Implementation of Improved Machine Learning Techniques for Plant Disease Detection and Classification

Ibrahim M. Adekunle

Department of ICT, Osun State University, Nigeria

***Abstract*−Agricultural production plays a big role in economic growth especially in developing countries; one of the biggest problems for a farmer is different type of diseases that affect crop on farm. This problem has destroyed plenty of crops which eventually led to shortage in annual agricultural production across the world. Quality and high production of crops could be determined by early detection of diseases in the crop. Diseases detection through manual observation can be somehow tedious, complex, expensive, and difficult, and subject to rigorous analysis. Although some researchers have worked in this field but most of the existing methods for solving problems in diseases detection have not been effective in terms of real time basis. This paper therefore presents an effective method for efficient detection of maize leaf diseases. The proposed method uses image processing techniques for the extraction of important features in order to showcase the characteristics properties of the image that could be used for the identification. Machine learning techniques are applied to classify the extracted features to separate diseased plant leaf from healthy ones. The experimental results show that the application of modified machine learning techniques could be effectively used for the classification of plant leaf diseases even with an accuracy of 96.7%. This approach would be very useful to farmers to prevent damages of crops, shortage of food production in the society and wasting of money on agricultural products like pesticides and so on.**

***Keywords*−Maize leaf disease, feature extraction, binary images, histogram equalization and machine learning**

I. INTRODUCTION

Pests and diseases results in damages of crops or part of the plant that can result to decrease in food production in the society. The knowledge of pest management and diseases are still lacking in different developing countries. Poor disease control, climate changes are one of the key factors in food production. When we talk of plant disease, this could also involve important plants such as crops, tomatoes, cassava, vegetables and other valuable plant that could lead to food production. It is therefore very important to monitor and take proper care of our agricultural products as this would increase our harvest in farming at the end of each year. Disease is one of the factors that can affect the amount of food production annually; how do we control these effects of pest and diseases on our crops? Disease identification in plants has been identified to be a very important technique for preventing the loss in the production of agricultural products. Of course it could be very difficult to monitor plant diseases manually at all time as this would require more energy and time in the process.

Early detection and classification of plant diseases can help in no small way to manage the agricultural products of crops in different farms. Many modern approaches have been implemented to minimize the effects of diseases in plants and also maximize the agricultural productivity but unfortunately extreme success has not been achieved in this area as expected. In spite of various technological approaches, expected production of local foods has not been met in various countries across the world today. Of course, governments are trying there best to ensure there are enough foods for their people at all time, some developing countries are using their natural resources to achieve this by importing foods from neighboring countries in order to meet people 's demand on foods and other agricultural products. This is the reason why researchers need to help in this area by developing efficient techniques to achieve success in minimizing some of the damages caused by pests and diseases to our agricultural products every year. Chromatography, mass spectrometry, thermography and hyper spectral techniques have been employed for disease identification. However, these techniques are not cost effective and are high time consuming. In recent times, server based and mobile based approaches for disease identification have been employed for disease identification. Several factors of these technologies being high resolution camera, high performance processing and extensive built in accessories are the added advantages resulting in automatic disease recognition.

Diseases in plant can cause economic and production losses in agriculture and forestry. For example, in [1], economic loss caused by a fungal disease in soybeans has led to a profit loss of about 11 million dollar due to plant infection. Early detection of infection in plants can play a very big role in the field of agriculture and the use of automatic disease identification system can be extremely important in this case. The viral, fungal and bacterial infections have been causing a lot of damages in agricultural products for the past few years. The paper in [1] extracts some useful features from plant leaf images using Local Tri-directional Patterns (LTriDP) for classifying different classes. This method uses discriminant properties represented by each image class with reduced dimensions for detection and classification of tomato plant diseases. The authors achieved great success as the experimental results showed that the proposed framework outperformed other previous approaches with an accuracy of 94%.

Detection of plant diseases using the physical appearance could sometimes pose a very big challenge compared to automatic detection that uses image processing methods since this approach takes less time and provides more accurate results. Generally speaking, some of the symptoms that can be seen in plant with serious diseases are yellow and brown with some scorches at early and late period. Image processing techniques could be used for the classification of the disease type for describing the affected area by extracting the characteristics properties of the plants [2-4]. This paper therefore presents a framework for computer-based system that consists of combination of image processing technique and machine learning approaches for early detection of diseases of maize crop. We propose to apply modified histogram equalization for feature extraction from leaf image after preprocessing. The rest of this paper is structured as follows. In Section II, brief review of existing research work is discussed. Section III presents the proposed framework for maize crop disease identification. Section IV reports the experimental results and in Section V, concluding remarks of this work are presented.

II. RELATED WORK

In [5], the authors applied image processing techniques and machine learning algorithms for solving constant problem in the agricultural sector that affects the production of blueberries. The results of the model showed that Deep Learning model could be effectively used to classify whether the blueberry plant has been affected or not.[6]applied Random Forest classifier in detecting healthy and diseased leaf from the dataset created. The authors used the machine learning techniques to train large datasets for the classification and analysis of diseases in plant and recognize abnormalities that occur on plants in their greenhouses or natural environment. The authors in [6] also introduced a Cotton Leaf disease identification using pattern recognition technique, which uses snake segmentation; Hu's moments are used to distinctively describe the characteristics properties of images. Active contour model has been used to limit the vitality inside the infection spot; BPNN classifier tackles the numerous class problems. The average classification result of the cotton leaf disease is about 85.52%.

Leaf disease detection and grading using computer vision technology & fuzzy logic. K-means clustering has been applied to segment the defected area; they applied artificial neural network (ANN) as a classifier which mainly helps to check the severity of the diseased leaf.[7] reviewed all the existing approaches in Deep Learning algorithms by analyzing the achievement of researchers so far in the detection and classification of symptoms of plant diseases. This review provides a comprehensive explanation and analysis of Deep Learning Techniques in visualizing various plant diseases. The authors identified some research gaps that could be used to obtain greater transparency in detecting diseases in plant.

In [8], classification of leaves was performed by using author-modified convolutional Neural Network (CNN) and Random Forest (RF) classifier among 32 species in which the performance was evaluated through classification accuracy at 97.3%. On the other hand, it was not as efficient at detecting occluded objects [9]. In order to detect the diseases in banana leaf, LeNet architecture was implemented and classification accuracy, F1- scores were used for the evaluation of the model in Colorand Gray Scalemodes [10]. [11] presents an effective method to detect paddy leaf disease by applying Gray-Level Co-occurrence Matrix (GLCM) technique for the extraction of features and Artificial Neural Network for the classification accuracy. This approach minimizes the excessive use of pesticides and ensures healthy crop is produced in each cycle.

[12] uses image analysis technique to identify plant parts and extracts relevant traits for efficient identification and characterization of plant diseases [13]; typically it is their agglomeration across a study that could provide suitable input for machine learning. They applied new techniques to represent the image data in a way that machine learning algorithms can use.

[14] trained a deep convolutional neural network to identify five different diseases in tomato leaves. The developed model clearly separated the infected tomato leaves from the healthy leaves with an accuracy of 99.84% on a held-out test set. The approach of training deep learning models on publicly available image datasets presents a clear path toward crop disease diagnosis on a massive global scale[15-16].Many rich countries in the world were once populated by small farmers who practiced subsistence agriculture and suffered massive yield losses due to diseases and pests. Farmers can determine the economic growth of a country, they can determine the wealth of a country and they can determine the standard of living of a country. But over the years the effect of diseases and pests has been making the dreams of so many countries not a reality. Lots of farmers have been frustrated out of the business while some are still struggling by battling with these effects of diseases and pests on crops production**.** This paper therefore aims at detecting, classifying and analyzing the diseases that affect maize crop by applying image processing and machine learning techniques. How can we achieve this? What are the symptoms of diseases that can be used for the detection and analysis? These ideas would help to differentiate between healthy and unhealthy plants by watching the physical appearance of the plants and extract some useful features using machine learning techniques. The details of feature extractions and implementation techniques of the algorithms would be discussed extensively in the next section.

III. MATERIALS AND METHODS

The online database used for this research consists of four different classes of corn leaves from plant village website. The images were not directly collected from the plant village website but through Mendeley website at [17]. The corn diseases symptoms can be categorized based on diseases severity of the affected areas in the crop. The four major classes on the maize leave crop are rust-leaf, spot leaf, blight leaf and healthy leave, each symptom can be physically sighted for proper identification of corn leave diseases symptoms on the affected parts. The database has a total image of 3852 ROIs of size 256x256 pixel patches from four different classes: normal healthy leaf (1162 images), corn leaf blight (985 images), corn leaf rust (1192 images) and corn leaf spot (513 images) [17].

The proposed method for early detection of fungal disease from physical appearance and other symptoms of maize crops comprises of the tasks such as image acquisition, preprocessing of images, feature extraction and so on. The summary of this approach for efficient detection of maize diseases is shown in Figure 1. The system overview of crop diseases detection and analysis involve several algorithmic steps as we have mentioned earlier. Each step of the algorithmic approach would be thoroughly discussed and analyzed using the data collected from the online database with the help of image processing techniques and machine learning methods.

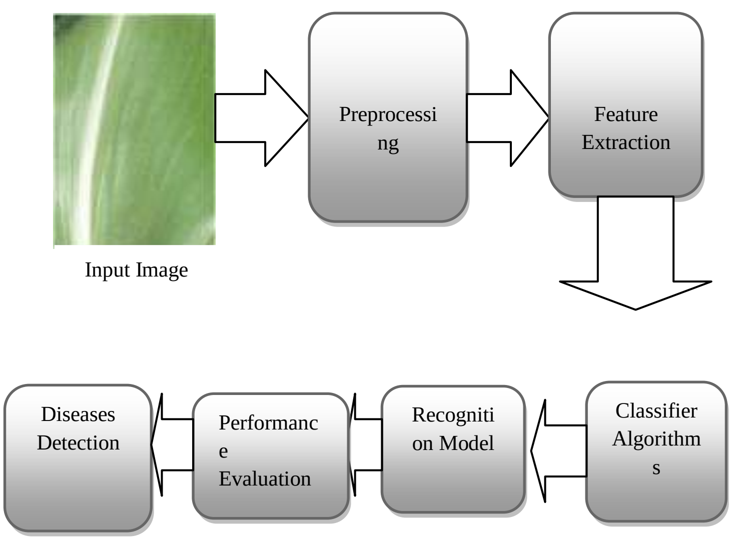

Figure 1: System models for diseases identification in Maize crop leaves

The method uses some characteristics features extracted from the input images to identify different symptoms on the crop plant. These diseases can affect different parts of the crop plant ranging from the leave, stem and other parts of the plant. The developed methodology as shown in Figure 1 consists of different stages; the first stage of the algorithmic steps is the acquisition of maize crop leaves from online database. The samples of the image leaves for different classes are shown in Figure 2 after cropping into smaller size with pixel dimension of 256 x 256. The images are stored in JPEG format and all in RGB color, the prototype uses MATLAB image processing library for this process.

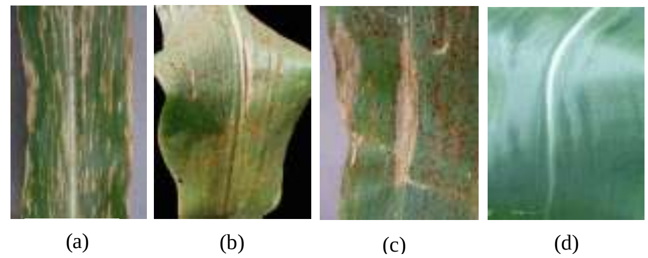

Figure 2: (a) Leaf with spot disease (b) Leaf with rust disease (c) Leaf with blight disease (d) Healthy leaf

After we have acquired the images and crop to appropriate pixel size to increase the speed of experimental computation, the next stage of the implementation process is pre-processing stage. Preprocessing of the input image is one of the important stages in the process of identifying diseases in maize crop leaf; this process has been applied to improve the quality of image and also remove the unwanted materials such as noise from the images. At this stage, the input images would be processed to obtain the equivalent binary image with little or no noise in order to improve the classification accuracy of disease detection. First and foremost, the RGB image (Figure 3a) would be converted to gray scale image as shown in Figure 3b, the resulted image has been processed further to generate a binary image with noise as presented in Figure 3c.Gaussian and median filters have been applied to get rid of the unwanted features that could introduce errors into our computational calculation, which results into image in Figure 3d.

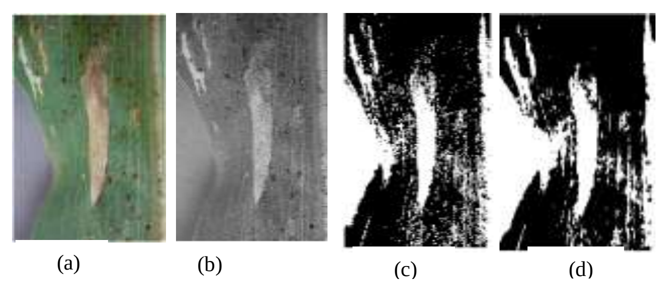

Figure 3: Illustrating Preprocessing stage of Maize leaf; (a) RGB image with blight disease; (b) Gray Scale Image; (c) Binary Image with Noise; (d) Binary Image with Noise free

As can be seen in Figure 3, several stages have been involved in reducing the noise level of the input image and improve the quality of the image enhancement for efficient detection and analysis of crop leaf. The process has really helped us to identify the affected parts of the leaf as presented in Figure 3. During this process, a 3D dimensional image has been converted to the corresponding 2D by applying some programming skills to normalize the intensity values of the image by transforming the RGB image to gray scale (Figure 3b). The image in Figure 3b exposes some of the blight diseases caused by bacteria in the affected areas, even the brown spot becomes more darken than other areas with gray color while the brown blight at the middle of the image becomes more lighter. This process converts the true color image RGB (Figure 3a) to the grayscale intensity image (Figure 3b) by eliminating the hue and saturation information while retaining the luminance. In RGB, visualization of color spaces is much harder since there are additional dimensions that the standard brain cannot visualize easily. Working around the image by thinking of each color variable as an intensity image could lead to grayscale image processing.

Additionally, Color information does not help us to identify important edges or other features that could enhance the image and result to efficient diseases detection and classification of crop leaf. Even the complexity of images and codes are another reasons why one should think of converting to grayscale since this process would drastically reduce the time and space complexity of processing and detecting diseases in plant leaf. It is a very excellent approach of solving problems since most previous and existing methods encountered this problem of high time consumption of algorithms. Starting with grayscale processing rather than color and understand how it can be applied to multichannel processing would go a long way in improving the speed, accuracy and overall efficiency of our approach.

The segmentation and pre-processing task are the initial stage before the image is used for the next process. The next stage of this process is to obtain binary images by selecting the lowest point between two classes of the histogram by considering the between-class variance. Figure 3c shows a binary image of maize leaf with some noise level as we can see, however in this Figure 3c, visualization of diseased part becomes easier compared to previous stage. With this binary image (Figure 3c), one can identify and analyze some set of pixels with different information. For instance, in Figure 3c the process separates the background level, which is zero from the diseased parts that is in the form of white level (1) since a binary image usually consists of pixels of exactly two colors (black and white). This means each pixel is stored as a single bit- 0 or 1. This process has normalized the intensity values of the grayscale images from 0-255 to 0-1. The problem with this stage is that it contains some noise or unwanted materials that can introduce errors into our experimental computation.

Median filter is implemented in this process. The average filter computes the mean (average) of the gray-scale values within a rectangular filter window surrounding each pixel. This has the effect of smoothing the image (eliminating noise). The filtered pixel has been calculated by:

$$r = (a1 + a2 + \ldots + an) / n \qquad (1)$$

The outcome of this implementation has been fantastic as presented in Figure 3d. Figure 3d now becomes more clearer compared to Figure 3c as the diseased parts indicated with white color are more prominent than the previous image.

IV. RESULTS AND DISCUSSION

In this work, histogram equalization techniques have been modified and used to extract useful features from segmenting images by improving the contrast image enhancement of the input image. In diagnosing and analyzing diseases patterns in crops, it requires good contrast images for better detection and accuracy purpose since this process contains maximum information of the diseases affecting plants. Most crops and plants images are of low contrast and this makes detection or visualization very difficult; therefore better contrast or localization of images is required. This approach would improve the image brightness and preserve the necessary information for further processing such as classification and so on.

The extracted features have been trained with random forests and support vector machine classifiers for comparative analysis. The two classifiers selected are very flexible in nature and can be used for classification and even regression analysis of crop plants. The extracted features from the processed images are divided into training and testing data, such that the feature vector is generated for the training dataset. The generated feature vector is trained with both classifiers to construct robust classification models for efficient classification of crop diseases. The feature vectors are extracted for test images to validate the performance of the classification models in order to evaluate and determine the accuracy of diseases detection in maize leaf. About 80% of the datasets prepared has been allocated for training the classification models while the remaining 20% are used for testing the algorithms. These vectors were processed with three different learning algorithms: Support Vector Machine [18-19], Random Forests [20-21] and Neural Networks[22-24]. The classification results are presented in Table 1 after the experiments.

Table 1: Classification Results for Maize Leaf Diseases

Machine Learning Models	Accuracy
Support Vector Machine	93.6%
Random Forests	96.7%
Neural Networks	95.3%

In Table 1, when SVM and characteristics features with histogram equalization were used, the algorithms recorded a classification accuracy of 93.6%, with random forest, we obtained a classification accuracy of 96.7% and lastly neural network classifier gave an accuracy of 95.3%. Overall classification accuracy of over 90% for maize leaf diseases is an excellent result. The results show the effectiveness of preprocessing approaches with the feature extraction methods in developing robust models and very powerful features for automatic detection and classification of plant leaf diseases. Random Forests performed best with an accuracy of 96.7% compared to support vector machine and neural networks. Of course support vector machine can sometimes perform better than Random Forests classifier but in this case, Random Forest seems to be better in terms of accuracy. This is probably because Random forest is well suitable for multiclass problems like in our datasets while SVM is more suitable for two-class problems and the use of binary techniques during preprocessing has really helped especially in the performance of Random forests classifier. With Neural Networks, it does not require big data to train the classification models unlike SVM that always work or train well with big data. The size of our data in conducting experiments in this research work could be one of the reasons why we have got better results with random forest compared to neural networks and SVM.

V. CONCLUSION AND FUTURE WORK

Accurate detection of symptoms of plant diseases with the help of image processing techniques and machine learning can help in supporting famers during their struggle against disease outbreaks. In this work, images of plant leaf that can visually display different symptoms of plant diseases have been used for different experiments to develop different recognition models for detecting diseases. The maize leaf used for experiments has four categories of diseases; leaf rust disease, leaf spot disease, leaf blight disease and healthy leaf. The images from the database have gone through rigorous preprocessing techniques and appropriate modified methods for extracting powerful features that could yield good results. The experimental results indicate that the developed models can significantly support accurate and automatic detection of leaf diseases. The results have significantly demonstrated effectiveness of modified histogram equalization in terms of contrast adjustment and enhancement of image quality in detection of maize leaf diseases with a classification accuracy of 96.7%. However, in future this research work can be improved by combining neural network classifier with deep learning techniques using other plant leaf to capture more information that could be useful for better classification accuracy. Additionally, a comprehensive study is required to understand the factors affecting the detection of plant diseases, such as the classes and size of datasets, learning rate, illumination and so on.

REFERENCES

[1]. Aziz S., Bashir M., Mughal O., khan M.U. and Khan A. (2019)“Image Pattern Classification for Plant Disease Identification using Local Tri-directional Features”,IEEE 10th Annual Information Technology, Electronic and Mobile Communication Conference (IEMCON).

[2]. Ghaiwat S. N. and Arora P. (2014), "Detection and classification of plant leaf diseases using image processing techniques: a review," International Journal of Recent Advances in Engineering & Technology, vol. 2, no. 3, pp. 1-7.

[3]. Dhaygude S. B. and Kumbhar N. P. (2013), "Agricultural plant leaf disease detection using image processing," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 2, no. 1, pp. 599-602.

[4]. Bashir S. and Sharma N. (2012), "Remote area plant disease detection using image processing," IOSR Journal of Electronics and Communication Engineering, vol. 1, no. 6, pp. 31-34.

[5]. Sullca C., Molina C., Rodrgíuez C., and Fernández T. (2019) “Diseases Detection in Blueberry Leaves using Computer Vision and Machine Learning Techniques”, International Journal of Machine Learning and Computing, Vol. 9, No. 5, pp. 656-661.

[6]. Maniyath S.R., Ram H. (2018) “ Plant Disease Detection Using Machine Learning”, International Conference on Design Innovations for 3Cs Compute Communicate Control. Pp. 41-45

[7]. Saleem M. H., Potgieter J. and Arif K. M. (2019) “Plant Disease Detection and Classification by Deep Learning”, Plants, MDPI, Vol. 8, No. 468, pp. 1-22.

[8]. Hall, D.; McCool, C.; Dayoub, F.; Sunderhauf, N.; Upcroft, B (2015). Evaluation of features for leaf classification in challenging conditions. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa Beach, HI, USA,pp. 797–804.

[9]. Kamilaris, A.; Prenafeta-Boldú, F.X (2018). Deep learning in agriculture: A survey. Comput. Electron. Agric. , 147, pp. 70–90.

[10]. Amara, J.; Bouaziz, B.; Algergawy, A (2017). A Deep Learning-based Approach for Banana Leaf Diseases Classification. In Proceedings of the BTW (Workshops), Stuttgart, Germany, pp. 79–88.

[11]. Usha K., Kanagasuba R. S., Jaichandran R., Priyadharshini C. (2019). Detection and Classification of Paddy Crop Disease Using Deep Learning Techniques, International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, No 3, pp. 4353-4359.

[12]. Tsaftaris, S, Minervini, M & Scharr, H (2016), “Machine Learning for Plant Phenotyping Needs Image Processing” Trends in plant science, vol. 21, no. 12, pp. 989–991.

[13]. Alajrami, M. A., & Abu-Naser, S. S (2018).” Onion Rule Based System for Disorders Diagnosis and Treatment”, International Journal of Academic Pedagogical Research (IJAPR), 2 (8), pp. 1-9.

[14]. Belal A. M. Ashqar, Samy S. Abu-Naser (2018), “Image-Based Tomato Leaves Diseases Detection Using Deep Learning”, International Journal of Academic Engineering Research (IJAER), Vol. 2 No. 12, pp. 10-16.

[15]. Krizhevsky A., Sutskever I., Hinton, Geoffrey E (2017). "ImageNet classification with deep convolutional neural networks". Communications of the ACM. 60 (6): pp. 84– 90.

[16]. Rastogi A., Arora A. and Sharma S. (2015),” Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic” 2nd International Conference on Signal Processing and Integrated Networks (SPIN).

[17]. Mendeley Homepage,https://data.mendeley.com/datasets/tywbtsjrjv/1#file-127d0761- 7c63-46f0-b08e-d0d9f7cad9da, last accessed 2020/06/02.

[18]. Maniyath S. R., Vinod P., Niveditha M., Pooja R., Shashank N., and Hebbar R. (2018), "Plant Disease Detection Using Machine Learning," in *2018 International Conference on Design Innovations for 3CsCompute Communicate Control (ICDI3C)*,: IEEE, pp. 41-45.

[19]. Brahimi, M.; Boukhalfa, K.; Moussaoui, A (2017). Deep learning for tomato diseases: Classification and symptoms visualization. Appl. Artificial. Intelligence., 31, 299–315.

[20]. Polder, G.; Blok, P.M.; de Villiers, H.A.C.; van der Wolf, J.M.; Kamp, J(2019). Detection in Seed Potatoes Using Deep Learning on Hyperspectral Images. Front. Plant Sci., 10.

[21]. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens., 54, 6232–6251.

[22]. Venkatesan R. and Li B. (2017), *Convolutional Neural Networks in Visual Computing*, CRC Press.

[23]. Sannakki S.S. and Rajpurohit V. S. (2015),” Classification of Pomegranate Diseases Based on Back Propagation Neural Network,” International Research Journal of Engineering and Technology (IRJET), Vol 2 Issue:02.

[24]. Ibrahim M.A., Ojo O.A. and Oluwafisoye P.A.(2017)**“** On Feature Selection Methods for Accurate Classification and Analysis of Emphysema CT Images.”, *International Journal of Medical Imaging*, USA, SciencePG Vol. 5, No. 6, 2017, pp. 70-78.