Detection of Sigatoka Disease in Plantain Using IoT and Machine Learning Techniques

F. O. Sweetwilliams¹, V. O. Matthews¹, E. Adetiba¹², D. T. Babalola³, V. Akande³

¹Department of Electrical and Information Engineering, Covenant University, Ota, Ogun State, Nigeria
²HRA, Institute for System Science, Durban University of Technology, Durban, South Africa
³Centre for Systems and Information Services (CSIS), Covenant University, Ota, Ogun State, Nigeria

Corresponding Author; faith.sweetwilliams@covenantuniversity.edu.ng

Abstract

Achieving United Nations Sustainable Development Goal 2 (UN SDG2) infers an imperative to urgently increase food production by up to 70%. However, concerns have risen that increases in food production have not kept pace with increase in world population, which is estimated to reach 10 billion people by the year 2050. In this paper, an IoT with machine learning based system was developed to acquire and process significant indicators such as temperature, moisture, humidity and leave images for the detection of Sigatoka disease in plantain. Appropriate sensors for detecting the stated disease indicators were interfaced with Raspberry Pi3 microcontroller module to collate and transmit the sensor data wirelessly to ThingSpeak, which is the selected cloud based IoT platform. The acquired leave images were further processed using two image descriptors, namely: Scalable Color Descriptor (SCD) and Histogram of Oriented Gradient (HOG) to extract discriminative color and texture features respectively. The features were then classified to detect the diseased or non-diseased class using Multilayer Perceptron Artificial Neural Network (MLP-ANN). The best accuracy of 98% was produced using the HOG descriptor.

Keywords: ANN, HOG, IoT, MLP, Precision Agriculture, SCD

1.0 Introduction

According to the Consultative Group on International Agricultural Research (CGIAR), plantains and bananas (Musa spp. L.) are staple foods for about 400 million people in many developing countries, particularly in Africa. Its overall global production ranks fourth after maize, rice and wheat. It is an important horticultural crop and is among the ten most important food security crops that feed the world (United States Department of Agriculture, 2012), [1]. However, its production is hampered by several pests and diseases, which include banana bunchy top virus, burrowing nematode, banana weevils, fusarium wilt, banana xanthomonas wilt and sigatoka. Sigatoka is a disease considered the most economically important disease of plantain worldwide, causing yield losses up to 50%. It is a leaf spot disease of plantains and banana caused by a windborne fungus, Mycosphaerella fijiensis with susceptibility by all plantain and most banana cultivars [2], [3]. The fungus M. fijiensis attacks the leaves, causing necrosis either partially or entirely, thus limiting the photosynthetic foliage area. As research efforts are been directed away from intensive agricultural practices such as high rate of fertilizer and pesticides application due to their detrimental effects on the environment, the need for automatic and highly precise methods to fight this threat has become very pertinent.
[4] identified the pathogen associated with Sigatoka disease of banana in South Africa due to a severe outbreak of the disease in the growing season of 1999/2000. They carried out an extensive survey of the five banana producing regions. In their work, they established that the causal organism of Sigatoka disease in South Africa was M. musicola. This was achieved by carrying out morphological and molecular studies on the 163 leaf samples showing disease symptoms collected from the five banana producing regions which were taken to the labs in envelopes and maintained at 5°C over 2 to 3 days of processing.

[5] proposed a software prototype system for detection of black sigatoka disease at a matured stage on the leaf of a banana plant. In their work, Image processing techniques were applied on captured banana leaf images. First, image growing and segmentation were done, then further feature extraction was carried out for the detection and computation of area of infection.

[6] identified isolates of mycosphaerella: M. musicola and M. fijensis which are the causal agents of Sigatoka disease of banana in South-East Asia. Difficulties and confusions encountered in the differentiation of the sigatoka disease causative agents establishes a high degree of relevance for the study. The work takes advantage of advances in molecular biology techniques, using the Polymerase Chain Reaction (PCR)-based method of Random Amplification of Polymorphic DNA (RAPD) to identify the pathogens. It was established that RAPD can be used to differentiate species of mycosphaerella, study the geographic distribution and spread of M.fijensis in South-East Asia.

[7] proposed to address challenges arising from disease diagnosis of mycosphaerella complex of banana. Prior diagnosis of the disease was primarily based on host symptoms and fungal fruiting structures, hence precautionary control approaches were stalled. In their work, they presented a robust spiec-specific molecular-based technique with Polymerase Chain Reaction (PCR). Conventional specie-specific primers and taqman real-time PCR assays were developed that could detect DNA at very low concentrations. They indicated future intentions of the usage of the proposed tools for the implementation of decision support systems for disease management.

[8] evaluated the effect of Microbial Fungicide (MF) based on Bacillus subtillis EA-CB0015 biomass and its metabolites for the control of Sigatoka Disease (SD) in greenhouse and field conditions. Experiments conducted established that the MF achieved a disease control level comparable to both protectant and systemic programs. When compared with the protectant fungicide, chlorothalonil in greenhouse and field, a reduction of 20.2% and 28.1% respectively was achieved. When the MF was incorporated into different programs of systemic fungicides, disease potency was reduced by up to 42.9% indicating no appreciable difference from conventional programs. Although public acceptance and high cost of biological products are identified as deterrents, the introduction of the MF for the control of SD, either as water suspension or combined with systemic fungicide will reduce the conventional fungicide load and solve the problem risk of fungicide resistance developing in a pathogen.

[9] proposed a study investigating various computer vision techniques leading to the development of an algorithm to detect Banana Bacterial Wilt (BBW) and Banana Black Sigatoka (BBS) in banana plant. The algorithm can be broken down into four primary parts: in the first part,
images of banana leaves are acquired by a high resolution digital camera. In the second part, the desired features are extracted from the captured images to obtain the vectors. In the third part, images are classified into two categories of healthy and diseased leaves. Seven classifiers were selected for the experiments, out of which the Extreme Randomized Trees gave the best performance of 0.96 Area Under Cover (AUC) for BBW and 0.91AUC for BBS. Based on the results the best model for automatic detection of the banana diseases was recommended.

The study at hand takes advantage of recent advances in IoT and machine learning to build an automatic Sigatoka detection model for plantain.

2.0 Materials and Methods

The architecture for implementing the Sigatoka disease detection in this study comprises of several blocks as shown in Figure 1. The materials and methods deployed to realize each of the blocks are succinctly described in this section.

Figure 1: System Architecture

2.1 Data Acquisition and Image Pre-processing

The Raspberry Pi 3 micro controller was selected to coordinate the deployed sensor nodes which include Raspberry Pi camera module Rev 1.3 of 5 megapixels, DS18B20 one wire temperature sensor, YL-38 soil moisture sensor and AM2301 humidity sensor. Real time values were captured and transferred to the ThingSpeak platform (cloud service platform) for visualization.

One hundred plantain leaf images were captured in subsets of 50 healthy leaves and 50 diseased leaves in JPEG format using the Raspberry PI camera from various trees at a Plantain plantation located around Ota, Ogun State, South-West of Nigeria. In this study, the acquired leaf images were pre-processed by resizing them to 250 x 150 pixels.
2.2 Features extraction

In pattern recognition literature, feature is a term that connotes a descriptor. Repeatability is a necessary characteristic of descriptors. This is the tendency that identical features will be discovered in different images. Features play a fundamental role in classification and image processing. Image features often comprise of color, shape and texture.

According to [10], it is possible to use color to describe and represent an image. The MPEG standard group has proposed a number of descriptors for both color and texture information. Examples of proposed and established color descriptors include Dominant Color Descriptor (DCD), Scalable Color Descriptor (SCD), Group of Frame (GoF), Color Structure Descriptor (CSD), Local Color Vector Binary Patterns (LCVBP) and Color Layout Descriptor (CLD). For simple image textural description, First Order Statistics (FOS) are often engaged by researchers. Meanwhile, they usually pose low performance because they provide information about the grey level distribution of image pixels alone with no consideration for pixel spatial information. A large variety of texture descriptors that are already reported in the literature with far reaching performances over FOS include, Haralick Gray Level Co-occurrence Matrix(GLCM), Sum and Difference Histogram(SDH), Local Binary Patterns (LBP) and its variants, Homogeneous Texture Descriptor (HDT), Local Directional Ternary Pattern (LDTP), Histogram of Oriented Gradient (HOG) and several others [11].

According to [12] out of all image features, color and texture are more visually expressive and are thus attractive descriptors for enhanced and efficient image retrieval. Hence, for this study, SCD and HOG color and texture descriptors respectively were selected for experiments [12], [13], [14].

2.3 Pattern Classification

Given a training sample $X$, the goal of pattern classification is to find the right class out of $M$ classes to which a query sample $y$ belongs [15]. Pattern classification is a machine learning approach that has attracted a lot of attention over the years due to its wide areas of applications in domains like communications, automation, speech recognition, data mining, bioinformatics, agriculture, medicine and etc. It is categorised into supervised or unsupervised based on the type of learning procedure that is engaged to generate the output value. Supervised pattern classification assumes the availability of training set with manually labeled instances with the corresponding classes while unsupervised classification assumes the availability of training set with unlabeled instances in which inherent patterns are learnt automatically. Examples of supervised pattern classifiers are decision tree, random forest, Support Vector Machine (SVM) and Artificial Neural Network (ANN) while Kohonen Self Organising Map (SOM) and Karhunen-Loeave expansion are examples of unsupervised methods [16], [14], [17], [18], [19], [20].

Among the supervised techniques herein listed, ANN has been attributed with quality performance across different problem domains and it is thus selected as the pattern classification method for this study. It is a computational model that is inspired by the structure, processing method and learning ability of the biological neural system. One of the most commonly engaged family of ANN for pattern classification is the feed-forward Multi-layer Perceptron (MLP). Its network architecture as shown in Fig. 2.0 comprises of the input neurons which form the input layer, the output layer and the hidden layers which are between the input and output layers. For the
ANNMLP to perform optimally, parameters such as the appropriate topology, initial weight, appropriate training algorithm and the training datasets are necessary.

![ANN-MLP Architecture](image)

**Fig. 2.0: ANN-MLP Architecture**

### 2.4 Experimental Setup

The signal from each of the sensor nodes was streamed to the ThingSpeak IoT platform throughout the period of the experiment and the outputs are presented in the results section. The discriminative color and texture features for each instance of the acquired leave images both for healthy and diseased plantains were extracted using the SCD and HOG descriptors. Thereafter MLP-ANN topologies were evolved based on each of the descriptors. The MLP-ANN trained using SCD contains 16 neurons in the input layer while the one trained using the HOG descriptor contains 81 neurons in the hidden layer. The features extraction and MLP-ANN codes and all experiments in this study were carried out in MATLAB R2017b environment. The Table below shows other generic specifications of the MLP-ANNs.

| S/N | Parameters | Description |
|-----|------------|-------------|
| 1   | Input layer neurons | Scalable Color Descriptor = 16  
     |            | HOG Descriptor = 81 |
| 2   | No of hidden layer | 1 |
| 3   | Hidden layer neurons | 10 - 100 neurons in step of 10 |
| 4   | Output layer neurons | 2 |
| 5   | Learning algorithm | Scale conjugate gradient |
Input neuron activation function | Purelin
---|---
Hidden and output neurons activation function | Tansig

3.0 Results and Discussion

Selected plots of the sensor data on ThingSpeak online IoT platform are presented in Fig. 3.0. The data acquisition spanned a period of one month and the variations in temperature, soil moisture and humidity are as illustrated. Sample of the leaf images acquired for this study are presented in Fig. 4.0. The greenness of all the leaves in Fig. 4(a) is indicative of the healthy state of the plantain plant while the black as well as yellow colorations of the leaves in Fig. 4(b) are indicative of the diseased state of the plantain plants from where they were cut.

![Figure 3.0: Plots of the Sensor Data on ThingSpeak](image)

![Figure 3.0: Leaf samples of healthy and diseased plantain](image)

Selected histogram plots of SCD (diseased and healthy) and HOG (healthy and diseased) are shown in Fig. 4.0. Some similarities in the shapes of the histograms for features in the same classes can be observed in the figures. The within-class similarity of features and vise versa provide a basis for appropriate classification of features into relevant classes by the succeeding pattern classifier.
The MLP-ANN results using the accuracy and MSE metrics for SCD as well as HOG descriptors are shown in Tables 2 and 3 respectively. As shown in Table 3, the MLP-ANN with HOG descriptor produced the best result with 98% accuracy and MSE of 0.0208 at 40 neurons in the hidden layer. The confusion matrix of this best performing MLP-ANN with HOG is presented in Fig. 5.0.
Table 2: MLP-ANN Result for the Scalable Color Descriptor (SCD)

| S/N | Hidden Layer Neuron | Accuracy | MSE  |
|-----|---------------------|----------|------|
| 1   | 10                  | 88       | 0.1268 |
| 2   | 20                  | 91       | 0.0857 |
| 3   | 30                  | 93       | 0.0653 |
| 4   | 40                  | 88       | 0.1192 |
| 5   | 50                  | 88       | 0.1079 |
| 6   | 60                  | 88       | 0.0884 |
| 7   | 70                  | 92       | 0.0868 |
| 8   | 80                  | 91       | 0.1159 |
| 9   | 90                  | 89       | 0.1024 |
| 10  | 100                 | 91       | 0.0902 |

Table 3: MLP-ANN Result for the Histogram of Oriented Gradient (HOG) Descriptor

| S/N | Hidden Layer Neurons | Accuracy | MSE  |
|-----|----------------------|----------|------|
| 1   | 10                   | 94       | 0.0560 |
| 2   | 20                   | 96       | 0.0293 |
| 3   | 30                   | 98       | 0.0251 |
| 4   | 40                   | 98       | 0.0208 |
| 5   | 50                   | 95       | 0.0458 |
| 6   | 60                   | 96       | 0.0475 |
| 7   | 70                   | 96       | 0.0390 |
| 8   | 80                   | 95       | 0.0521 |
| 9   | 90                   | 93       | 0.0534 |
| 10  | 100                  | 98       | 0.0223 |
Figure 5.0: The Confusion Matrix of the MLP-ANN with the Best Accuracy

To know the divergence between HOG and rectangular filters, [21] conducted experiments equating the two types of filters. The best features of the filters were carefully chosen and analyzed. The features were calculated on each of the 2,418 training images and their mean measured. The correlation between both filter features were also computed for each image. The results proved that HOG gave a better performance. This claim supports the results obtained in the study at hand.

A performance analysis carried out by [13] on the overall performance of HOG compared with Haar wavelet, PCA-SIFT and shape context using the MIT pedestrian database revealed that HOG gave much better performance.

4.0 Conclusions

An IoT and machine learning based platform for early detection of Sigatoka disease in plantain has been presented in this paper. Real time monitoring of critical parameters such as temperature, soil moisture and humidity data through the ThingSpeak IoT platform was achieved. The appropriate combination of machine learning based algorithms (HOG and MLP-ANN) to classify leaf images into healthy and diseased state was also achieved. In the future, we hope to increase the number of sensing components in order to extend the coverage of the soil and environmental parameters that correlate with healthy or diseased status of the crop. We will also acquire larger dataset to achieve better disease classification vis-a-vis accuracy for the machine learning module. Furthermore, state-of-the-art algorithms for machine learning such as deep learning and extreme learning machine will be explored.
Acknowledgement

The publication of this work is fully sponsored by Covenant University Centre for Research, Innovation and Development (CUCRID), Covenant University, Ota, Nigeria.

References

[1] Fagbola, B. O., Adejoro, M. A., & Odubanjo, A. O. (2008, October). Research focus on banana and plantain (Musa spp.): Nigerian perspectives. In IV International Symposium on Banana: International Conference on Banana and Plantain in Africa: Harnessing International 879 (pp. 859-864).

[2] Swennen, R. (1990). Plantain cultivation under West Africa conditions: A reference manual. IITA.

[3] Mobambo, K. N., Gauhl, F., Vuylsteke, D., Ortiz, R., Pasberg-Gauhl, C., & Swennen, R. (1993). Yield loss in plantain from black sigatoka leaf spot and field performance of resistant hybrids. Field Crops Research, 35(1), 35-42.

[4] Surridge, A. K. J., Viljoen, A., Crous, R. W., & Wehner, F. C. (2003). Identification of the pathogen associated with Sigatoka disease of banana in South Africa. Australasian plant pathology, 32(1), 27-31.

[5] Tanjim-Al-Akib, M., & Islam, N. (2003). Feature Optimization of Whitefly Detection Algorithm using Image Segmentation and Feature Analysis.

[6] Johanson, A., Crowhurst, R. N., Rikkerink, E. H. A., Fullerton, R. A., & Templeton, M. D. (1994). The use of species-specific DNA probes for the identification of Mycosphaerella fijiensis and M. musicola, the causal agents of Sigatoka disease of banana. Plant Pathology, 43(4), 701-707.

[7] Arzanlou, M., Abeln, E. C., Kema, G. H., Waalwijk, C., Carlier, J., Vries, I. D., ..., & Crous, P. W. (2007). Molecular diagnostics for the Sigatoka disease complex of banana. Phytopathology, 97(9), 1112-1118.

[8] Gutierrez-Monsalve, J. A., Mosquera, S., González-Jaramillo, L. M., Mira, J. J., & Villegas-Escobar, V. (2015). Effective control of black Sigatoka disease using a microbial fungicide based on Bacillus subtilis EA-CB0015 culture. Biological control, 87, 39-46.

[9] Amara, J., Bouaziz, B., & Algergay, A. (2017). A Deep Learning-based Approach for Banana Leaf Diseases Classification. In BTW (Workshops) (pp. 79-88).

[10] Laencina Verdaguer, S. (2009). Color based image classification and description (Master's thesis, Universitat Politècnica de Catalunya).

[11] Sandid, F., & Douik, A. (2016). Robust color texture descriptor for material recognition. Pattern Recognition Letters, 80, 15-23.
[12] Dorairaj, R. and Namuduri, K.R., 2004, November. Compact combination of MPEG-7 color and texture descriptors for image retrieval. In Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004. (Vol. 1, pp. 387-391). IEEE.

[13] Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In international Conference on computer vision & Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE Computer Society.

[14] Adetiba, E. and Olugbara, O.O., 2015. Improved classification of lung cancer using radial basis function neural network with affine transforms of voss representation. PloS one, 10(12), p.e0143542.

Adetiba, E. and Olugbara, O.O., 2015. Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. The Scientific World Journal, 2015.

[15] Xu, J., An, W., Zhang, L. and Zhang, D., 2019. Sparse, collaborative, or nonnegative representation: Which helps pattern classification?. Pattern Recognition, 88, pp.679-688.

[16] Rao, M.S. and Reddy, B.E., 2011. Comparative analysis of pattern recognition methods: An overview. Indian Journal of Computer Science and Engineering (IJCSE), 2(3), pp.385390.

[17] Popoola, S.I., Adetiba, E., Atayero, A.A., Faruk, N. and Calafate, C.T., 2018. Optimal model for path loss predictions using feed-forward neural networks. Cogent Engineering, 5(1), p.1444345.

[18] Goyal, H., Joshi, N. and Sharma, C., 2018. An Empirical Analysis of Geospatial Classification for Agriculture Monitoring. Procedia computer science, 132, pp.1102-1112.

[19] Kohout, J., Komárek, T., Čech, P., Bodnár, J. and Lokoč, J., 2018. Learning communication patterns for malware discovery in HTTPS data. Expert Systems with Applications, 101, pp.129-142.

[20] Alaba, P.A., Popoola, S.I., Olatomiwa, L., Akanle, M.B., Ohunakin, O.S., Adetiba, E., Alex, O.D., Atayero, A.A. and Daud, W.M.A.W., 2019. Towards a more efficient and costsensitive extreme learning machine: A state-of-the-art review of recent trend. Neurocomputing, 350, pp.70-90.

[21] Zhu, Q., Yeh, M. C., Cheng, K. T., & Avidan, S. (2006). Fast human detection using a cascade of histograms of oriented gradients. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 1491-1498). IEEE.