10.31341/jios.46.1.1
Open Access

UDC 004.65:004.85:636.09(669.1)
Original Scientific Paper

Early Cocoa Blackpod Pathogen Prediction with Machine Learning Ensemble Algorithm based on Climatic Parameters

Olofintuyi Sunday Samuel *olofintuyi.sundaysamuel@gmail.com*
Doctor of Philosophy at Achievers University
Owo, Ondo State Nigeria

Abstract

Machine learning has been useful for prediction in the various sectors of the economy. The research work proposed an ensemble SA-CCT machine learning algorithm that gives early and accurate prediction of blackpod disease to farmers and agricultural extension officers in South-West, Nigeria. Since data mining put into consideration the types of pattern in a given dataset, the study considered the pattern in climatic dataset retrieved from Nigeria Meteorological agency (NIMET). The proposed model uses climatic parameters (Rainfall and Temperature) to predict the outbreak of blackpod disease. The ensemble SA-CCT model was formulated by hybridizing a linear algorithm Seasonal Auto Regressive Integrated Moving Average (SARIMA) and a nonlinear algorithm Compact Classification Tree (CCT), the implementation was done with python programming. The proposed SA-CCT model gives the following results after evaluation. Precision: 0.9429, Recall 0.9167, Mean Square Error: 0.2357, Accuracy: 0.9444

Keywords: Blackpod Disease, Machine Learning, Data Mining, SA-CCT Model; SARIMA; CCT

1. Introduction

Data mining is a technique which makes use of patterns and knowledge in a given dataset. The information that is extracted from the dataset helps to improve existing research, and thus leading to productivity. Various fields such as education, health-care, engineering, agriculture and market analysis has found the application of data mining relevant [1]. Virtually, all sectors in Nigeria rely on information and communication technologies for an optimum productivity; this is because complex problem can be tackled with the aid of a computer. [2] pinpoints that the greatest problem of farmers across the globe is infestation of crops with disease and pest and how to protect their crop that is not far-fetched because of lack of specialized tools that will help in predicting when the conditions for the outbreak of the diseases is near. [3] stated in their work that there are two major requirements that must be fulfill during prediction of plant diseases, and these requirement includes accuracy and fastness. Cocoa has been a major economic crop and served as a major source of income for Nigeria over the decades and has drastically reduced because of low production that is caused by the incidence of diseases on the crop and farmers not willing to produce cocoa because of the outbreak of blackpod disease. Blackpod is the deadliest and destructive disease of cocoa globally, this disease affects every aspect of cocoa plant. And all the parts that are affect includes the pods, stems and roots, leaves and flower cushion. A great number of cocoa pods have been lost to diseases across the globe due to the unavailability of timely reports of outbreaks, epidemics in developing regions like West Africa is still a serious challenge [4]. Also, in the time past, number of algorithms have been developed and implemented to extract information and knowledge patterns which are used for prediction of diseases from a time series dataset but they are unable to accurately make the prediction correctly [5]. Therefore, there is a need to improve on the prediction model and also develop an ensemble system that gives timely report. Maina [6] proposed a vision-based model that was developed to classify plant diseases, the use of such models is applicable in situations where the crop has fully matured and its phenotypic characteristics are visible; hence such models may not be feasible for early detection and accurate determination of crop diseases. This is because a disease at advanced stage would have spread more than such models could capture. Therefore, there is indeed the need to utilize information technology in providing a model for early and accurate detection of crop diseases. It has been proven using both theoretical and empirical findings that when models are combined, there is an improvement in the performance of the hybridized model compared to the individual model(s). The performance of the hybridized model increases when the ensemble model are of different types. Decision support system for farmers will reduce the incidence of disease and improve food production [7], [8]. [9] presented C4.5 algorithm in machine learning for prediction of crop yield in Indian, the authors used climatic parameters in the proposed research work. [10] gave a new perspective on plant disease by using deep learning. This research work proposed an ensemble model that decompose a time series data into it respective linear and non-linear component. Research has it that Seasonal Auto Regressive Integrated Moving Average (SARIMA) is a good statistical model that is good for predicting the linear component of a dataset while Compact Classification Tree (CCT) is also good for predicting the non-linear component of the dataset. The hybridized predictive was formulated using SARIMA and CCT to form SA-CCT model.

2. Literature Review

[11] uses Convolutional Neural Network (CNN) in modelling the response to crop management. The author(s) compares the proposed structure with regression linear model, random forest and support vector machine using the same set of dataset but the proposed model gave a better accuracy than the other models. [12] compares the performance of three machine learning algorithms along with multiple regression for predicting the risk in stagonaspora nodorium in wheat. After evaluation, it was indicated that random forest gave an accuracy of 93%. [13] compared different mathematics-based techniques for data processing and prediction of possible fruit disease infection. Six significant weather variables and one variable representing the month in the year are selected as predictor variables. Prediction includes two most

important diseases of cherry fruit: monilinia laxa and coccomyces hiemalis. Authors failed to implement the predictive model to give farmers notification of diseases through mobile or internet network. Various researchers have conducted research between a linear model and non-linear model to see their performance on the same dataset. According to [14] in their research work, answered the question that at what condition does a non-linear model outperformed a linear model. They provided an answer by saying maybe the dataset is non-linear without much disturbance, therefore cannot expect linear model to outperform the non-linear model [15]. When a non-linear model is used to model a linear problems, a mixed results is obtained and hence, it is wise to develop a hybridized model for a time series dataset. The concept behind hybridization is to use the component of each model uniquely to capture what they are good at. Since both findings of empirical and theoretical suggest hybridization of model, most especially when the models are different in nature in order to improve the performance of the hybridized model [16]. In addition, knowing the characteristics of real world dataset is difficult, hence, hybridization of model that has both linear and non-linear capabilities would be a very good approach for forecasting. There has been an increment in the literature of hybridizing model since the work of [17]. This was further supported by [18] where they hybridized Radial Bais Function (RBF) and ARIMA models. [19] conducted a research work by hybridizing a rule based system and neural network for making prediction on the price of changes in stock index futures. [20] in their work hybridized ARIMA model with SVM for forecasting stock prices. Also, [21] did a research where they integrated SARIMA with SVM for making prediction. Furthermore, using a hybridization technique that breaks a time series data into its respective linear and non-linear component has been on the increase in literature which they also have shown better result than a single model. The linear SARIMA is used jointly with non-linear CCT to form SA-CCT to detect the relationship that exist in the dataset. There are three cogent reasons why the models used for prediction are hybridized. Firstly, it is extremely cumbersome to pinpoint whether the time series data is generated from non-linear or linear underlying process. With this, it is difficult to determine which model will be suitable hence, the concept of hybridizing the model for better prediction. Secondly, real world dataset is purely a linear and non-linear that is, they have both linear and non-linear pattern respectively. Neither SARIMA nor CCT will be adequate for modelling of the process. Hence, this problem can be solve by hybridizing the two technique. Lastly, it has being agreed from literature that hybridization perform well than a single model when it comes to prediction [16]. Prediction of plant diseases can be performed based on weather data by using the appropriate machine learning technique [22]. [23] concluded in their research that there will be estimated increase of 20% of cocoa diseases which will cost lost in the total worldwide production of cocoa production. There is a correlation between some of the cocoa diseases and weather condition. The next section discusses data visualization, creation of the SA-CCT model and the comparison of the proposed model with SARIMA and CCT.

3. Methodology

This section discusses the proposed model building and simulation. Dataset used for model training and implementation were collected from Nigeria Meteorological agency (NIMET) which covers five major cocoa producing states in South-West Nigeria. After data collection, the dataset were visualized with Principal Component Analysis (PCA). Models creation was based on climatic variables which determines the blackpod disease. After visualization, three different models were trained and evaluated. The models include SARIMA, CCT and an ensemble of SARIMA and CCT to form SA-CCT. Figure 1 below depicts the proposed model. Implementation of the work was done with python programming language.

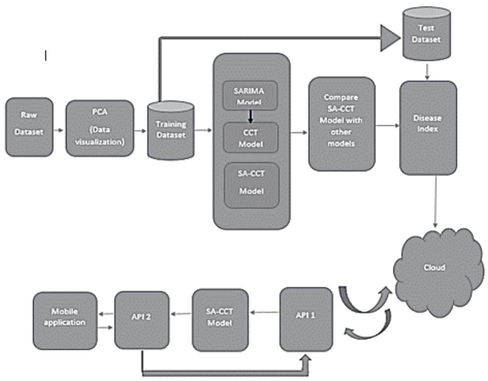

Figure 1. Proposed Model

3.1. Dataset

Dataset used for this research work was retrieved from Nigeria Metrological Agency (NIMET). The dataset covers five cocoa producing states (Osun State, Ogun State, Ekiti State, Oyo State and Ondo State) in South-West, Nigeria. Also, the dataset covers thirty year, from 1988 – 2018.

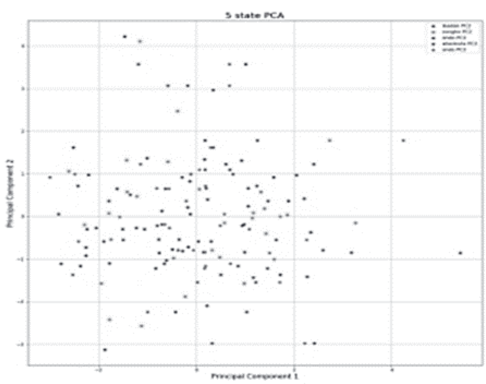

Figure 2. PCA for five states

Two major features were considered from the dataset which include temperature and rainfall. Data preprocessing and visualization was followed which was done with PCA. The dataset was put into the subspace for dimensionality reduction with the PCA using 2D plot diagram. Figure 2 below shows the plot diagram for the five states. The dataset was divided into training dataset and testing dataset which was used to test the accuracy of the three models.

3.2. SARIMA Model

SARIMA (p d, q, P, D, Q)$_s$ has two parts; Non-seasonal part (p, d, q) and seasonal parts (P, D, Q)$_s$

Where:

1. p: order of non-seasonal AR terms
2. d: order of non-seasonal differencing
3. q: order of non-seasonal MA terms
4. P: order of seasonal AR
5. D: order of seasonal differencing
6. Q: order of seasonal MA terms

SARIMA Process

SARIMA (p, d, q, P, D, Q)$_s$ has the form

$$\emptyset_p(B^s)\emptyset_p(B)(1-B^s)^D(1-B)^d X_t = \theta_Q(B^s)\theta_q(B)Z_t \quad (1)$$

Below is the polynomial form of Equation 1

$$\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q \quad (2)$$

$$\theta_Q(B^s) = 1 + \theta_1 B^s + \theta_2 B^{2s} + \cdots \theta_q B^{QS} \quad (3)$$

$$\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \emptyset_p B \quad (4)$$

$$\emptyset_p(B^s) = 1 - \emptyset_1 B^s - \emptyset_2 B^{2s} - \cdots \emptyset_p B^{ps} \quad (5)$$

Seasonal differencing, If D = 1

Then $\forall_s X_t = (1-B^s)X_t = X_t - X_{t-s}$

If D = 2

$$\forall_s^2 X_t = (1-B^s)^2 X_t = (1-2B^s+B^{2s})X_t$$
$$= X_t - 2X_{t-s} + X_{t-2s}$$

SARIMA popularity is due to its flexibility that is used on varieties of time series data. The basic assumption made to use this model is that it considers only the linear component of a time series data. SARIMA can only make prediction for the linear form associated with time series data.

3.3. Compact Classification Tree

CCT construct its decision tree making use of historical dataset. Classification of new dataset are then done with the constructed decision tree. When the class value is know, classification tree are then used for each of the instance. During the construction of classification tree, several parameters are considered such as the splitting rule, the rule that considered the splitting of data when been classified into smaller part. Research has it that there are various impurity functions such as Twoing splitting rule, Gini splitting rule and Deviance [24]. CCT is suitable for prediction making; the internal node is used to split instance space into various subspaces using some function of the attributes valves inputted. In some cases, a single attribute is tested and the space is divided with respect to the value of the attributes value. Considering a numeric attributes, the range of the condition is been used. For us to obtain the most appropriate targeted value, each leaf will be assign to a class. Alternatively, the probability vector that indicates the targeted attributes values is uphold by a leaf. Classification of instance is done from the root of the tree down to the leaf based on the test condition. Pruning method and stopping criteria are used to construct the complexity of the tree. Tree depth, total number of nodes, number of attribute used and total number of leaves are used to measure the tree's complexity.

Algorithm for CCT
TreeGrowth (A, B)

1. if stopping_cond(A,B) = *true* then
2. *leaf* = createdNode().
3. *Leaf.label* = Classify(A)
4. Return *leaf*.
5. Else
6. *root* = createdNode()
7. *root.test_cond* = find_best_split (A, B).
8. Let S = {s/s is a possible outcome of *root.test*_cond}
9. For each $s \in S$ do
10. A_s={sroot.test_cond(a)= $s, \alpha \epsilon A$
11. *Child* = TreeGrowth (A_s,B).
12. add *child* as descendent of root and label the edge *(root-child*) as s
13. end for
14. end if
15. Return root

3.4. SA-CCT Model

Two models were used in formation of the proposed model. SARIMA and CCT. SARIMA a statistical tool which had gained attention of researchers for linear problems and CCT machine learning algorithm which had also gained the attention of researchers over the years were hybridized to form SA-CCT model. SARIMA a statistical tool is good tool for detection and prediction of linear pattern in a given dataset while CCT is an algorithm which is also good for detection and prediction of non-linear pattern in a given dataset. Both linear and nonlinear models have achieved successes in their own linear or nonlinear problems. However, none of them is a universal model that is suitable for all situations. In the research conducted by [17] they said that a combined model having both linear and nonlinear modeling abilities will be a good alternative for forecasting the time series data. Both the linear and nonlinear models have different unique strength to capture data characteristics in linear or nonlinear domains, so the combined model proposed in this study is composed of the linear component and the nonlinear component. Therefore, the combined model can model linear and nonlinear patterns with improved overall forecasting performance. It may be reasonable to consider a time series to be composed of a linear autocorrelation structure and a nonlinear component.

Mathematical Representation of the Proposed Model

If W_t is a function of X_t and Y_t of the order of n

Where X_t is the linear component of the proposed model and Y_t is the non-linear component of the proposed model.

W_t can be written in the form below:

$$W_t = X_t^n f\left(\frac{Y_t}{X_t}\right) \dots \dots \quad (6)$$

Differentiating Equation 6 partially w.r.t 'X', we have

$$\frac{\partial W_t}{\partial X_t} = nX_t^{n-1}.f\left(\frac{Y_t}{X_t}\right) + X_t^n.f^1\left(\frac{Y_t}{X_t}\right)$$

$$\frac{\partial W_t}{\partial X_t} = nX_t^{n-1}.f\left(\frac{Y_t}{X_t}\right) - X_t^{n-2} \quad (7)$$

Multiplying both sides by X, we have:

$$\frac{\partial W_t}{\partial X_t} = nX_t.f\left(\frac{Y_t}{X_t}\right) - X_t^{n-1}Y_t.f^I\left(\frac{Y_t}{X_t}\right) \quad (8)$$

Differentiating Equation 6 partially w.r.t 'Y' we have:

$$\frac{\partial W_t}{\partial Y_t} = X_t^n f^I\left(\frac{Y_t}{X_t}\right)\frac{1}{X_t} \quad (9)$$

Multiplying both side by y, we get

$$X_t\frac{\partial W_t}{\partial Y_t} = X^{n-1}Y.f^I\left(\frac{Y_t}{X_t}\right) \quad (10)$$

Adding Equation 8 and 10 we have

$$= X_t\frac{\partial W_t}{\partial X_t} + Y_t\frac{\partial W_t}{\partial Y_t} = nX_t^n f\left(\frac{Y_t}{X_t}\right)$$

$$= X_t\frac{\partial W_t}{\partial X_t} + Y_t\frac{\partial W_t}{\partial Y_t} = nW_t \quad (11)$$

Where nW_t is the predicted values from both linear component and non-linear component.

3.5. Android Client

The android client represents an application (App) that will allow farmers to receive near real time information from the server. The android client is composed of an android mobile handset with an application that will allow farmers to interact with the system. Alert messages or notifications will be transmitted via an application-programming interface to the mobile phone warning farmers on the possibility of cocoa blackpod disease and possible control mechanism such as type of fungicide application. A farmer will use the android application to register their details via a simple registration form and the data will be sent to the server for storage. Figure 3 depicts the registration interface for farmers to register.

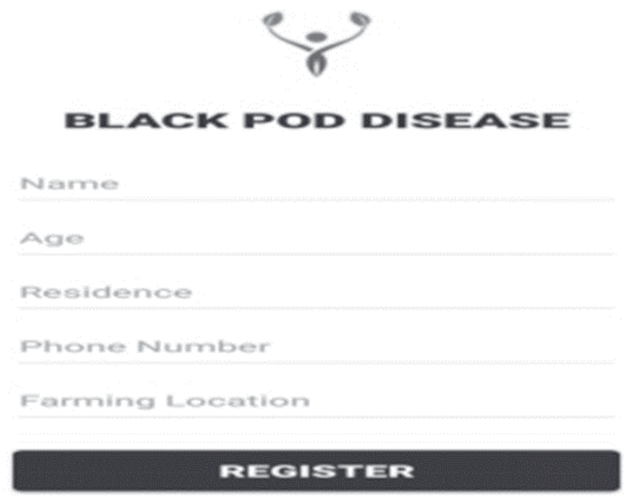

Figure 3. Interface for Registration

3.6. Server

The server acted as a host for both the farmer and climatic database. The database stored information received from the farmers. The climate database will store information on black pod cocoa disease and their symptoms. The server will as well host the data mining application that will use data from both the farmer database and disease database to determine the likelihood of occurrence of a crop disease.

3.7. Application Program Interface (API)

The API 2 gets user data from the application (APP) database. Through the API 2, the API 2 parsed the date to the API 1, once API 1 gets the date and state, it loads the data from the cloud and fit it into the Model.PKL, which make prediction and send it to the APP through API 1 for display. Figure 4 reveals the diagram of the model connection with two APIs and APP. The type of API used is a RESTFUL API that was provided by flask. Flask is a python web framework and it also provides service like Rest API. The API stands as a communication link between the third party software (such as mobile application and web application) and the model built. This third party software then sends a request in json format to the API.

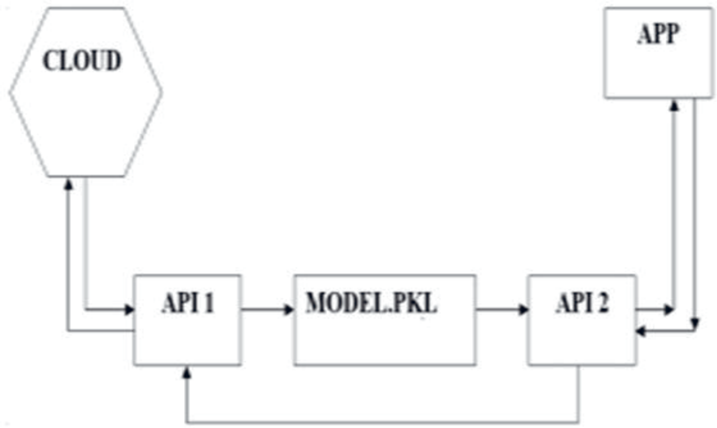

Figure 4. Diagram of the Model Connection with two APIs and APP

3.8. Performance Evaluation

Machine Learning has several ways of evaluating the performance of learning algorithms and the classifiers they produce. But for the essence of this research, the matric such as accuracy, recall, precision and means square error were used to evaluate our proposed model. They are thus expatiated below.

Recall: it talks about events that are truly positive and are announced as positive.

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{12}$$

Precision: it describes events that are negative and are announced as negative event.

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{13}$$

Accuracy: it describes the general effectiveness of the proposed model.

$$Accuracy = \frac{TP + TN}{TP + FN + FN + TN} \tag{14}$$

Means Square Error (MSE): it is an estimator of the overall deviations between predicted and the actual values.

$$MSE = \frac{1}{n} \sum_{t=1}^{n} e^2 t \tag{15}$$

Note: Where $e_t = y_t - f_t$ is the forecast error, and *n* is the size of the test set.
Where y_t is the actual value
And f_t is the forecasted value

4. Result and Discussion of SARIMA Model

Immediately after visualizing the climatic dataset from data space using the Principal component analysis, the dataset is then divided into two sections which are the training set and testing set. 70% of the dataset is used for training while 30% were used for testing. 10 folds validation was also used during the data separation. SARIMA model was simulated and evaluated with the following parameters; precision, recall, accuracy and mean square error. Also, two features were considered during the training and validation stage. The following evaluation was done after prediction: Accuracy: 0.8333%, Precision: 0.6316, Recall: 0.6316 and Mean Square Error: 0.4082. Figure 5 depicts the graph for SARIMA prediction.

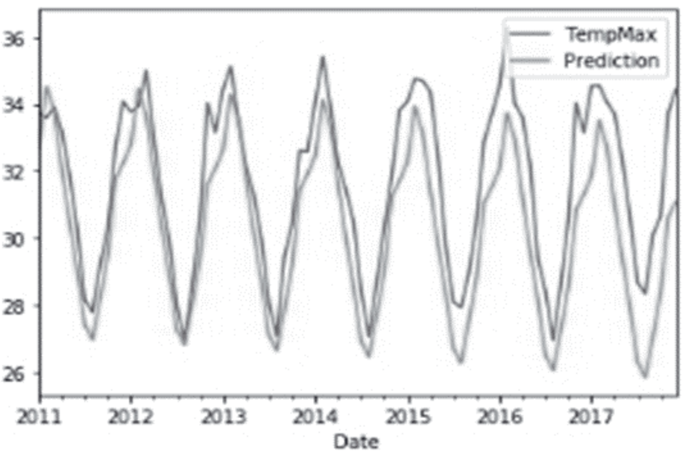

Figure 5. Outcome of SARIMA Prediction of Blackpod Disease

NB: The blue line shows the actual output and the orange shows the predicted output of the model when we fit in the test set

4.1. Result and Discussion of Compact Classification Tree Model

The model was developed using compact classification tree, the first thing done was to import the necessary library. After the successful load of the library, the pandas was use to load the dataset into the system, after that the date was set as the index for the database. The database was splitted into input and output after which the train_test_split library was imported from sklearn.model_selection and the input and output is split into train and test set. After that we load the compact classification tree classifier and fit the x_train and y_train into the classifier to train the classifier, after successful training the x_test was fit for prediction. The following evaluation was done after prediction: Accuracy: 0.911%, Precision: 0.8888, Recall: 0.8888 and Mean Square Error: 0.2981. Figure 6 depicts the graph for compact classification tree prediction.

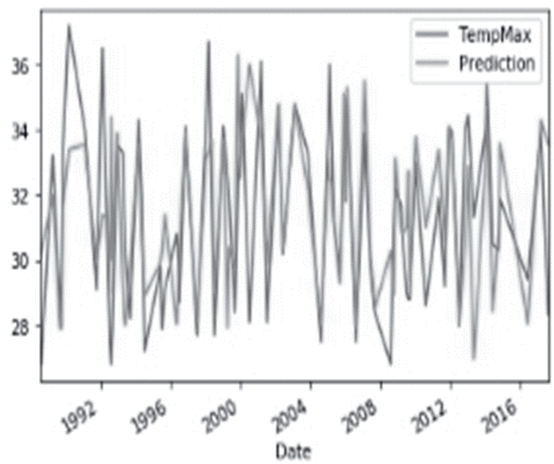

Figure 6. Outcome of Compact Classification Tree prediction of Blackpod Disease

4.2. Result and Discussion of SA-CCT Model

The model was developed using SA-CCT which is the combination of SARIMA and CCT. The following evaluation was done after prediction: Accuracy: 0.944%, Precision: 0.9429 and Recall: 0.9429. Figure 7 depicts the graph for compact classification tree prediction. After evaluation, we import pickle which is used to create model, the model is created and saved in the stated directory with .pkl extension and the model can be used to make predictions.

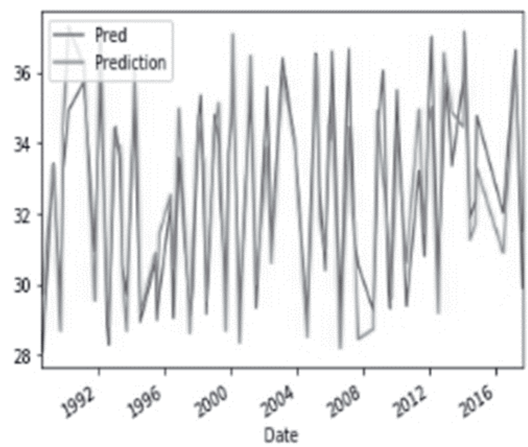

Figure 7. Outcome of SA-CCT Model for prediction of Blackpod Disease

NB: The blue line shows the actual output and the orange shows the predicted output of the model when we fit in the test set.

4.3. SARIMA, CCT and SA-CCT Performance Evaluation Comparison

The proposed SA-CCT model outperformed the other two models by giving the lowest mean square error that is, it gives the least deviations between predicted and actual values and gave the highest accuracy between the three models. As a result, SA-CCT was considered for making prediction to all registered farmers. Table 1 below shows the results of the proposed model, CCT and SARIMA model while figure 8 reveals the graphical representation of the results.

Metrics	SARIMA	CCT	SA-CCT
Precision	0.6316	0.8888	0.9429
Recall	0.6316	0.8888	0.9167
Accuracy	0.8333	0.9111	0.9444
MSE	0.4082	0.2981	0.2357

Table 1. The compared results of SARIMA, CCT and SA-CCT Models

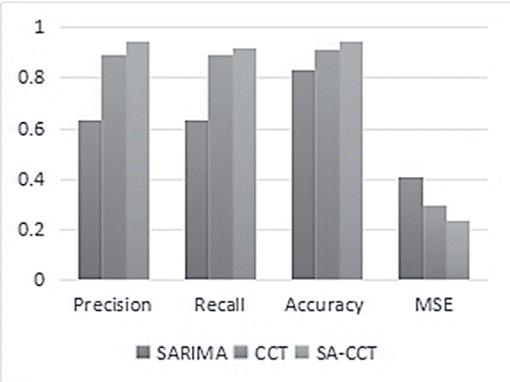

Figure 8. The compared results of SARIMA, CCT and SA-CCT Models

5. Conclusion

This work introduces an ensemble SA-CCT novel approach for predicting of blackpod disease by giving farmers in South-West Nigeria monthly notification of blackpod disease via their mobile application developed. Farmers are expected to install the application on their mobile phone and stay connected so as to receive the monthly notification. The research work also reveals that SA-CCT model performs better than the individual SARIMA (linear) and CCT (nonlinear).

References

[1] C. Aggarwal, Data mining: the textbook. Switzerland: Springer International Publishing Switzerland, 2015.

[2] A. Amin, A. Yousif, and F. Abdulrahman, “First use of radio telemetry to assess behaviour of red palm weevil, Rhynchophorus ferrugineus (Olivier) in the presence and absence of pheromone traps,” Computer and electronics in Agriculture, pp 170, 2020, 105252

[3] E. Arsevska, M. Roche, P. Hendrikx, D. Chavernac, S. Falala, and R. Lancelot, “Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web,” Computers and electronics in agriculture, 123, 2016, 104-115.

[4] M. Peter, A. Abiodun, O. Oyedeji, and R. Adegboyega, “Black pod disease profile: monitoring its outbreak in Southwest, Nigeria,” Plos one, doi.org/10.137/journal.pone, 2020, 0209306

[5] U. Ayub, and S. Moqurrab, “Predicting crop diseases using data mining approaches: Classification. 2018 1st International Conference on Power, Energy and Smart Grid (ICPESG),” https://doi:10.1109/icpesg, 2018, 8384523

[6] C.N. Maina, “Vision-based model for maize leaf disease identification: a case study in Nyeri County,” 2016

[7] U. Angel, P. Marco, D. Céline, C. Cesar, A. Abdou, C. Luigi, and C. Javier, “A decision support tool to enhance agricultural growth in the Mékrou river basin (West Africa),” Computers and Electronics in Agriculture, 154, 2018, 467-481.

[8] Z. Zhaoyu, F. Jose, B. Victoria, and L. Nestor, “Decision support systems for agriculture 4.0: Survey and challenges,” Computers and electronics in Agriculture. 170, 2020, 105-256

[9] S. Veenadhari, B. Misra, and C. Singh, “ Machine learning approach for forecasting crop yield based on climatic parameters,”International Conference on Computer Communication and Informatics 2014, Coimbatore, INDIA

[10] H. Sue, G. Herve, B. Pierre, and J. Alexis, “New perspectives on plant disease characterization based on deep learning,”Computers and Electronics in Agriculture 170, 2020, 105-220

[11] B. Alexandre, T. Rodrigo, H. Naira, M. and Nicolas, “Modeling yield response to crop management using convolutional neural networks,” Computers and Electronics in Agriculture, 170, 2020, 105-197

[12] M. Lucky, C. Christina, G. Kevin, and O. Peter, (2016) “Predicting pre-planting risk of Stagonospora nodorum blotch in winter wheat using machine learning models,” Frontiers Plant Sci. 7,

[13] I. Milos, I. Sinisa, J. Srdjan, and P. Stefan, “Early cherry fruit pathogen disease detection based on data mining prediction,” Computers and Electronics in Agriculture, 150, 2018, 418-425

[14] Z. Tang, and P. Fishwick, "Feedforward neural nets as models for time series forecasting," ORSA Journal on Computing 5(4), 1993, 374–385.

[15] G. Zhang, B. Patuwo, and M. Hu, "Forecasting with artificial neural networks: The state of the art;" International Journal of Forecasting, 14, 1998, 35-60

[16] G. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," Informatic Science, 177, 2007, 5329-5346.

[17] J. Bates, and W. Granger, "The combination of Forecast operational research quarterly," 20(4), 1969 451-468

[18] D. Wedding, and K. Cios, "Time series forecasting by combining RBF networks, certainty factors, and the Box–Jenkins model," Neurocomputing, 10, 1996, 149–168.

[19] R. Tsaih Y. Hsu, C. Lai, "Forecasting S&P 500 stock index futures with a hybrid AI system; Decision Support Systems," 23, 1998, 161–174

[20] P. Pai, and C. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting; Omega 33, 2005, 497–505.

[21] K. Chen, and C. Wang "A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan," Expert Systems with Applications, 32, 2007, 254– 264.

[22] V. Chaurasia, and S. Pal, "Early prediction of heart diseases using data mining techniques. Caribbean Journal of Science Technology. 1, 2013, 208–217 ISSN: 0799-3757.

[23] C. Rettinassababady, and C. Jeyalakshimi, (2014) "Bio-fungicides: The best alternative for sustainable food security and ecosystem," Doi:10.1007/978-81-322-1801-2_35

[24] Math Works, Compact Classifification Tree. Retrieved from https://www.mathworks.com/help/stats/compactclassifificationtreeclass.html?requestedDomain=www.mathworks.com, on March 20, 2019,