses Bayesian inference for regression and probabilistic classification. In Support Vector Machine, the data points with maximum margin are chosen and separated by hyper plane. Radial Basis Function has three layers namely, input, hidden and output layer. It is used for function approximation which depends only on the distance from the origin. Random Forest performs both classification and regression which is also known as ensemble machine learning algorithm. It adapts divide and conquer approach used to improve the performance. It creates a forest with more decision trees which helps in strong prediction. Decision trees otherwise known as Classification and Regression Trees [31] are used, when dependent variable is continuous. Table II demonstrates the performance of classification algorithms for plant leaf disease detection.

TABLE II
OVERALL PERFORMANCE OF CLASSIFICATION ALGORITHMS FOR PLANT LEAF DISEASE DETECTION

First Author, Year	Classification Algorithms	Reported Accuracy	Pros	Cons	Future Research Direction
Sandika Biswa, 2014	FCM clustering and neural network	93%	Robust	Difficult in segmentation	To achieve accurate disease severity identification
A. A. Joshi, 2016	KNN classifier	87.02%	Simple Implementation, Learns complex models easily	High Computational complexity	a) To work with other kind of rice diseases. b) To work with other crops by slightly modifying the technique.
	Minimum Distance Classifier	89.23%			
John William Orillo, 2013	Back Propagation ANN	93.33%	Color based characterization of leaf can be extracted properly.	High computational cost	To attain more accuracy
Pooja Pawar, 2016	Artificial Neural Network	80.45%	Works well for more than one crop of different types.	Feature Selection is difficult.	To integrate Gabor filter for feature extraction
Harshal Waghmare, 2016	Multi class Support Vector Machine	96.6%	Performs accurate classification.	Accuracy improves only when the testing and training ratio increases.	To work with different kinds of diseases
Balasubramanian VijayaLakshmi, 2016	FRVM– Accuracy	99.87%	It produces better accuracy.	a) Segmentation is a challenging task. b) Leaves with like shape and size are tough to classify. c) Difficulty in classification of leaves with complicated backgrounds	a) To perform classification of leaves by including different leaf features b) To work with medicinal plants
	FRVM– Sensitivity	99.5 %	It is not sensitivity to noise.		
	FRVM– Specificity	99.9 %	It reduces time complexity and there is no limitation in speed and size.		
Dheeb Al Bashish, 2011	Neural Network	93%	Significantly accurate and very effective in leaf disease recognition, Reduces the computational complexity	Eliminates the intensity texture features	a) To automatically detect the severity of the leaf disease b) To develop better segmentation and classification algorithm
Kaur, 2016	Support Vector Machine with Ant Colony Optimization	96.77% to 98.42%	Detects fungal and bacterial diseases	Memory Intensive	a) To work with viral plant diseases b) To integrate the work with swarm intelligence
Mukherjee, 2017	Back Propagation Multi-Layer Perceptron Neural Network	80%	Classifies medical plants like Kalmegh and Tulsi based on their morphological features	Dithering present at the edges makes the task tougher.	To work with other different kinds of medicinal plants
D. Luna, 2017	Artificial Neural Network	98.61%	Better Accuracy, Simple Implementation	Large Computational time.	a) To work with Philippine herbal and to identify under-utilized and native fruits leaves with more samples b) To consider location of the herbal medicine for distinct characterization
Bin Liu, 2017	Deep Convolution Neural Network	97.62%	Accurate, robust, prevents over fitting of the CNN model, high feature extraction capability	Difficulty in identifying the structure of the model	To work with high quality apple leaves images to attain more accuracy
Sun, 2017	Deep Learning Model	91.78%	Fast, robust, simple to implement	Less Scalability	To perform yield prediction, disease segmentation
H. Al-Hiary, 2011	Neural Network	83% to 94%	Provides Accuracy with less computational effort	Dependent on certain features of the leaf	To integrate genetic algorithm and neural network to estimate the severity of the leaf disease
K. Renugambal, 2015	Artificial Neural Network	85%	More feature analysis and segmentation techniques are applied	Accuracy is comparatively low	To focus on developing fuzzy optimization algorithms
	Linear SVM	91%			
	Non Linear SVM	94%			
Santanu Phadikar, 2008	Self-Organizing Map	92%	Simple and Computationally efficient	Image transformation in frequency domain does not offer better classification	To achieve more accuracy

IV. CONCLUSION

This paper reviews and summaries various techniques used for classifying and detecting various bacterial, fungal and viral plant leaf diseases. The classification techniques helps in automating the detection of plant leaf diseases and categorizing them centered on their morphological features. The future work of this paper focuses on identifying the mulberry plant leaf diseases with CNN as classifier. It is also intended to focus on increasing the recognition rate and classification accuracy of severity of leaf diseases by using hybrid algorithms.

REFERENCES

[1] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "Identification and classification of fungal disease affected on agriculture/horticulture crops using image processing techniques," *IEEE International Conference on the Computational Intelligence and Computing Research*, 2014.

[2] Balasubramanian Vijayalakshmi and Vasudev Mohan, "Kernel based PSO and FRVM: An automatic plant leaf type detection using texture, shape and color features," *Computer and Electronics in Agriculture*, vol. 125, pp. 99-112, 2016.

[3] X. Wang, M. Zhang, J. Zhu and S. Geng, "Spectral prediction of Phytophthora infestans infection on tomatoes using artificial neural network (ANN)," *International Journal of Remote Sensing*, pp. 1693–1706, 2008.

[4] Dong Pixia and Wang Xiangdong, "Recognition of Greenhouse Cucumber Disease Based on Image Processing Technology," *Open Journal of Applied Sciences*, vol. 3, pp. 27-3, Mar. 2013.

[5] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi and S. Vishnu Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Commission Internationale du Genie Rural(CIGR) journal*, vol. 15, no. 1, pp. 211-217, 2013.

[6] Harshal Waghmare, Radha Kokare and Yogesh Dandawate, "Detection and Classification of Diseases of Grape Plant Using Opposite Colour Local Binary Pattern Feature and Machine Learning for Automated Decision Support System," *3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2016.

[7] T. H. Jaware, R. D. Badgujar and P. G. Patil, "Crop disease detection using image segmentation," *Proceedings of Conference on Advances in Communication and Computing (NCACC'12)*, April, 2012.

[8] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad, "A Framework for Detection and Classification of Plant Leaf and Stem Diseases," *International Conference on Signal and Image Processing*, 2010.

[9] A. A. Joshi and B. D. Jadhav, "Monitoring and Controlling Rice Diseases Using Image Processing Techniques," *2016 International Conference on Computing, Analytics and Security Trends (CAST) College of Engineering Pune*, Dec 2016.

[10] Bauer, S. D. Korc and F. Forstner, "Investigation into the classification of diseases of sugar beet leaves using multispectral images," *Wageningen Academic Publishers,* pp. 229-238, 2009.

[11] K. R. Gavhale, U. Gawande and K. O. Hajari, "Unhealthy region of citrus leaf detection using image processing techniques," *International Conference on Convergence of Technology I2CT*, pp. 1–6, 2014.

[12] S. B. Jadhav and S. B. Patil, "Grading of Soybean Leaf Disease Based on Segmented Image Using K-means Clustering," *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 4, no. 6, Jan 2015.

[13] K. Renugambal and B. Senthilraja, "Application of image processing technique in plant disease recognition," *International journal of engineering research and technology*, vol. 4, no. 03, Mar. 2015.

D. Luna, G. Robert G. Renann, Baldovino, E. A. Cotoco, Anton Louise P. de Ocampo, C. Ira,Valenzuela, A. B. Culaba, and E. P. Dadios Gokongwei, "Identification of philippine herbal medicine plant leaf using artificial neural network," *9th IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1-8, 2017.

[14] Sandika Biswas, Bhushan Jagyasi, Bir Pal Singhy and Mehi Lal, "Severity Identification of Potato Late Blight Disease from Crop Images Captured under Uncontrolled Environment," *IEEE Canada International Humanitarian Technology Conference - (IHTC)*, 2014.

[15] Bin Liu, Yun Zhang, Dong Jian He and Yuxiang Li, "Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks," *Symmetry,* vol. 10, no. 11, 2017.

[16] Habibollah Agh Atabay, "Deep residual learning for tomato plant leaf disease identification," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 24, 2017.

[17] S. Gaikwad and K. J. Karande, "Image Processing Approach For Grading And Identification Of Diseases On Pomegranate Fruit," *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol. 7, no. 2, pp. 519-522, 2016.

[18] Wan Mohd Fadzil, Shah Rizam, R. Jailani and M. T. Nooritawati, "Orchid leaf disease detection using Border Segmentation technique," *IEEE Conference on Systems, Process and Control (ICSPC)*, vol.1, pp.168-179, Dec. 2014.

[19] Jobin Francis, D. Anto Sahaya Dhas and B. K. Anoop, "Identification of leaf diseases in pepper plants using soft computing techniques," *Conference on emerging devices and smart systems (ICEDSS),* pp. 168-173, 2016.

[20] Qin, F. Liu, D.X. Sun, B.D. Ruan, L. Ma and Z. Wang, "Identification of alfalfa leaf diseases using image recognition technology," *PLoS ONE*, vol.11, Dec. 2016

[21] Pooja Pawar, Varsha Turkar, and Pravin Patil, "Cucumber disease detection using artificial neural network," *IEEE International Conference on Inventive Computation Technologies (ICICT)*, vol. 3, pp. 1-5, 2016.

[22] Kaur, Iqbaldeep, Gifty Aggarwal, and Amit Verma, "Detection and Classification of Disease Affected Region of Plant Leaves using Image Processing Technique," *Indian Journal of Science and Technology*, vol. 9, no. 48, 2016.

[23] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, "Fast and accurate detection and classification of plant diseases," *Machine learning*, vol. 14, no. 5, 2011.

[24] Santanu Phadikar and Jaya Sil, "Rice disease identification using pattern recognition techniques," *11th IEEE International Conference on Computer and Information Technology*, pp. 420-423, 2008

[25] Sun, Yu, Yuan Liu, Guan Wang, and Haiyan Zhang, "Deep learning for plant identification in natural environment," *Computational intelligence and neuroscience*, vol. 2017, 2017

[26] John William Orillo, Gideon Joseph Emperador, Mark Geocel Gasgonia, Marifel Parpan, and Jessica Yang, "Rice plant nitrogen level assessment through image processing using artificial neural network," *IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM),* pp. 1-6. 2014.

[27] Mukherjee, Gunjan, Arpitam Chatterjee and Bipan Tudu, "Morphological feature based maturity level identification of Kalmegh and Tulsi leaves," *Third IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pp. 1-5, 2017

[28] M. C. Ramya, V. Lokesh V, T. N. Manjunath and R. S. Hegadi, "A Predictive Model Construction for Mulberry Crop Productivity," *International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015).*

[29] Siddharth Singh Chouhan, Ajay Kaul, Uday Pratap Singh and Sangeev Jain, "Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for identification and classification of plant leaf disease: An automatic approach towards plant pathology," *IEEE Access*, vol. 6, pp. 8852-8863, 2018

[30] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," *Computers and Electronics in Agriculture*, vol. 156, pp. 96-104, 2019.