# A Review: Plant Diseases Prediction Using Image Processing

Anusuiya, Ramesh Mishra

***Abstract*** **-** This survey literature discusses on mechanisms to early detect agricultural plant leaves' diseases for the quantitative and qualitative safety of the products using machine learning and image processing techniques. As a result, in this survey, different machine learning techniques, such as Artificial Neural Network, Support vector machine, Convolutional neural network, K-nearest neighbor, Nave bayes, Radial basis function, Self organizing map, Fuzzy inference system, minimum distance criterion, least square support vector machine, Feed forward neural network and learning vector quantization were used and their significances are also described.

***Keywords*** **-** *Machine learning, Image processing, Classifier, performance, Precision, recall, accuracy and F1 score*

## I. INTRODUCTION

As 80 to 85% [1] of Ethiopia's society lives in the rural areas (are agrarian), a care should be taken to assure the quality and quantity of the agricultural products. It is common to see different agricultural plants' (plants used as a source of food) diseases in different parts of the country and in different seasons. Hence, this paper discusses a survey on different plant diseases prediction or detection techniques to avoid (at least to minimize) the severity of the disease.

## II. PREDICTION OF AGRICULTURAL PLANTS' DISEASES USING MACHINE LEARNING TECHNIQUES

One of the research studies done on plant disease prediction is Ethiopian Coffee plant diseases prediction by [1]. To conduct their research, the authors collected a dataset (images) using a standard camera from the place where coffee plant is plenty to predict the three (Coffee Leaf Rust, Coffee Berry Disease and Coffee Wilt Disease) types of coffee plant leaf diseases. The researchers have collected 9100 coffee plant leaf diseased images for both training (70%) and testing (30%) using MATLAB2013Ra. Using these images, the researchers conducted an image preprocessing task, such as removing low frequency, background noise, normalize the intensity of the individual particles on the images, removing reflection and masking portion of the images. As a next step, image segmentation was conducted using K-means technique and genetic algorithm was used to select the features (level co-occurrence matrix and color) to classify the types of Ethiopian coffee plant diseases. The researchers also used sobel edge detection method to find the borders of the acquired coffee leaf images. During the experimental stage, the researchers trained the model using 70% of the dataset and tested the performance of recognition of the classifiers (Artificial Neural Network, K-Nearest Neighbor, Naive Bayes and combination of Radial basis function and self organizing map) i.e. the experiments were conducted using texture and color features separately to classify the diseased into three classes (the three diseases types). Finally, the researchers found that color features have more representing power than texture features and the combination of RBF and SOM has better classification performance as can be depicted below.

**Table 1. Classifiers' performance**

| S/No. | Classifier | Performance (%) |
|---|---|---|
| 1 | Nearest Neighbor classification (KNN) | 58.16 |
| 2 | Artificial Neural Network (ANN) | 79.04 |
| | | |
| 3<br>4 | Naive Bayes<br>Hybrid of RBF and SOM (Radial basis function and Self 90.07 organizing map). | 53.47 |

As stated in [2], the authors upgraded (improved) previous work by achieving 20% improvement. To do so, different image processing and machine learning techniques were used. Basically, the research was done to investigate agricultural plant leaf diseases, namely, early scorch, cotton mold, Ashen mold, late scorch and tiny whiteness from a collected leaf images using digital camera. Due to the reason that visually identifying plant leaf diseases is tiresome, expensive, time consuming, difficult and inefficient, machine learning was found to be the mere solution with the help of image processing tasks (image acquisition, preprocessing, segmentation and feature extraction) for early stages detection and treatment of the diseases.
Focusing on speed and accuracy to detect the above mentioned agricultural plant leaf diseases, the authors used K-means clustering to identify the infected objects, extracted the features set of the infected objects using color co-occurrence methodology for texture analysis and detected and classified the type of the diseases using Artificial Neural Networks by putting the leaves in infected or non-infected classes.

Finally, the researchers have evaluated their work using MATLAB and found an accuracy of 94.67% as can be seen in the table below.

**Table 2. Classification results per class for neural network**

| From_species | Early Scorch | Cottony Mold | Ashen Mold | scorchLate | Tiny Whiteness | Normal | Accuracy |
|---|---|---|---|---|---|---|---|
| Early Scorch | 25 | 0 | 0 | 0 | 0 | 1 | 100 |
| Cottony Mold | 0 | 24 | 0 | 1 | 0 | 0 | 96 |
| Ashen Mold | 0 | 0 | 25 | 0 | 0 | 1 | 100 |
| Late Scorch | 0 | 0 | 0 | 22 | 1 | 0 | 88 |
| Tiny whiteness | 0 | 1 | 0 | 0 | 23 | 0 | 92 |
| Normal | 0 | 0 | 0 | 2 | 1 | 23 | 92 |
| Average | | | | | | | **94.67** |

Researchers at [3] detected agricultural plant leaf diseases by proposing image acquisition, preprocessing, segmentation, feature extraction and classification as a methodology to minimize effects of diseases that harm the quality and quantity of agricultural products. The researchers acquire images of plant leaves using standard camera, found the infected portion of the image (leaf) using K-means clustering approach, extracted texture based feature using gray level co-occurrence matrix (GLCM) for model training and image classification and classified the leaves as normal or diseased using support vector machine (SVM). The researchers have used MATLAB 2013 for conducting the experiment of the proposed system while applying K-means clustering; they have got three clusters of which one contains the diseases affected leaves. After this, features are extracted using GLCM and fed into support vector machine to classify the leaves as normal and diseases affected ones. In [4], the researchers have conducted a research to detect agricultural plant leaf diseases using image processing and deep learning techniques on smart phones. They conducted their research on 54,306 plant leaves to identify 14 crop species with 26 diseases. The dataset taken to conduct the research had assigned into 38 crop-disease pair class labels and deep convolutional neural networks model was used to classify the diseases into the class labels. Although different performance measures were used to investigate the ability to predict, the best performing model achieved a mean F1 score of 99.35%. To evaluate the applicability of deep convolutional neural network approach, the researchers focused on AlexNet and GoogleNet architectures. As AlexNet has 5 conventional, 3 fully connected and a softmax layers, GoogleNet architecture has 22 layers. The researchers have used these convetional neural network architectures in different configurations by training the model from the scratch in one case, and then by adapting already trained models using transfer learning. Finally, GoogleNet is found to be better performing than AlexNet architecture. In case of the dataset used (color, grayscale and segmented), the models performed best in the case of the colored version of the dataset. As stated in [5], a deep learning approach was proposed to classify banana plant leaf diseases to early detect and diagnose two types of banana leaf (sigatoka and speckle) diseases for the protection of banana yield losses. Both of these diseases are fungal diseases. To early detect and predict ` these diseases, the researchers have used computer vision and machine learning techniques by using convolutional neural network (CNN) with LeNet architecture because CNN requires small image preprocessing tasks and can be taught visual features directly from images. So, in conducting the research work, the researchers have introduced a framework that consists of image preprocessing and deep learning based classification. In the image preprocessing part, image acquisition using standard camera (images of healthy and infected leaves are captured), color feature extraction and image resizing (60X60 pixels) were done. In the second part (classification), CNN with LeNet architecture was used due to its multilayer feature. The feature extraction model has convolution and pooling layers for which the convolution layer extracts the color features and the pooling layer resizes (reduces size) the size of the image. The classification model, as part of CNN, uses a fully connected layer that classified the images into three class labels as healthy, sigatoka or speckle infected using the softmax activation function. Finally, the researchers have conducted an experiment using 3700 images of healthy, sigatoka and speckle infected banana plant leaves using deeplearning4j library in different percentage values (80%, 60%, 50%, 40% to 20%) for training and classification. Lastly, green color banana leaves were found to be healthy and the black ones diseased. The effectiveness of the model was measured using a combination of accuracy, precision, recall and F1-score and the results are shown in the table below.

**Table 3: Precision, recall, accuracy and F1 score for the corresponding experimental configuration.**

| | | Color | | | | Grayscale | | | |
|---|---|---|---|---|---|---|---|---|---|
| Train | Test | Accuracy | Precision | Recall | F1-score | Accuracy | Precision | Recall | F1-score |
| 20% | 80% | 0.9861 | 0.9867 | 0.986 | 0.9864 | 0.9444 | 0.9479 | 0.9444 | 0.9462 |
| 40% | 60% | 0.9861 | 0.9865 | 0.9859 | 0.9863 | 0.9757 | 0.9764 | 0.975 | 0.976 |
| 50% | 50% | 0.9972 | 0.9970 | 0.9972 | 0.9971 | 0.8528 | 0.889 | 0.8527 | 0.8705 |
| 60% | 40% | 0.9676 | 0.969 | 0.9677 | 0.9683 | 0.9282 | 0.9314 | 0.9283 | 0.9298 |
| 80% | 20% | 0.9288 | 0.9299 | 0.9288 | 0.9294 | 0.8594 | 0.8678 | 0.8594 | 0.8636 |

The paper in [6] states that the researchers have conducted a research to find out a machine learning technique to early detect 13 plant leaf diseases in order to minimize the production loss, and sometimes death of the plants, using the plants' leaves images. To do this, the researchers have employed convolutional neural network approach that can be trained by caffe deep learning framework developed by Berkley Vision and learning centre and wrote their scripts in python programming language. The dataset used for training and testing were downloaded from the Internet by the name of the plants or diseases in different languages. As a result, 30,880 images for training and 2589 images for testing were used to classify the images into 15 classes as stated below.

**Table 4: Dataset for image classification of leaf disease.**

| Class | Number of original Images | Total number of images: original and augmented | Number of images from the dataset used for Validation |
|---|---|---|---|
| (1) Healthy leaf | 565 | 4523 | 331 |
| (2) Pear, cherry, and peach, porosity | 265 | 2124 | 152 |
| (3) Peach, powdery mildew | 108 | 1296 | 90 |
| (4) Peach, *Taphrina deformans* | 152 | 1552 | 156 |
| (5) Apple, pear, *Erwinia amylovora* | 232 | 2368 | 205 |
| (6) Apple, pear, *Venturia* | 183 | 2200 | 151 |
| (7) Apple, powdery mildew | 120 | 1440 | 118 |
| (8) Apple, Rust | 163 | 1960 | 163 |
| (9) Pair, *Gymnosporangium sabinae* | 267 | 2142 | 185 |
| (10) Pair, gray leaf spot | 122 | 1464 | 198 |
| (11) Grapevine, wilt | 287 | 2300 | 114 |
| (12) Grapevine, mites | 250 | 2000 | 230 |
| (13) Grapevine, powdery mildew | 237 | 1900 | 183 |
| (14) Grapevine, downy mildew | 297 | 2376 | 201 |
| (15) Background images | 1235 | 1235 | 112 |
| | **4483** | **30880** | **2589** |

In this step, augmentation process was used to increase the dataset and reduce over fitting. In the image preprocessing step, cropping, making square around the leaves (to highlight region of interest), removing images of smaller resolution and dimension less than 500 px, resizing images into 256X256 to reduce training time were done.
Holding the testing step using the validation dataset, its validation was measured using 10 folds cross validation technique and an effectiveness of 96.3% was found.

The authors in [7] have conducted a research to find out diseases of plant leaves using support vector machine approach using the plant leaves images as inputs. To do so, they have acquired plant leaves images, did image enhancement, segment each image to identify area of interest, extract features, classify the images using SVM, found the infected portion of every leaf and measure the accuracy of the classifier. They have finally got an accuracy of 96.77% in classifying the healthy leaves and 98.38% in classifying the leaves as cercospora leaf spot diseases. The researchers have lastly concluded that their experimental results are by far better than the previously obtained research results. A research on pomegranate leaf fungal diseases is done by [8]. In this work, the researchers have acquired images, preprocess the images, segment the images, extract image features, classify the leaves as healthy or diseased ones and finally recommend treatment measures. In this research, the authors have conducted their research to identify (classify) the leaves into class labels as "healthy" or "diseased" of which the diseased class contains another two subclasses of diseases (alternaria and cercospora). The researchers have taken images from National Research Center on pomegranate , India and resize the images into 512X512 without affecting the quality of the images. After resizing, the authors have applied image filtering to remove noise from the images using Gaussian Low Pass filter with a size of 3X3 and standard deviation of 0.5. In order to get the region of interest of the leaves, the researchers have applied image segmentation using K-means clustering and thresholding based approaches. Many features that will be used for classification were extracted using First-level Haar wavelet transform and the classification step into the class labels was done using Fuzzy Inference system (FIS). Lastly, the classifier recommends (informs) the farmers (agronomists) on what fungicides to be sprayed, quantity of the spray and spray intervals as can be shown in the figure below.


Fig. 1. Treatment Measures

As it is described in [9], the authors have studied to find a way to detect diseased parts of leaves of plants using image processing techniques. They conducted their research to detect diseases of leaves of banana, beans, Guava, Jackfruit, lemon, mango, potato, sapota and tomato and classify the images into healthy or any one of the diseases' classes. To do this, 500 plant leaves of 30 different native plant species were used using digital camera. The obtained images were converted into HSI format. Segmentation step was used to extract the infected portion of each leaf. Since image features are important for both training and classification, texture features were extracted using color co-occurrence method. Because this research was conducted to early detect early scorch, yellow spots, brown spots, late scorch, bacterial and fungal diseases, support vector machine (with accuracy of 94.74%) and minimum distance criterion (with accuracy of 86.77%) were used as classifying algorithms though support vector machine was found to be better.

As in other plants, a research on tomato leaves is conducted to detect its diseases [10]. To do this, tomato leaves images were taken as inputs and the inputs were changed into HIS color space representation. In this step, due emphasis was given to the H value and the green colored pixels are masked and removed to reduce the computation time because green color indicates healthy pixels. Having the infected portions of the leaves from the above step, segmentation was done using the K-medoid clustering. Color and texture feature were extracted using color co-occurrence method to train the model and finally the diseased leaves were classified using least square support vector machine taking the features as inputs to avoid the necked eye observation of plant leaf diseases problems by a botanist.

The study in [11] presents methods to detect leaves' diseases of bean and bitter gourd. In this research, the researchers have used 118(63-bean leaves and 55-bitter gourd) diseased leaves taken from the field using digital camera. To conduct this study, shape and texture features of diseased 118 leaves were used. To classify the diseased leaves, feed forward neural network algorithm (FFNN), learning vector quantization (LVQ) and radial basis function network (RBF) were used. As it is natural to measure the performance of any model, the researchers have employed accuracy, precision, recall and F-measure to evaluate its performance. Finally, the researchers have experimented that 58 bean leaves were correctly classified and 5 leaves were wrongly classified using feed forward neural network algorithm; 60 bean leaves were correctly classified and 3 been leaves were wrongly classified using learning vector quantization and all bean leaves were correctly classified using radial basis function algorithm. On the other hand, 49 bitter gourd leaves were correctly classified and 6 leaves were incorrectly classified using feed forward neural network algorithm; only 7 bitter gourd leaves were correctly classified and 48 were wrongly classified using learning vector quantization and 21 bitter gourd samples were correctly classified and 34 bitter gourd leaves were wrongly classified using radial basis function algorithm. Hence, feed forward neural network classification approach provided better result.

## III. CONCLUSION

This survey investigated methods to predict agricultural plant leaves diseases to reduce the effects of the diseases on agricultural production using machine learning techniques. Though the proposed techniques can help farmers at least in minimizing the effects of the diseases, none of these techniques describe the severity of each disease, are only used to predict plant leaves diseases and are applied only to specific types of plant leaf diseases. Table 5 summarizes the features of machine learning techniques based plant leaves disease prediction techniques.

**Table 5. Machine Learning Technology**

| No. | Type | Target to Predict | Proposed Techniques | Outputs |
|---|---|---|---|---|
| 1. | Coffee | Coffee Leaf Rust, Coffee Berry Disease and Coffee Wilt Disease | Artificial Neural Network, K-Nearest Neighbor, Naive Bayes and combination of Radial basis function and self organizing map | Coffee Leaf Rust, Coffee Berry Disease and Coffee Wilt Disease |
| 2. | Leaf | early scorch, cotton mold, Ashen mold, late scorch and tiny whiteness | Artificial Neural Network | Infected(any one of the diseases) or normal |
| 3. | Leaf | normal and diseases affected | support vector machine | normal or diseased |
| 4. | Leaf | 14 crop species with | Deep convolutional | 38 crop-disease |

| | | | | |
|---|---|---|---|---|
| | | 26 diseases | neural network | pair class labels |
| 5. | Banana | sigatoka and speckle | Convolutional neural Network | healthy, sigatoka or speckle infected |
| 6. | Leaf | 13 plant leaf diseases as in table 4 | Convolutional neural network | Classified into 15 classes as in table 4. |
| 7. | Leaf | Cercospora | Support vector machine | Cercospora |
| 8. | Pomegranate | alternaria and cercospora | Fuzzy Inference system | healthy or diseased |
| 9. | Banana, Beans, Guava, Jackfruit, Lemon, Mango, Potato, Sapota and Tomato | early scorch, yellow spots, brown spots, late scorch, bacterial and fungal diseases | support vector machine and minimum distance criterion | healthy or any one of the diseases' classes |
| 10 | Tomato | Leaf disease | least square support vector machine | diseased leaves |
| 11 | Bean and Bitter gourd | Leaves disease | Feed forward neural network algorithm, learning vector quantization and radial basis function network. | diseased leaves |

## IV. REFERENCES


[1]. Abrham Debasu Mengistu1*, Dagnachew Melesew Alemayehu2 and Seffi Gebeyehu Mengistu3, "Ethiopian Coffee Plant Diseases Recognition Based on Imaging and Machine Learning Techniques", International Journal of Database Theory and Application Vol.9, No.4, pp.79-88, 2016.

[2]. H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of

[3]. Bindushree H B, Dr. Sivasankari G G, "Detection Of Plant Leaf Disease Using Image Processing Techniques", International Journal of Technology Enhancements and Emerging Engineering Research, vol 3, issue 04, issn 2347-4289.

[4]. Sharada P.Mohanty1, 2,3, DavidP.Hughes4, 5,6 and MarcelSalathé 1, 2,3*, "Using Deep Learning for Image-Based Plant Disease Detection", Frontiers in Plant Science | www.frontiersin.org, September 2016 | Volume 7 | Article 1419.

[5]. Jihen Amara,1 Bassem Bouaziz,2 and Alsayed Algergawy 3 , “A Deep Learning-based Approach for Banana Leaf Diseases Classification”, B. Mitschang et al. (Hrsg.): BTW 2017 – Workshopband, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 79
[6]. Srdjan Sladojevic,1 Marko Arsenovic,1 Andras Anderla,1 Dubravko Culibrk,2 and Darko Stefanovic1 , “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification”, Hindawi Publishing Corporation Computational Intelligence and Neuroscience Volume 2016, Article ID 3289801, 11 pages http://dx.doi.org/10.1155/2016/3289801
[7]. Sujeet Varshney and Tarun Dalal, “A Novel Approach for the Detection of Plant Diseases”, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 5, Issue. 7, July 2016, pg.44 – 54.
[8]. S. S. Sannakki, V. S. Rajpurohit, “An Approach for Detection and Classification of Leaf Spot Diseases Affecting Pomegranate Crop”, International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 2, Special Issue (NCRTIT 2015), January 2015. ISSN 2348 – 4853.
[9]. S. Arivazhagan, R. Newlin Shebiah*, S. Ananthi, S. Vishnu Varthini, “Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features”, Agric Eng Int: CIGR Journal, 15(1): 211－217, 2013.