Plant Diseases Recognition Techniques

Hailay Beyene (PhD scholar)[1], Narayan A. Joshi[2], Ketan Kotecha[3]
[1]Parul University, [2]Dharmsinh Desai University, [3]Symbiosis Institute of Technology

Abstract- This review of article discusses about the mechanisms to early detect plant cereals' leaves diseases to ensure the quantitative and qualitative safety of the products using image processing and machine learning techniques. Hence, in this review, a number of machine learning and image processing techniques, such as K-nearest neighbor, Nave bayes, Radial basis function, Convolutional neural network, Self organizing map, Fuzzy inference system, Artificial Neural Network, Support vector machine, minimum distance criterion, least square support vector machine, Feed forward neural network, learning vector quantization, deep convolutional neural network approach and GLCM, First-level Haar wavelet and thresholding were used respectively.

Keywords- Training**,** image processing, F1 score, classifier, performance, Precision, recall, accuracy, machine learning, image segmentation and image extraction.

I. INTRODUCTION

The sole aim of this review is to identify techniques that can be used for protecting-at least minimizing, the severity of cereals' diseases to ensure the food security of the agrarian (people leading their life by agriculture) because 80 to 85% [5] of Ethiopia's society lives in the rural areas (are agrarian). The quality and quantity of the agricultural products must be maintained by taking care of the agricultural plants early from different diseases. 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. However, there are no adequate ways to early detect the diseases and ensure the quality and quantity of the products. Hence, the objective of this review is to identify the existing techniques to early investigate agricultural plant diseases so that the techniques can be adopted or can pave a way to another techniques to early protect the cereals, such 'cactus'.

1. Agricultural Plants' Diseases Prediction Techniques

A research study was done by [5] to investigate Ethiopian Coffee plant diseases. It was conducted in such a way that, the authors collected a dataset (images) using a standard camera from the place where coffee plant is plenty to predict Coffee Leaf Rust, Coffee Berry Disease and Coffee Wilt Disease of coffee plant leaf diseases. For this purpose, the researchers have collected 9100 coffee plant leaf diseased images for both training and testing. They used 70% of the images for training and 30% of them for testing using MATLAB2013Ra. To investigate the diseases, 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 were conducted.

Since image preprocessing is not enough to get good result during the classification stage, the researchers have used K-means technique for image segmentation and genetic algorithm was used to select the features (level co-occurrence matrix and color) to classify the types of Ethiopian coffee plant diseases. They also used Sobel edge detector to find the edges of the infected portions of the acquired coffee leaf images.

As it is stated above, the authors have used 70% of the image (dataset) for training their model during the experimentation step. Artificial Neural Network, K-Nearest Neighbor, Naive Bayes and combination of Radial basis function and self organizing map were used as classifiers and color and texture features were used as features of classification to classify the three types of coffee plant diseases in to three class labels.

Lastly, the authors have found that color features have good representing power than texture features and the combination of SOM and RBF has better classification performance (result) as can be seen below.

Table 1. Classifiers' performance

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

As stated in [1], the previous work was improved by 20% using different image processing and machine learning techniques. This study was done to investigate scorch, cotton mold, ashen mold, late scorch and tiny whiteness from a collected leaf images using digital camera. As visually identifying diseases is expensive, tiresome, time consuming, difficult and inefficient, machine learning was found to be the mere solution with the help of image acquisition, preprocessing, segmentation and feature extraction for early stages of diseases detection and treatment of the diseases.

To detect these diseases accurately and faster, 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. Artificial Neural Network classifier was used to detect and classify the type of the diseases by putting the leaves in infected or non-infected class labels.

The implementation was done using Matlab and the researchers have got an accuracy of 94.67% as can be seen in the table below.

Table 2. Classification results per class for neural network

From Species	Late scorch	Tiny Whiteness	Normal	Early Scorch	Cottony Mold	Ashen Mold	Accuracy (%)
Late Scorch	22	1	0	0	0	0	88
Tiny whiteness	0	23	0	0	1	0	92
Normal	2	1	23	0	0	0	92
Early Scorch	25	0	1	0	0	0	100
Cottony Mold	1	0	0	0	24	0	96
Ashen Mold	0	0	1	0	0	25	100
Average							**94.67**

[12] also conducted a research work to detect plant leaf viral, fungus and bacterial diseases that can directly or indirectly affect the human life. In this work, many plant diseases prediction techniques, such as support vector machine, K-Nearest Neighbor, Artificial Neural Network and Probabilistic Neural Network were discussed and best technique was proposed. The researchers have proposed new plant diseases detection algorithm to identify the infected part of the leaf after the training using a training data set (image) is taken place. To do this, the researchers intensively used support vector machine algorithm to accurately analyze (such as contrast enhancement) the leaf image. So, the system is trained by a training data set (image) and affected parts of the leaf are easily detected and found the accuracy of greater than 90% using four leaves images.

A research conduct by authors at [2] states that the authors have detected agricultural plant leaf diseases. To do so, they have proposed image acquisition, preprocessing, segmentation, feature extraction and classification as a methodology to minimize effects of diseases that can affect the quality and quantity of the plants' products. The images were acquired using a standard digital camera from the field, K-means clustering approach was used to identify the infected portion of the image (leaf image), a texture based feature was extracted using gray level co-occurrence matrix (GLCM) for model training and image classification and Support vector machine (SVM) was used to classify the leaves images as normal or diseased using MATLAB 2013. I.e. MATLAB 2013 was used for conducting the experiment of the proposed system for image preprocessing, feature extraction and diseases classification in to normal and diseased classes.

A research in [3] was conducted to identify 26 plant leaves' diseases using 54,306 leaves on 14 crop species using deep learning and image processing techniques on smart phones. 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.

The research study done by [13] shows that weeds are the severe headaches to drastically reduce the productivity of agricultural products. As is stated by the researchers, applying excessive herbicides uniformly throughout the agricultural field may affect the environment at large. As a result, a system (machine) that can differentiate crops from weeds using digital images that can minimize (solve) the problems is required. To do this, the researchers used three (Naive Bayes, support vector machine and C4.5 (decision tree algorithm)) classification algorithms to detect weeds by employing digital image analysis. They used too many images as training and testing data sets using the machine learning algorithms. Finally, they were able to detect the weeds and Naive Bayes algorithm was found to be the most accurate (99.3%) to correctly classify the weeds.

The authors at [4] have proposed a deep learning approach to early investigate healthy or infected banana leaves using computer vision and machine learning techniques. The study states that there are two common banana leaf diseases, namely, sigatoka and speckle that decreases the banana yields. It is also stated that these diseases are fungal diseases. In this study convolutional neural network (CNN) with LeNet architecture is used because CNN requires small image preprocessing tasks and can be taught visual features directly from images. In conducting the research work, a framework that consists of image preprocessing and deep learning based classifications were introduced. During image processing, 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 classification step, 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.

During the experimentation step, 3700 images of healthy, sigatoka and speckle infected banana plant leaves were used and the experimentation was done using deeplearning4j library in different percentage values (80%, 60%, 50%, 40% to 20%) for training and classification. Finally, 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	Recall	Precision	Accuracy	F1-score	Recall	Precision	Accuracy	F1-score
20%	80%	0.986	0.9867	0.9861	0.9864	0.9444	0.9479	0.9444	0.9462
40%	60%	0.9859	0.9865	0.9861	0.9863	0.975	0.9764	0.9757	0.976
50%	50%	0.9972	0.9970	0.9972	0.9971	0.8527	0.889	0.8528	0.8705
60%	40%	0.9677	0.969	0.9676	0.9683	0.9283	0.9314	0.9282	0.9298
80%	20%	0.9288	0.9299	0.9288	0.9294	0.8594	0.8678	0.8594	0.8636

Another research work done [14] on agricultural plant diseases detection indicates that plant diseases decrease productivity, quality, and growth of the plants. To early overcome these problems, there must be techniques in which the diseases can be detected by extracting some features, such as, color, shape and edge from the plants' leaves or stems. So, these researchers were able to detect plant viral, fungus and bacterial diseases using different machine learning algorithms and image processing techniques though the accuracy of performance they found is not explicitly explained.

The authors in [6] conducted a research to early detect 13 plant leaf diseases as can be seen from table.4 using computer vision and machine learning techniques in order to minimize the early loss of the production and sometimes death of the plants. To do this, convolutional neural network approach that can be trained by caffe deep learning framework developed by Berkley Vision and learning centre was employed and the scripts were written in python programming language. 30,880 images for training and 2589 images for testing were downloaded and used as a dataset from the Internet by the name of the plants or diseases in different languages. Finally, the images were classified into 15 classes as depicted 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) Background images	1235	1235	112
(2) Grapevine, downy mildew	297	2376	201
(3) Grapevine, powdery mildew	237	1900	183
(4) Grapevine, mites	250	2000	230
(5) Grapevine, wilt	287	2300	114
(6) Pair, gray leaf spot	122	1464	198
(7) Pair, *Gymnosporangium sabinae*	267	2142	185
(8) Apple, Rust	163	1960	163
(9) Apple, powdery mildew	120	1440	118
(10) Apple, pear, *Venturia*	183	2200	151
(11) Apple, pear, *Erwinia amylovora*	232	2368	205
(12) Peach, *Taphrina deformans*	152	1552	156
(13) Peach, powdery mildew	108	1296	90
(14) Pear, cherry, and peach, porosity	265	2124	152
(15) Healthy leaf	565	4523	331
	4483	**30880**	**2589**

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 research studied by authors in [7] was improvement of previous study. This research was done to investigate cereal plant leaf diseases using support vector machine (SVM) approach by acquiring the plant leaves' images, enhancing the plant leaves' images, segmenting each image to identify the area of interest, extracting the necessary features of the plant leaves' images. Finally, the images were classified using SVM, found the infected portion of every leaf and measured the accuracy of the classifier. The authors have finally got an accuracy of 98.38% in classifying the leaves as cercospora leaf spot diseases and 96.77% in classifying the healthy leaves. These results have lead the researchers to decide that their experimental results are by far better than the previously obtained research results.

Researchers at [8] have conducted a research study on pomegranate leaf fungal diseases to classify the leaf images into three class labels, namely, 'healthy', 'alternaria' or 'cercospora' because if the pomegranate leaf is infected, it is either alternaria or cercospora. In this work, the researchers have acquired images, preprocess the images, segment the images, extract image features, classify the leaves as healthy or diseased (alternaria or cercospora) ones and finally recommend treatment measures. The images were taken from National Research Center on pomegranate, India and resized 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 thresholding and K-means clustering 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.

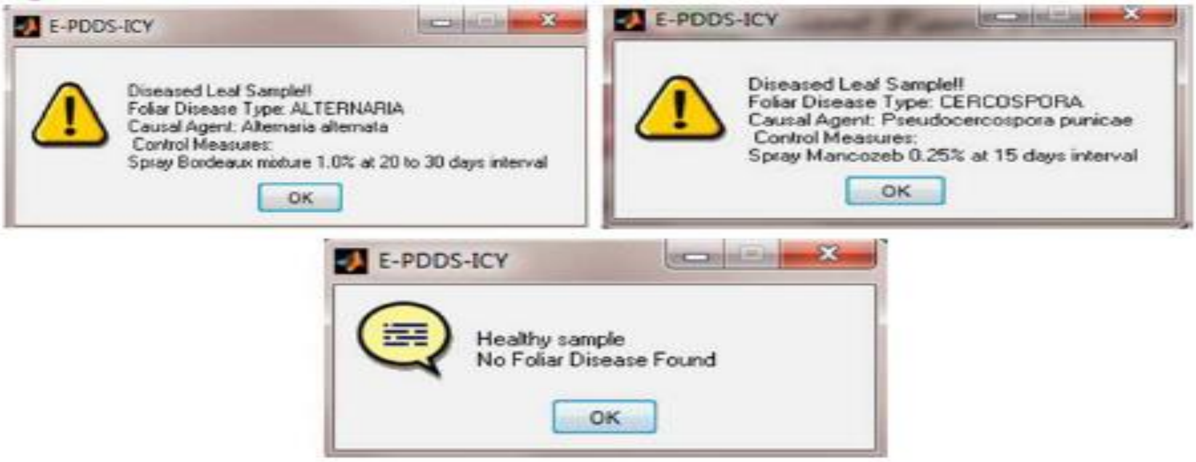

Fig.1: Treatment Measures

As it is described in [9], researchers have conducted a study to early detect early scorch, yellow spots, brown spots, late scorch, bacterial and fungal diseases that affect of leaves of banana, beans, Guava, Jackfruit, lemon, mango, potato, sapota and tomato. To conduct the study, the researchers have used image processing and machine learning techniques. Images of 500 plant leaves of 30 different native plant species were taken using digital camera from the field. These images were converted into HIS (hue saturation intensity) format and infected portions of the images (leaves) were extracted using segmentation technique. Because image features are important for both training and classification, texture features were extracted using color co-occurrence method. 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 have better performance (accuracy).

As this research study focuses on cereal plant diseases detection, a research on tomato leaves is conducted to detect its diseases [10]. Tomato leaves images were used as inputs and changed into HSI (hue saturation intensity) 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. After this, image segmentation was done using K-medoid clustering. Color co-occurrence method was used to extract color and texture features 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 authors in [11] have held a research to detect leaves' diseases of bean and bitter gourd. To conduct this research, 118(63-bean leaves and 55-bitter gourd) diseased leaves were taken from the field using digital camera. Shape and texture features of diseased 118 leaves were used to conduct this study. Feed forward neural network algorithm (FFNN), learning vector quantization (LVQ) and radial basis function network (RBF) were used to classify the diseased leaves.

The researchers have employed accuracy, precision, recall and F-measure to evaluate the performance of their model. Lastly, the authors 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.

Besides this, the researchers have conducted an investigation and 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. Finally, feed forward neural network classifier provided better result.

II. CONCLUDING REMARKS

In this review, different image processing and machine learning techniques that can be used to early recognize agricultural plants' leaves to reduce the impacts of the diseases on agricultural products are identified. It is found that these proposed techniques are important to help farmers at least in minimizing the effects of the diseases though most of these techniques do not describe the severity of each disease. The techniques that are identified in this review are only used to predict plant leaves diseases and are applied only to specific types of plant leaf diseases.

III. REFERENCES

[1]. 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 Computer Applications (0975 – 8887), Volume 17– No.1, March 2011.

[2]. 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.

[3]. Sharada P.Mohanty[1, 2,3], DavidP.Hughes[4, 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.

[4]. 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

[5]. Abrham Debasu Mengistu[1*], Dagnachew Melesew Alemayehu[2] and Seffi Gebeyehu Mengistu[3], "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.

[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.

[10]. Vidyaraj K1, Priya S2, " Developing an algorithm for Tomato leaf disease detection and classification", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 3, Special Issue 1, February 2016.

[11]. K. Muthukannan1, P. Latha2, R. Pon Selvi1 and P. Nisha1 "CLASSIFICATION OF DISEASED PLANT LEAVES USING NEURAL NETWORK ALGORITHMS", VOL. 10, NO. 4, MARCH 2015.

[12]. R. Zhou, S. Kaneko, F. Tanaja, M. Kayamori, M. Shimizu: "Matching Based Cercospora Leaf Spot in sugar beet" The 11th International Conference on Quality Control by Artificial Vision, pp.99-106. 2013.

[13]. Md . Mursalin and Md . Motaher Hossain et. Al, "Performance Analysis among Different Classifier Including Naive Bayes, Support Vector Machine and C4.5 for Automatic Weeds Classification", Global Journal of Computer Science and Technology Graphics & Vision, Volume 13 Issue 3 Version 1.0 Year 2013.

[14]. Santanu Phadikar and Jaya Sil, "Rice Disease Identification using Pattern Recognition Techniques", Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008), 25-27 December, 2008, Khulna, Bangladesh, pg.no 420- 423, (IEEE).

Acknowledgments **Author's** Profile

I gratefully thank Dr. Narayan A. Joshi and Dr. Ketan Kotecha for their continuous and constructive advice. They dedicated their precious time and energy for this work to happen. I also express my special thanks to Ethiopian Ministry of Education for its financial support to conduct my study without any financial burden. Last but not least, special thanks goes to my family for giving me strength and patience to concentrate on my study.

Hailay Beyene.

- BSC degree in Computer Science from Mekelle University, Ethiopia.
- MSC degree in Computer Science from Addis Ababa University, Ethiopia.
- Research area: *Information retrieval, Machine learning, image processing*