Heliyon 9 (2023) e15245

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

An ensemble deep learning approach for predicting cocoa yield

Sunday Samuel Olofintuyi [a,*], Emmanuel Ajayi Olajubu [b], Deji Olanike [c]

[a] Department of Computer Science, Achievers University, Owo, Nigeria
[b] Department of Computer Sciences and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
[c] Department of Agricultural Extension and Rural Development Obafemi Awolowo University, Ile-Ife, Nigeria

ARTICLE INFO

Keywords:
CNN-RNN with LSTM
Cocoa yield prediction
Ensemble model
Machine learning

ABSTRACT

One important aspect of agriculture is crop yield prediction. This aspect allows decision-makers and farmers to make adequate planning and policies. Before now, various statistical models have been used for crop yield prediction but this approach experienced some hiccups such as time wastage, inaccurate prediction, and difficulties in model usage. Recently, a new trend of deep learning and machine learning are now adopted for crop yield prediction. Deep learning can extract patterns from a large volume of the dataset, thus, they are suitable for prediction. The research work aims to propose an efficient deep-learning technique in the field of cocoa yield prediction. This research presents a deep learning approach for cocoa yield prediction using a Convolutional Neural Network and Recurrent Neural Network (CNN-RNN) with Long Short Term Memory (LSTM). The ensemble approach was adopted because of the nature of the dataset used. Two different sets of the dataset were used, namely; the climatic dataset and the cocoa yield dataset. CNN-RNN with LSTM has some salient features, where CNN was used to handle the climatic dataset, and RNN was employed to handle the cocoa yield prediction in southwest Nigeria. Two major problems generated by the CNN-RNN model are vanishing and exploding gradients and this was handled by LSTM. The proposed model was benchmarked with other machine learning algorithms based on Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). CNN-RNN with LSTM gave the least mean of absolute error as compared to the other machine learning algorithms which shows the efficiency of the model.

1. Introduction

One major concern for agriculture scientists over the globe is how to enhance yield production for some major cash and food crops. This is a result of the demand for food from the ever-increasing population of the world [1,2]. Some of the major crops include cocoa, maize, millet, and wheat [3,4]. The attention of researchers is growing rapidly concerning these crops because they constitute significant sources of nutrition worldwide also, these crops face some hiccups such as a change in climate and natural resources depletion which has directly affected their products globally [5,6]. Cocoa, which is a major cash crop in Nigeria, contributes a significant percentage to the country's foreign exchange and a lot of products such as cocoa butter, cocoa powder, chocolate liquor, milk chocolate, and chocolate spread are produced from cocoa. Crop yield prediction is of great advantage to farmers and agricultural extension workers because it helps in timely import and export decisions and knowledgeable financial and management decisions are

* Corresponding author.
E-mail address: olofintuyi.sundaysamuel@gmail.com (S.S. Olofintuyi).

https://doi.org/10.1016/j.heliyon.2023.e15245
Received 1 September 2022; Received in revised form 29 March 2023; Accepted 30 March 2023
Available online 5 April 2023

made based on it [7]. The high demand for cocoa and unexpected environmental fluctuation sends a warning signal to agricultural scientists to increase its production [8]. Environmental factors such as climate among other factors affect the yield of cocoa produced in southwest Nigeria and this makes the prediction of cocoa yield more challenging because the climatic dataset comes in both linear and non-linear forms, hence the difficulties in estimating its accuracy. Recently, the hope of farmers and agricultural extension workers was rekindled by the introduction of sensor technology [9,10], data science [11], and Machine Learning (ML) [12] which has been used to improve their production. Various ML such as Artificial Neural Networks (ANN), Bayesian Network (BN), Random Forest (RF), and Regression Trees (RT) have been used for yield prediction [13–15].

ML has found its application in the following field of agriculture, namely; crop disease prediction [16–18] and yield prediction [19] based on climatic factors and agricultural practices [20]. Two major factors that affect yield prediction, these factors are controlled and uncontrolled. The controlled factors are practices the farmers adopted during farming practices such as the type of seeds that were planted, fertilizer used, number of times irrigation was done on the farmland, weeding, and other practices the farmer engaged in on the farmland [21]. Climatic factors (rainfall, humidity, and temperature) and soil characteristics happen to fall into the categories of uncontrolled factors that affect the prediction of crop yield. The predicted yield which is output from ML is dependent on the implicit input variable, the implicit input variable which is the climatic component is a nonlinear function and complex [7]. All the aforementioned factors and crop yield exhibits non-linearity among themselves, hence, a robust model is needed to handle the task. ANN has shown promising results in the area of crop yield prediction [22] because of its inherent ability to handle both linear and non-linear forms of data, with this capability of ANN, many researchers now adopt ANN for their research work. Recently, deep learning models are now being used over other ML models because of their in-built capability which enhances feature learning and automatic feature extraction from raw datasets thus producing better accuracy for predicting crop yield [23]. It was noted from previous studies that statistical tools such as Auto regression Integrated Moving Average (ARIMA) and Seasonal Auto regression Integrated Moving Average (SARIMA) have been used for predicting cocoa yield in Nigeria but the accuracy results have not been encouraging, because of these, this research work takes advantage of the state of the art to improve the accuracy by adopting an ensemble deep learning approach for predicting cocoa yield based on the climatic dataset and cocoa yield dataset. Also, Convolutional Neural Network (CNN) can be used to extract climatic parameters because it is suitable to process data with multiple array formats such as one-dimensional data (Signals and Sequences), two-dimensional data (images), and three-dimensional data (videos). The temporal dependencies of the climatic dataset used were captured by one – dimensional convolution. With CNN all the necessary information and the high-level features from the input data can be extracted by CNN when paired with the pooling operation. All features of the input dataset can be summarized by CNN when a filter is applied to the input data with these, CNNs are easier to train. While RNNs are suitable for tasks that involve sequential data so as to capture their time dependencies. The usage of the historical cocoa yield dataset as an input in this work also enables our model to predict cocoa yield. Finally, motivated by the high predictive performance of CNNs and ensemble models in ecology [7,24–28], we propose an ensemble model for cocoa yield prediction in southwest Nigeria.

Generally, optimization techniques and more advanced hardware are needed whenever deep learning models are to be trained because the loss of function is extremely high which makes the optimization of such function more difficult, and also, there is a problem of vanishing and exploding gradients [29]. The research work aimed to present an ensemble deep learning approach to predict cocoa yield in southwest Nigeria which is against the previously used statistical method and compared the results with other machine learning algorithms. The major contributions of the approach are listed below. 1) To predict cocoa yield more accurately, we proposed a new ensemble deep neural network model which captures time dependencies of climatic and yield datasets. 2) The CNN-RNN with LSTM combines CNN and RNN with LSTM. CNN was used to capture and preprocess climatic dataset that serves as the independent variable while RNN was used to capture and preprocess cocoa yield which serves as the dependent variable 3)The study also suggests an ensemble deep learning for yield prediction instead of statistical method which is tedious, inaccurate and time-consuming.

The area considered for this research are states that are within the Southwestern region of Nigeria. These are Osun State, Oyo State, Ondo State, Ekiti State, and Ogun State. They all have ecosystems that are characterized by humid and sub-humid. The states are

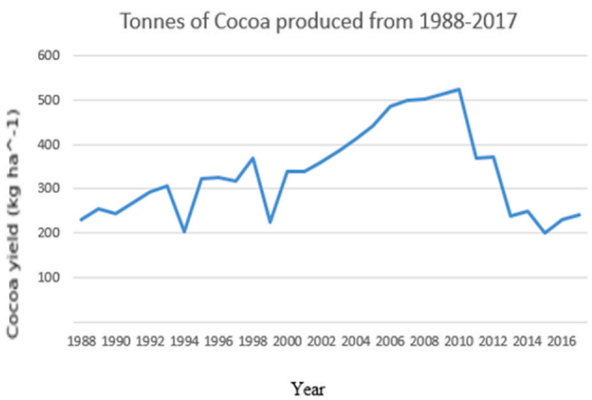

Fig. 1. Tonnes of cocoa produced from 1988 to 2017.

surrounded by Kwara and Kogi in the North and East respectively while they are bordered by the Republic of Benin in the West, and the Atlantic Ocean serves as a boundary in the South. Research has it that most of the farmers in Southwestern Nigeria are into crop production, so also in the savanna and rainforest where about 90% are dominantly into crop production while about 37.8% in the region that are swampy area are involved in fish farming. In this study, Lagos was excluded from the states sampled in Southwestern Nigeria because the state is regarded as the hub of urban business and most of its land is not used for farming. Fig. 1 shows the tons of cocoa produced from 1988 to 2017. Cocoa is usually planted around April in South West Nigeria and the harvesting is done around October. The data for these months were targeted for both the yield and climate. The following are the main cash crops that are planted in the zone: oil palm, rubber, and cocoa. The focus of this research was based on one of the cash crops – precisely, cocoa.

The next section takes a look at the various approaches that have been used for yield prediction with machine learning and deep learning. Section two describes the methodology used and how the model operates while sections four and five describe the results from our findings and its conclusion respectively.

22. Related work

Researchers over the globe have adopted the ML approach for crop yield prediction and the results obtained have been promising. Yield prediction of potato, wheat, and maize was carried out by Ref. [30] authors used random forest and multiple linear regression for prediction, and experimental results show that random forest outperformed multiple linear regression. [3] in his research work adopted SARIMA for prediction but the result obtained provides room for improvement. [31] uses SARIMA model to forecast dengue cases. Experiment results show that the interval between the historical data and the predicted results is not too close but can be improved. Also, [32] did a comparative analysis among various ML algorithms, namely; Support Vector Regression (SVR), Multiple Linear Regression (MLR), k-nearest neighbor, M5-Prime regression tree, and perceptron multilayer neural networks. After the experimental results, M5-Prime gave the least error in yield prediction thereby making it the most suitable for yield prediction among the models. In another study, the random forest was used by Ref. [33] to predict the yield of mango fruit under irrigation regimes. It was concluded in the study that RF is applicable for predicting mango yield even in underwater management. [34] Classified the yield of durum wheat with association rule mining and decision tree, experimental results show that association rule mining gave the best result in all the locations. [35] Conducted a research where ML was used on remote sensing data to estimate nitrogen status and yield prediction. Experimental results show that among all ML algorithms used, the M5- Prime regression algorithm was more efficient for yield prediction whereas, for nitrogen estimation, the least square Support vector machine gave the best result. It was concluded in the research that there is an increase in accuracy and efficiency by fusing sensor technology into machine learning. [36] adopted ANN for the prediction of rice yield with some climatic parameters such as minimum and maximum temperature and precipitation in some selected districts of Maharashtra, India. It was concluded that ANN gave a specificity of 0.9819, a sensitivity of 0.9633, and an accuracy of 0.975. [37] compared the performance of Default ANN (D-ANN) and customized ANN(C-ANN) for wheat yield prediction, and the report reveals that C-ANN was more efficient than D-ANN for wheat yield prediction. In other studies, [38] adopted advanced sensing techniques and machine learning for wheat yield prediction. The wheat yield prediction was based on various non-climatic factors and the features that affect yield growth was captured with supervised self-organizing maps. Supervised Kohonen Networks (SKNs) were compared with other models such as XY-fused Networks (XY-Fs) and counter-propagation ANN (CP-ANNs) for wheat yield prediction accuracy, SKN gave an accuracy of 81.65%, CP-ANNs gave 78.3% and XY-Fs gave 80.92%. From the results derived, SKN is more suitable for wheat yield prediction than the other models. In another study, [16] proposed an ensemble model for cocoa black pod disease prediction in southwest Nigeria based on climatic parameters. The ensemble model was made of a Compact Classification Tree (CCT) and Seasonal Autoregressive Integrated Moving Average (SARIMA). The climatic parameters consist of minimum and maximum temperature and rainfall for five cocoa-producing states. The proposed model was compared with other models and gave a better accuracy compared to other models after simulation. Ref. [39] proposed a Spiking Neural Network (SNN) architecture for crop prediction. Spatial and temporal variation for crop yield was achieved based on the remote sensing data used, and the accuracy achieved by SNN was higher compared to other models. The results produced by ANN in the field of crop prediction have necessitated its tremendous usage, making researchers explore other advanced areas of neural networks such as deep learning.

In recent times, deep learning has been used by researchers to predict crop yield. A Convolutional Neural network, which is a good example of deep learning has been used extensively by researchers on large image datasets, thus, is useful for crop yield classification and crop yield prediction. Ref. [40] Predicted crop yield with a deep learning approach and the performance of the model was compared with regression tree, Lasso, and shallow neural networks. Results reveal that the model used outperformed other models. In another study, Ref. [28] predicted soya beans yield with Convolution Neural Network-Recurrent Neural Network (CNN-RNN), the model proposed was unique because a dimensionality reduction technique was applied to the remotely sensed images dataset used for its training and prediction. Ref. [41] predicted crop yield with a deep neural network using metrological datasets that span from 2006 to 2015. The prediction was based on optimized input variables which were derived from satellite products. In similar work, a soya bean was predicted based on a deep learning framework. Argentina was used as a case study for the research and the results that were derived were satisfactory when a transfer learning approach was adopted to predict soya bean harvest based on small amounts of data in Brazil [42]. In another study, Ref. [27] did a study to investigate the ability of CNN in forecasting rice yield grain. Datasets of remotely sensed images were used for the research work, experimental results show that CNN gave a robust grain forecast in all the ripening stages. Corn yield loss was predicted with CNN which cut across different locations in Canada and the United States [43]. In a similar study, the yield of orchards was estimated with CNN, and the estimation of the plant growth was based on its various stages. The area from the image that represents was extracted by CNN because it serves as an object detector [44]. Ref. [42] adopted the concept of transfer learning because deep learning needs a large dataset, and an extensive amount of dataset is needed whenever a deep learning approach is to be adopted, the results derived from this approach have been encouraging.

2.1. Existence of the state of the art

Researchers over the globe have explored the various deep learning models, the effectiveness and efficiencies of each deep learning model have been explored in the area of prediction of crop yield, and they have all proven to be suitable. However, there is one of the models RNN seems to be efficient because of its capacity in dealing with time series datasets. Time series data plays an important role in the field of crop yield prediction because yield prediction depends on temperature and rainfall. RNN seems to be a perfect model to handle problems posed by time series data for yield prediction. In recent times, some researchers the likes of [45] used a fusion of CNN and Long Short Term Memory (LSTM) to estimate the wheat crop. A raw imagery dataset was used during the experiment and the results obtained were promising. The proposed fusion of CNN and LSTM outperformed other convolution methods by 74% accuracy and outperformed other deep learning models by an accuracy of 50%.

Ref. [46] compared the performance of RNN and Feed Forward Neural Network (FFNN) for the prediction of wheat yield, the metric was based on RMSE. Experimental results reveal that RNN gave better performance compared to FFNN. Another study was done by Ref. [20] Crop yields were predicted by LSTM in Iowa states. The efficiency of RNN was justified by the result derived from the experiment. In another study, Ref. [47] predicted soybean and maize yield crops, where climatic data and soil data were used. The prediction was based on a neural network model. The soil data were firstly fed into the fully connected layer while the climatic dataset was provided through the recurrent LSTM layer. The results obtained from the proposed model were encouraging compared to other results. Ref. [48] employed the LSTM model for the prediction of wheat yield, some smoothening functions were introduced to the proposed LSTM model, and this was done to increase the efficiency of the model. Experimental results reveal that the proposed model gave better accuracy.

Ref. [49] compared the performance and efficiency of various classifiers such as the xgboost classifier, KNN classifier, random forest classifier, logistic regression, and RNN with LSTM. The performance was based on the prediction of crop yield, two major parameters rainfall and temperature were considered for yield prediction. RNN with LSTM gave better results as compared to other classifiers. In a similar study, Ref. [50] compared the efficiency of CNN, ANN, and RNN with LSTM for crop yield prediction. RNN with LSTM gave an accuracy of 89% which outperformed other techniques used. The choice of the proposed CNN-RNN with the LSTM model emerged from the fact that two sets of data are to be used namely; the climatic dataset and the cocoa yield dataset. Also from the review, RNN and CNN are the trending, prevailing and latest models in the field of deep learning which are used but not fully explored in the area of crop yield prediction.

3. Materials and methods

3.1. Proposed model

This study proposed an ensemble model of Convolutional Neural Network and Recurrent Neural Network (CNN-RNN) with LSTM for cocoa yield prediction. The research adopts VGG-like architecture for the design of the proposed model. The model has four segments. The first segment constitutes CNN and the climatic dataset. CNN was used to extract the features of the climatic dataset (Rainfall, maximum temperature, and minimum temperature). CNN was used in this segment because various studies have proven the effectiveness of CNN for capturing the linear and nonlinear dependencies in a climatic dataset [7,24,25].

In the second segment, RNN was used to extract the features of cocoa yield (years and yield outcome). RNN was used because other

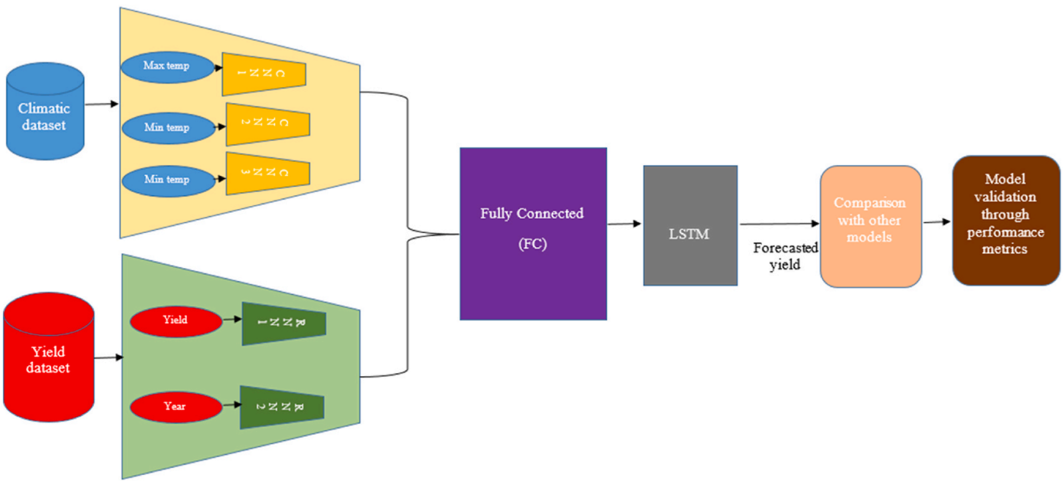

Fig. 2. Proposed model for cocoa yield prediction.

researchers have adopted it for yield prediction [8]. Also, it has been useful in a time series dataset. In the third segment, a fully connected layer was used to aggregate all the features of the datasets and then forwarded to the LSTM. Four layers of LSTM were used and for each layer of LSTM, 50 neurons were used and 0.2 dropout regularization was adopted to avoid overfitting of the model. To overcome gradient explosion during training, the learning rate was reduced and the number of batches was increased to make the model robust. A learning rate of 0.0005 and 20 batch size was used. Cocoa yields were predicted based on the input variables from the yield dataset and climatic datasets. To find the input variables, the guided backpropagation method was used to back-propagate the positive gradients which help to maximize the activation of the targeted neurons [40]. All the validation samples were firstly fed into the CNN-RNN with the LSTM model and the average activation of all the neurons was computed into the LSTM at time step t. The gradient of the activation neurons was set to 1 while the other neurons were set to 0. For us to find the important input variables based on the magnitude of the gradient, the gradients of the activated neurons were propagated into the input space. Fig. 2 below depicts the structure of the proposed model.

3.2. Dataset

This research work adopts two sets of data, namely; the climatic dataset and the yield dataset. The climatic dataset was collected from Nigeria's metrological agency (Nimet) covering the five major cocoa-producing states in Nigeria. The climatic dataset spans from 1988 to 2017 with minimum temperature, maximum temperature, and rainfall as its features. The climatic dataset used for this research covered only the five major cocoa-producing states in southwest Nigeria, the states include; Ondo state, Ekiti state, Ogun state, Ibadan, and Osun state. Before any dataset is fed into any machine learning, the dataset needs preprocessing, to put it in a format that the machine can use. Since the planting of cocoa in southwest Nigeria starts around April and the harvesting starts around October, the study focus on both the yield data and climatic data for these months between 1988 and 2017. The climatic dataset has 31,320 samples with three features namely; rainfall, minimum temperature, and maximum temperature. While the yield dataset was collected from the work of [51] and spans from 1988 to 2017, it has two features namely the "yield outcome" and "year" with a total of 58 samples. The total dataset samples used for training is 21,640 while 9738 samples were used for testing, both datasets are examples of time series datasets. Both data were properly scrutinized to ensure that there are no missing values. The dataset was normalized using the popular normalization method known as min-max normalization. This approach was used to bring the dataset into a close range and also to perform the linear transformation of the dataset. The same amount of aforementioned training and the testing dataset was used after data preprocessing because there were no missing values and non-numeric values in both dataset. Based on the minimization of loss, the optimal iteration was selected from the series of experiments conducted. Equation (1) is the formula for normalization used.

$$\theta_n \frac{\varepsilon_n - Min(\varepsilon)}{Max(\varepsilon) - Min(\varepsilon)} * 100 \tag{1}$$

Where:

θ_n = In the dataset, the nth normalized value
ε_n = the nth value in the dataset
Min (ε): Minimum value in the dataset
Max (ε): Maximum value in the dataset

3.3. Operation of CNN from the proposed model

Another aspect of the emerging researchable area is the aspect of time series data. CNN is a type of feed-forward neural network

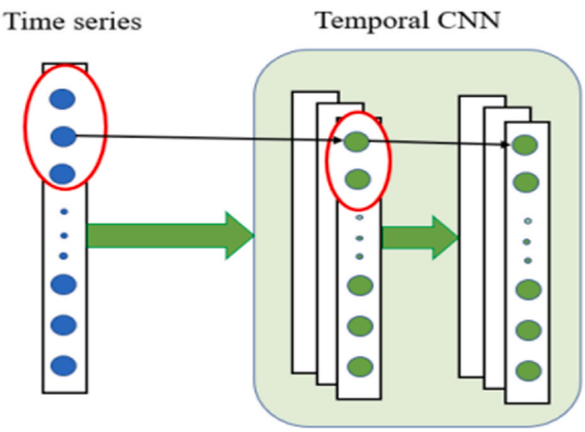

Fig. 3. CNN training climatic dataset.

that uses a convolutional structure to extract features from a time series dataset. Multilayer neural network works in such a way that each neuron is connected to the other, an input neuron is connected to another output neuron. For instance, given "a" as inputs and "b" outputs, an element of a*b will be generated as the weight matrix, but for CNN, a convolutional kernel of K*K will be used to reduce the weight of a*b to K*b, as a result of this reduction, only a weight matrix of K*b will be learned during the process of training the layers. There is an improved parameter optimization which leads to a hidden layer of the deep neural network to be trained at the same computational complexity. Useful information is thus extracted and trained by the CNN thus, improving the output of the CNN. A temporal CNN (Fig. 3) was used in this research to extract useful information from the time series dataset. The temporal CNN adopts K*1 as its kernel size against the traditional CNN that uses K*K. One important parameter in temporal CNN is the kernel size because it determines the length of the time series for each read and the information contained in the time series dataset was extracted by the convolution operation. The kernel matrix for 2-D convolutions is a two-dimensional matrix while the kernel matrix for temporal convolution is a one-dimension matrix. Equation (2) below depicts the input of the time series data into the temporal CNN and Equation (3) depicts the kernel function which is a vector of the parameters. During training, the vectors are constantly updated. Equation (4) and Equation (5) depicts the convolutional mapping between the input and kernel when the step size is assume as β

$$w(t) : \{1, \ldots\ldots\ldots, £\} \rightarrow R \tag{2}$$

$$f(t) : \{1, \ldots\ldots.., k) \rightarrow R \tag{3}$$

Assumes that the step size is β, the convolutional mapping between the input and kernel will be

$$q(t) : \left\{1, \ldots\ldots.. \frac{(\Theta - k)}{\beta + 1}\right\} \rightarrow R \tag{4}$$

Which can be expressed as

$$\mathrm{q(y)} = \sum_{k=\Theta}^{k} \mathrm{f(t).w(y.\beta - t + k - \beta + 1)} \tag{5}$$

The climatic data are preprocessed by the temporal CNN with the one-dimensional operation and are fed into the LSTM. Table 1 depicts the predictor variables selected from the climatic dataset. At this point, the dataset is suitable for LSTM for prediction.

3.4. Operation of RNN from the proposed model

RNN was used to preprocess the cocoa yield dataset. Table 1 above depicts the predictor variables used for cocoa yield. RNN is one of the variants of ANN. Its operations are similar to ANN with some differences. Firstly, RNN has three layers which are the input layer, hidden layer, and output layer. The input layer takes the input variables from the dataset, and the hidden layer receives the input and then sends it to the output layer. One of the differences between ANN and RNN is that output for the next layer is produced from the values of the previous layers of the hidden layer. RNNs are good for sequence modeling, this concept makes RNN more efficient because the sequence data used will be more understood. Fig. 4 below depicts the working structure of RNN, where X_0 servers as input into the RNN and Y_0 is the corresponding output to X_0. Y_0 Combines with X_1 to serve as RNN and Y_1 will be the output for the input from X_1 and Y_0. Y_1 combines with X_2 to serve as input into RNN and a corresponding Y_2 is produced. This process continues for X_3 … … … X_t during the training process and it helps RNN to memorize the content. Despite the efficiency of RNN, there were two major

Table 1
Predictor variables used for few years from the dataset(s).

Year	Cocoa yield (kgha^{-1})	April Avg. Max. Tem (°C)	Aug Avg. Max. Tem (°C)	Oct Avg. Max. Tem (°C)	April Avg. Min. Tem (°C)	Aug Avg. Min. Tem (°C)	Oct Avg. Min. Tem (°C)	April Avg. Rainfall (mm)	Aug Avg. Rainfall (mm)	Oct Avg. Rainfall (mm)
1988	230	31.18	26.36	29.28	25.3	23.0	24.0	173.5	152.3	201.8
1989	256	31	26.64	28.54	25.1	23.1	23.7	122.0	220.6	164.2
1990	244	30.93	26.54	28.7	25.2	23.0	23.8	184.8	168.3	167.2
1991	268	30.34	26.23	28.02	24.2	22.8	23.1	175.2	223.3	190.6
1992	292	31.41	25.89	28.88	25.0	22.7	23.6	130.9	165.7	159.5
1993	306	31.54	27.16	29.33	25.2	23.2	23.9	62.5	205.7	139.9
1994	203	31.02	26.64	28.70	25.2	23.1	23.7	96.9	147.3	192.1
1995	323	30.95	27.19	28.72	25.3	23.6	23.9	135.8	269.9	187.3
1996	325	30.81	26.63	28.61	24.7	23.1	23.5	147.8	222.7	176.3
1997	318	30.05	26.90	29.26	24.2	22.8	23.4	211.5	154.9	210.4
1998	370	33	26.61	28.82	26.4	22.9	24.0	91.2	87.2	178.9
1999	225	30.67	26.88	28.14	24.4	23.2	23.2	100.8	237.6	258.8
2000	338	30.90	26.50	28.70	25.0	23.0	24.0	96.1	244.8	146.7
2001	340	30.76	25.86	29.41	25.1	23.0	24.3	169.9	98.1	73.1
2002	362	30.67	26.57	28.54	25.0	23.3	23.7	104.4	229.2	212.6
2003	385	30.74	26.87	29.65	25.1	23.5	24.4	164.1	125.4	170.7

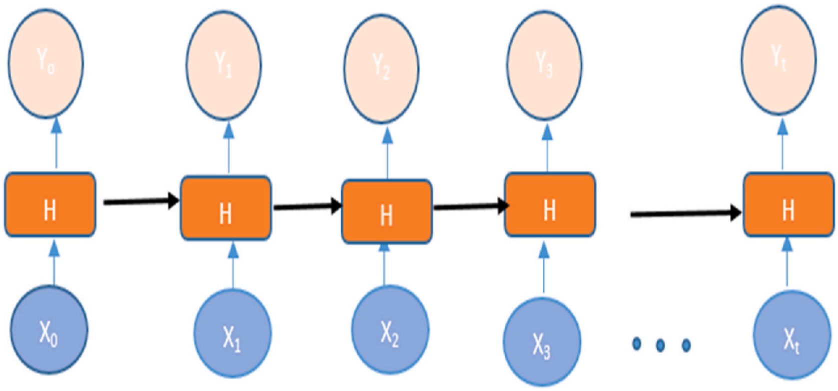

Fig. 4. Structure of RNN

setbacks suffered by RNN. Firstly, they suffered from the problem of exploding gradients where weights are unnecessarily allotted to the input by the algorithm. Secondly, they suffered from a vanishing gradient problem where the algorithm stops learning because the gradient is too small which is a result of repeated iteration. LSTM was used in this research work to solve the problems observed with the RNN. LSTM is of the form of RNN designed to handle vanishing gradient occurrence. Fig. 4 below depicts the structure of RNN.

3.5. Long Short Term Memory (LSTM)

LSTM has a long short-term memory cell which enables the network to bring to remembrance the output for any time frame. Basically, LSTM has 3 gates, namely; input gates (it), forget gates (ft), and output gates (ot) [52] Equations (6)–(10) below depicts the mathematical representation of the three gates respectively. From Equation (6) to 10 the following are the parameters used; tanh () was used as the hyperbolic tangent function, ʘ as the element-wise multiplication, ɵ as a logistic sigmoid function. bi, bf, and bo are the biases to be learned during training while wi, wf, and wo are used as the weight of the neurons. The input gate determines whether or not to add a new input into the network. The forget gate confirms from the output gate whether another output can be used to send to the output gate, it also determines which current information in the forget gate is to be deleted, which determines whether the information is useful or not useful. Any information that is not useful in the forget gate is deleted and new information is received from the input gate and others are passed to the output gate on request. All the information received from the memory is stored in the memory cell.

$$i_t = \theta(w_i[h_{t-1}, x_t] + b_i) \tag{6}$$

$$f_t = \theta(\ w_f[h_{t-1}, x_t] + b_f \tag{7}$$

$$o_t = \theta(w_o[h_{t-1}, x_t] + b_o) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\ (w_c[h_{t-1}, w_t] + b_c \tag{9}$$

$$h_t = o_t \odot \tanh(c_t) \tag{10}$$

Stack LSTM were used to receive input from CNN and RNN respectively. Where one takes input from the RNN and the other from the CNN as depicted in Fig. 2 above. The climatic output preprocessed data from CNN serves as input into one of the LSTM while the preprocessed yield output data serves as input into the other LSTM. The LSTM does the actual prediction and the results were compared to other algorithms. The final hidden stage for the LSTM is depicted in Equation (11). ε denotes the average of the hidden stages in Equation 11

$$h_t = \varepsilon\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) \tag{11}$$

3.6. Machine learning used for comparison

i Multilayer Neural Network (MLN)

MLN is another machine learning algorithm that was used to predict cocoa and the results were compared with the proposed CNN-RNN with the LSTM model. MLN has three layers which include; the input layer, hidden layer, and output layer. All these neuron layers are interconnected with each other. The input layer takes in the dataset and sends it to the hidden layer and the hidden layer sends it to the output layer which finally does the prediction as depicted in Fig. 5 below. MLN used has five input variables which were the features of the datasets used and a Tanh(x) was used as the activation function with 18 neurons on the hidden layer, the output layer of the MLN does the prediction.

ii Naïve Bayes

BN is an example of a machine-learning algorithm that was used to predict cocoa yield. The research work adopts Naïve Bayes which is constrained form of BN. As shown in Fig. 6 below, there exists a relationship between the dataset and its variables. The relationship is plotted and connected by arcs in a graph. From the graph, we have the independent and dependent variables, where the independent is the term parent node and the dependent is the term child node. The probability density of all child nodes is solely dependent on the parent node. The algorithm uses probability distribution which is derived from Bayes theory as stated below. As indicated in Fig. 6 below, the climatic dataset and yield dataset are the parent nodes. The min temp, max temp, and rainfall are all the features of the climatic node. Year and yield outcome are also features of the yield dataset. The yield predictor node is dependent on all the features to determine its final output. Equation (12)–(16) depicts how Bayes' theorem was apply to the dataset.

$$P(a|b) = \frac{P(b|a) * P(a)}{P(b)} \tag{12}$$

a represents the input variables from the dataset while b represents the output from the model. Where;

$$b = (b_1 b_2 b_3 \dots\dots\dots b_n)$$

Putting naïve assumption to Bayes' theorem, which is independent among the features, we then have;

$$P(a|b_1 b_2 \dots\dots b_n) = \frac{P(b_1|a)P(b_2|a)\dots\dots\dots P(b_n|a)P(a)}{P(b_1)P(b_2)\dots\dots\dots P(b_n)} \tag{13}$$

Which can be represented as

$$P(a|b_1 b_2 \dots\dots b_n) = \frac{P(a)\prod_{i=1}^{n} P(b_i|a)}{P(b_1)P(b_2)\dots\dots\dots P(b_n)} \tag{14}$$

We can remove the denominator, since it remains constant for a given input. We then have;

$$P(a|b_1 b_2 \dots\dots b_n) \alpha\ P(a)\prod_{i=1}^{n} P(b_i|a) \tag{15}$$

For the set of all the possible inputs, we picked the maximum probability when the set of inputs for all the possible values of the class variable y are picked. This can be represented mathematically as;

$$a = argmax_a\ P(a)\prod_{i=1}^{n} P(b_i|a) \tag{16}$$

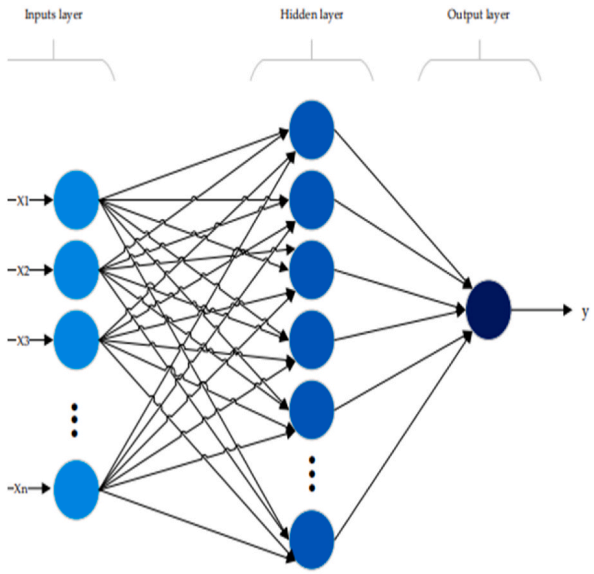

Fig. 5. Structure of Multi-Layer Perceptron neural network.

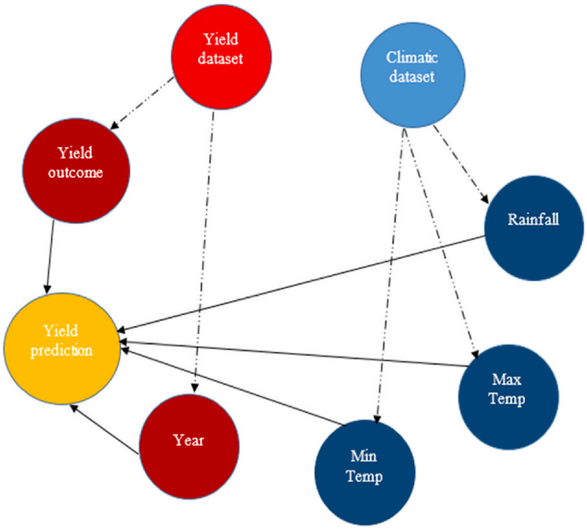

Fig. 6. Structure of naïve Bayes.

iii Seasonal Auto-Regression Moving Average (SARIMA)

SARIMA evolved from ARIMA and ARIMA evolved from ARMA. ARMA is a combination of AR (Auto-Regression) and MA (Moving Average). Given a discrete time series Z_t that has different variables, the general Auto regression for Z_t is represented in Equation (17) below

$$\text{AR (1)};\ Z_t = \lambda_0 + \lambda_1 Z_{t-1} + \lambda_2 Z_{t-2} + \ldots + \lambda_p Z_{t-p} + \varepsilon_t \tag{17}$$

Where Z_t represents the variable at $t, Z_{t-1}, Z_{t-2}, Z_{t-p}$,
Where λ_0 represents the coefficient
And ε_t represents the error factor.

Also, the generalization of the Moving Average (MA) is represented in Equation (18) as depicted below.

$$\text{MA (m)};\ Z_t = \theta_0 + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_p Z_{t-p} + \varepsilon_t \tag{18}$$

If we replace.

Z_t with $\Delta Z_t = (Z_t - Z_{t-1})$, ARMA then becomes ARIMA (l, m, n).
l is the weighted moving average from past observations. It also indicates the order of Auto regression.
m indicates the order of differencing (Integration)
n indicates the order of moving average

Equation (19) below depicts the general form of the ARIMA model

$$Z_t = \lambda_0 + \lambda_1 Z_{t-1} + \lambda_2 Z_{t-2} + \ldots + \lambda_p Z_{t-p} + \varepsilon_t + \theta_0 + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_p Z_{t-p} \tag{19}$$

Seasonal ARIMA (SARIMA) model was used because there is variation in the time series dataset.

SARIMA was also used to predict cocoa yield and compared with the proposed algorithm because SARIMA does well when it comes to the time series dataset. SARIMA has two major parts which are the seasonal part (L, M, N) and the non-seasonal (l, m, n) [53]. Where:

l: order of non-seasonal AR terms
m: order of non-seasonal differencing
n: order of non-seasonal MA terms
L: order of seasonal AR
M: order of seasonal differencing
N: order of seasonal MA terms
l = m = n = range (0, 2)
l, m, n were iterated as lmn = list (itertools. product (lmn))

SARIMA (l, m, n, L, M, N)s has the form where;

Order = (l, m, n) = (1, 1, 0)
Seasonal order = (L, M, N)s = (1, 1,1, 12)

Where s is the seasonality = 12.

The SARIMA model (1, 1, 0), (1, 1, 0, 12) were selected because it has the lowest aic.
SARIMA used was supported in python via Statsmodels library.
Model = sm. tsa.statespace.

SARIMAX (y_train, order = (1, 1, 0), seasonal_order = (1, 1, 0, 12), enforce stationary = false, enforce_invertibility = false) were used to fit on the training dataset using the model. fit () function. The model was trained with 70% of our dataset. Once that is achieved, we then use the SARIMA model to make prediction by calling the predict () function. For evaluation metric used for SARIMA, Equation (20) shows that SARIMA (l, m, n, L, M, N) $_s$ has the form

$$(\lambda^s)\varphi_l(\lambda)(1-\lambda^s)^M(1-\lambda)^n L_t = \theta_n(\lambda^s)\theta_n(\lambda)N_t \tag{20}$$

3.7. Evaluation metrics

The proposed model, machine learning algorithms, and deep learning were evaluated based on the following metrics; Mean Absolute Error (MAE), Means Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The aforementioned parameters were selected because there are the most suitable parameters when it comes to evaluating the error rate and performance of a model (Soudeh et al., 2021).

3.7.1. Mean absolute error (MAE)

It is computed by finding the average of the absolute error. It is being calculated with the formula depicted in Equation (21) where $e_t = y_t - f_t$ which is the error forecast, the size of the test set is denoted by n, the actual value is denoted by y_t and f_t is the forecasted value.

$$MAE = \sum_{t=1}^{n}\frac{e_t}{n} \tag{21}$$

3.7.2. Mean square error (MSE)

MSE estimates the total deviation between the predicted values and the actual values. It is being calculated with the formula from Equation (22) where $e_t = y_t - f_t$ which is the error forecast, the size of the test set is denoted by n, the actual value is denoted by y_t and f_t is the forecasted value.

$$MSE = \sum_{t=1}^{n}\frac{e_t^2}{n} \tag{22}$$

3.7.3. Root Means Square Error (RMSE)

This is calculated by finding the rooted average of mean square error and the average of the square error. Where $e_t = y_t - f_t$ which is the error forecast, the size of the test set is denoted by n, the actual value is denoted by y_t and f_t is the forecasted value. Equation (23) depicts the formula for RMSE

$$RMSE = \sqrt{\frac{\left(\sum_{t=1}^{n} e_t^2\right)}{n}} \tag{23}$$

3.7.4. Mean Absolute Percentage Error (MAPE)

This is used to measure the accuracy of a forecast system in percentage. Whenever the forecast error percentage is closer to or above 100%, such a model is regarded as inaccurate but if it tends toward 0%, then the such model is okay. Where $e_t = y_t - f_t$ which is the error forecast, the size of the test set is denoted by n, the actual value denoted by y_t and f_t is the forecasted value. Equation (24) depicts the formula for MAPE

$$MAPE = \sum_{t=1}^{n}\frac{e_t}{y_t} * 100 \tag{24}$$

3.7.4.1. Experimental setup. For our experimental setup, the Anaconda platform was used in a Python environment. The dataset was partitioned into training and testing datasets. 70% of the dataset was used for training while the remaining 30% was used for testing, this ratio was concluded after a series of multiple experiments were carried out to validate the efficiency of all the models used. The proposed model was compared with other machine learning models and other deep learning models. Getting good experimental results entails setting up a suitable optimizer to solve the problem of parameter optimization. In the deep learning, library used, different optimization algorithms can be adopted such as Root Mean Square prop (RMSprop) and stochastic Gradient Descent (SGD). One of the ways to overcome the problem of overfitting is to choose the right optimizer. A comparative experiment was carried out comparing the mean square error (MSE) loss of different optimizers when training the CNN-RNN with the LSTM model. From the comparative experiment, the Adam optimizer gave the best result which was finally used for the model training. For the activation function, Rectified Linear unit (ReLu) was selected. Also, an epoch of 200 was used in a batch size of 20 this is because there were minimum losses. Since the training dataset is 20 years, a timestep of 20 was used indicating that the model can learn from the back 20 years and predict for the subsequent years.

3.7.4.2. Validation mechanism. The historical dataset from 1988 to 2008 was used to train all the models used in this research work. The dataset from 2009 to 2017 was reserved as the test dataset and the data from the years before each of them formed the training dataset. It is to be noted that the training dataset was also partitioned into a 30%–70% ratio. Where 30% was excluded strictly for validation while the rest was used for training. The dataset reserved for validation was not used for model training. To ensure that the validation data has similar distribution with the training datasets, a stratified splits cross-validation approach was adopted where all the observations in the training data were binned into 5 linearly spaced based on the corresponding yield value. The training data were resampled 10 times (with replacement), enabling us to train the model 10 times. As a result of this, the models were trained 10 different times on the newly created training dataset.

4. Results from CNN-RNN with LSTM

The proposed model was trained with the yield and climatic dataset respectively, Based on the correlation captured and the training of the proposed model, the prediction of cocoa yield for the next nine years was done. The model uses a time step which helps the model to learn and based on what it has learned, the output is generated. See Table 18 for hyper parameters setting.

The model uses a time step of 20 years to learn from the given dataset and predict the next subsequent years. See other parameters setting for the proposed model in section 3.1 Table 2 and Fig. 7 below depict the actual and predicted cocoa yield from 2009 to 2017 while Table 3 reveals the evaluation metrics of CNN-RNN with LSTM.

4.1. Result from multilayer perceptron (MLP)

Table 4 and Fig. 8 depict the actual and predicted cocoa yield of MLP from 2009 to 2017. Different numbers of neurons were used in the hidden layer during model training ranging from 10 to 18 with an interval of 4, this was done to select the best architecture for predicting the cocoa yield. Tanh(x) was selected as the activation function compared to other activation functions because it gave a higher performance. 70% of the dataset was used to train the MLP which was sorted out randomly by the model. See Table 18 for hyper parameters setting. The model training was repeated three different times and was tested with a different number of neurons. The error rate of the model was controlled as the neuron were changed, MLP with 18 neurons gave the best result as compared with MLP with 10 and 14 neurons respectively. After the prediction of cocoa yield, MLP was evaluated and the results are represented in Table 5.

4.2. Naïve Bayes

Table 6 and Fig. 9 depict the actual and predicted cocoa yield of NB from 2009 to 2017. The climatic dataset node and the yield dataset node are the parent nodes respectively. The evaluation results for NB are depicted in Table 7.

4.3. Results for Seasonal Auto-regression Integrated Moving Average (SARIMA)

Table 8 and Fig. 10 depict the actual and predicted cocoa yield of SARIMA from 2009 to 2017. SARIMA, a linear model was trained with the yield and climatic dataset from 1988 to 2008. The climatic features (Rainfall, Minimum, and Maximum Temperature) from

Table 2
Actual and predicted cocoa yield of CNN-RNN with LSTM from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield kg ha^{-1})
2009	513	500.5206
2010	525	512.4174
2011	370	344.7426
2012	371	345.3445
2013	238	225.6457
2014	248	210.7325
2015	200	164.3216
2016	230	200.6718
2017	240	195.4672

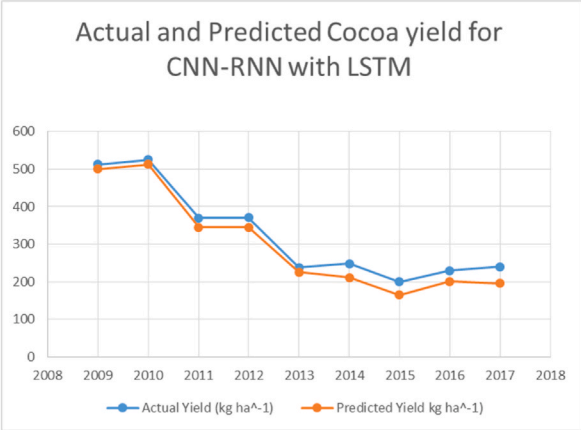

Fig. 7. Graphical Representation of Actual and predicted cocoa yield of CNN-RNN with LSTM from 2009 to 2017.

Table 3
Evaluation metrics of CNN-RNN with LSTM.

Technique	MAE	MSE	RMSE	MAPE
CNN-RNN with LSTM	26.1111	808	28.4175	9.7705

Table 4
Actual and predicted cocoa yield of MLP from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield ((kg ha^{-1})
2009	513	452.7568
2010	525	465.9856
2011	370	289.0765
2012	371	295.6732
2013	238	201.4521
2014	248	175.0438
2015	200	145.5431
2016	230	196.769
2017	240	130.457

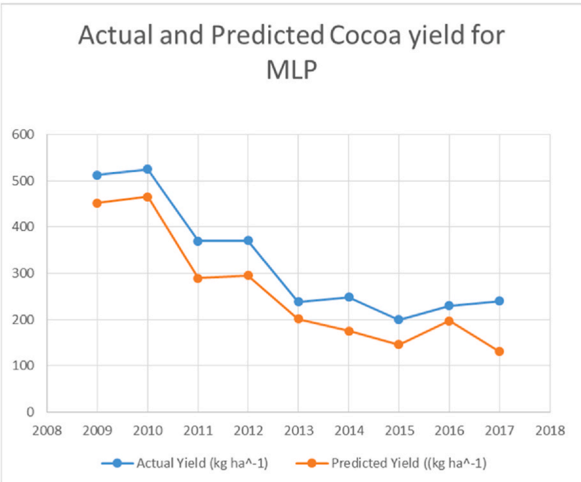

Fig. 8. Graphical Representation of Actual and predicted cocoa yield of MLP from 2009 to 2017.

Table 5
Evaluation metrics of MLP.

Technique	MAE	MSE	RMSE	MAPE
MLP	64.6666	37,667	194.0809	21.9169

Table 6
Actual and predicted cocoa yield of NB from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield ((kg ha^{-1})
2009	513	450.6873
2010	525	462.1327
2011	370	280.8641
2012	371	295.7722
2013	238	200.3883
2014	248	173.9046
2015	200	144.1276
2016	230	190.9836
2017	240	131.782

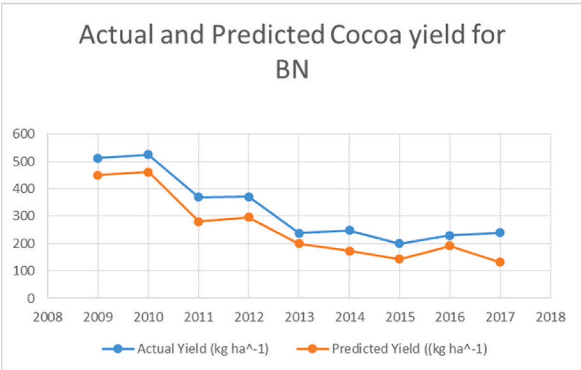

Fig. 9. Graphical Representation of Actual and predicted cocoa yield of NB from 2009 to 2017.

Table 7
Evaluation metrics of NB.

Technique	MAE	MSE	RMSE	MAPE
NB	67.1111	40,583	201.4526	22.6845

Table 8
Actual and predicted cocoa yield of SARIMA from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield (kg ha^{-1})
2009	513	425.0192
2010	525	461.5409
2011	370	279.4536
2012	371	201.6902
2013	238	171.3219
2014	248	141.9021
2015	200	134.4285
2016	230	192.852
2017	240	125.9931

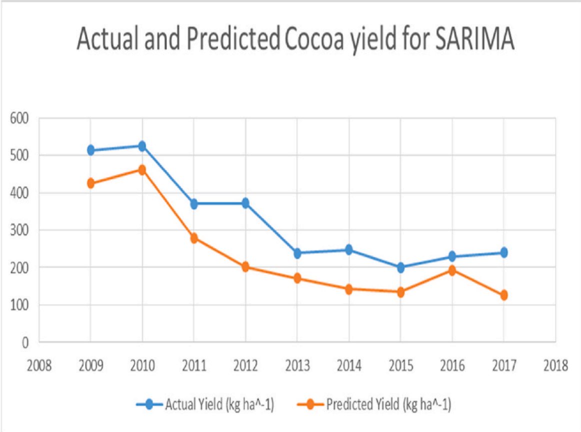

Fig. 10. Graphical Representation of Actual and predicted cocoa yield of SARIMA from 2009 to 2017.

April to October were considered as the independent variable, and the cocoa yield was taken as the independent variable. After the prediction of cocoa yield by SARIMA, it was evaluated and the results are represented in Table 9.

4.4. CNN-LSTM

Table 10 and Fig. 11 depict the actual and predicted cocoa yield of CNN-LSTM from 2009 to 2017. CNN-LSTM model has been proposed by Sharma et al. (2020) for yield prediction before, the model was compared with our proposed model using climatic and yield datasets for model training and validation. CNN-LSTM was trained separately to compare the results to the proposed model. As used, CNN was used to process our data and the result was forwarded to the LSTM. The output was followed by a fully connected layer. CNN-LSTM model has four LSTMs and for each LSTM, 50 neurons were used. To avoid the problem of overfitting, 0.2 dropout regularization was applied. Adam optimizer was used along with 200 epochs in a batch of 20. Also, the ReLu activation function was applied to the model learning at a rate of 0.0005. The kernel size used was 3 by 3 with a stride number of one. Table 11 depicts the evaluation metrics for the CNN-LSTM model. See Table 18 for hyper parameters setting.

4.5. RNN-LSTM

Table 12 and Fig. 12 depict the actual and predicted cocoa yield by RNN-LSTM from 2009 to 2017. The RNN-LSTM model was used for wheat yield prediction by Nishu and Anshu, (2021). Our proposed model was compared with the RNN-LSTM model. This model has two parts, RNN which were used to process the dataset and then passed to the LSTM, the model uses a fully connected layer. The model also has four LSTM and 50 neurons used for each. The following parameters were used 0.2 dropout regularization, an epoch of 200 in a batch size of 20, Adam optimizer, and ReLu activation function. Since we are training the model with 20 years dataset, a timestep of 20 was used and the model learning at a rate of 0.0005. See Table 18 for hyper parameters setting. Table 13 depicts the evaluation metrics for the RNN-LSTM model.

4.6. LSTM

The proposed model was also compared with the LSTM model. LSTM has been used by Jiang et al. (2018) for corn yield prediction. Table 14 and Fig. 13 depicts the actual and predicted cocoa yield by LSTM from 2009 to 2017. In the quest to get the best model for predicting cocoa yield in southwest Nigeria. LSTM was trained to determine the best model. Five LSTM layers were used, and two layers of LSTM were used to extract the features from the datasets as the input layer, the dense layer, and the output layer. Each layer of LSTM used has 50 nodes and a dropout of 0.2 was used with a timestep of 20. Adam optimizer was used and the model learning was at a rate of 0.0005. Table 15 depicts the evaluation metrics for the RNN-LSTM model. See Table 18 for hyper parameters setting.

5. Discussion

After the experiment was conducted, the following inference was concluded.

i. SARIMA which is a linear model gave the highest mean absolute error which suggests that the other two models (MLP and NB) gave the least absolute error as compared with SARIMA only.
ii. Secondly, it was observed that among the other three machine learning models used, the means absolute error interval between the three models is minimal as compared to the proposed model. This suggests that the proposed model performed very well considering the error interval between the proposed model and the other used machine learning models.

Table 9
Evaluation metrics of SARIMA.

Technique	MAE	MSE	RMSE	MAPE
SARIMA	89	71,253	266.9328	29.6203

Table 10
Actual and predicted cocoa yield of CNN-LSTM from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield (kg ha^{-1})
2009	513	487.4169
2010	525	486.0142
2011	370	300.7284
2012	371	304.2168
2013	238	220.3468
2014	248	204.6748
2015	200	177.4432
2016	230	205.0342
2017	240	200.9346

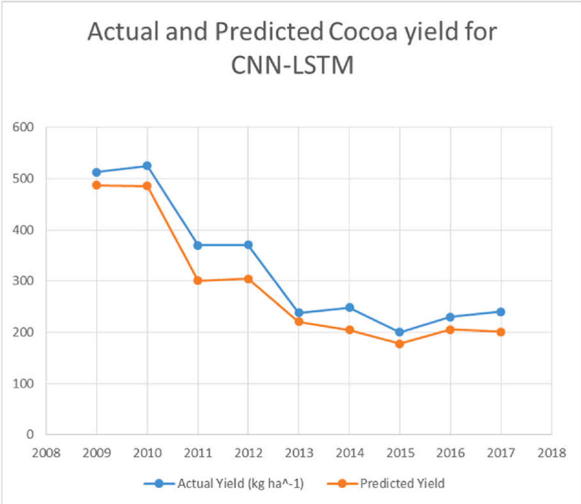

Fig. 11. Graphical Representation of Actual and predicted cocoa yield of CNN-LSTM from 2009 to 2017.

Table 11
Evaluation metrics of CNN-LSTM.

Technique	MAE	MSE	RMSE	MAPE
CNN with LSTM	38.6666	13,471	116.0634	12.4926

Table 12
Actual and predicted cocoa yield of RNN-LSTM from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield (kg ha^{-1})
2009	513	498.3406
2010	525	484.8946
2011	370	350.4714
2012	371	343.001
2013	238	205.7946
2014	248	204.1423
2015	200	175.0014
2016	230	198.4271
2017	240	192.4785

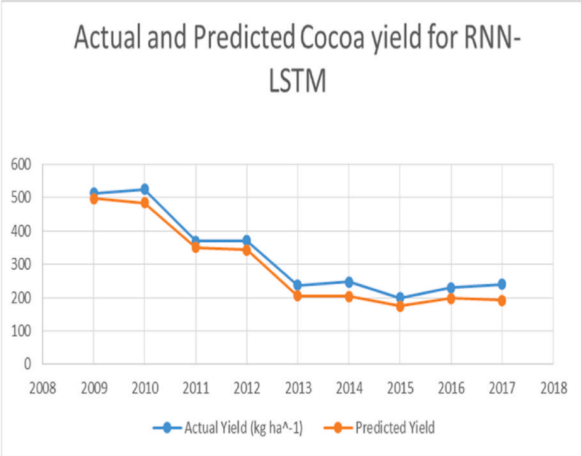

Fig. 12. Graphical Representation of Actual and predicted cocoa yield of CNN-LSTM from 2009 to 2017.

Table 13
Evaluation metrics of RNN-LSTM.

Technique	MAE	MSE	RMSE	MAPE
RNN-LSTM	31.3333	8864	94.1495	11.1739

Table 14
Actual and predicted cocoa yield of LSTM from 2009 to 2017.

Years	Actual Yield (kg ha^{-1})	Predicted Yield (kg ha^{-1})
2009	513	469.8426
2010	525	476.321
2011	370	294.3214
2012	371	299.9497
2013	238	217.8263
2014	248	195.0426
2015	200	165.0427
2016	230	200.9216
2017	240	186.3296

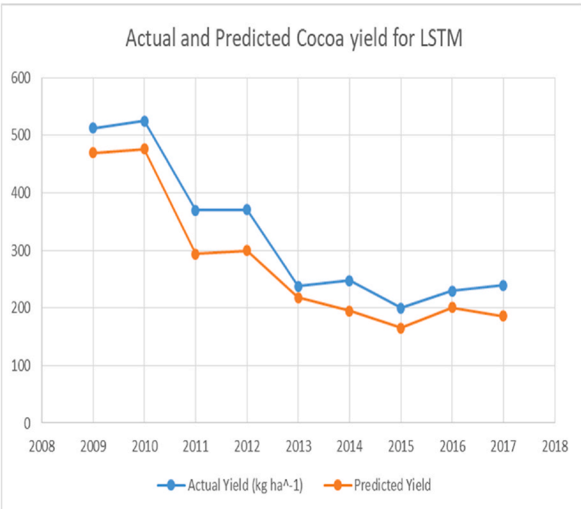

Fig. 13. Graphical Representation of Actual and predicted cocoa yield of LSTM from 2009 to 2017.

Table 15
Evaluation metrics of LSTM.

Technique	MAE	MSE	RMSE	MAPE
LSTM	47.6666	20,487	143.1341	15.5115

Table 16
Performance Comparison of the proposed technique and other machine learning techniques.

Techniques	MAE	MSE	RMSE	MAPE
CNN-RNN with LSTM	26.1111	808	28.4175	9.7705
NB	67.1111	40,583	201.4526	22.6845
SARIMA	89	71,253	266.9328	29.6203
MLP	64.6666	37,667	194.0809	21.9169

Table 17
Performance Comparison of the proposed techniques and other deep learning techniques.

Techniques	MAE	MSE	RMSE	MAPE
CNN-RNN with LSTM	26.1111	808	28.4175	9.7705
LSTM	47.6666	20,487	143.1341	15.5115
CNNwith LSTM	38.6666	13,471	116.0634	12.4926
RNN-LSTM	31.3333	8864	94.1495	11.1739

Table 18
Hyper parameters settings of models used.

Model	Optimizer	Epoch	Dropout Regularization	Learning rate	Activation function	Evaluation Metric	Batch Size	Number of hidden layers	Number of neurons	Time step
CNN-RNN with LSTM	Adam	200	0.2	0.0005	Relu	MAE, MSE, RMSE,MAPE	20	6	64	20
LSTM	Adam	200	0.2	0.0005	Relu	MAE, MSE, RMSE,MAPE	20	4	50	20
RNN-LSTM	Adam	200	0.2	0.0005	Relu	MAE, MSE, RMSE,MAPE	20	4	50	20
CNN-LSTM	Adam	200	0.2	0.0005	Relu	MAE, MSE, RMSE,MAPE	20	4	50	20
NB						MAE, MSE, RMSE,MAPE				
MLP	Adam	300			Tanh(x)	MAE, MSE, RMSE,MAPE		2	18	
SARIMA						MAE, MSE, RMSE,MAPE				

iii. Models with LSTM performed significantly better than MLP, NB, and SARIMA. This indicates that LSTM has a significant advantage over other conventional methods especially when it comes to time series datasets.
iv. Comparing the LSTM with CNN-LSTM and RNN-LSTM, CNN-LSTM and RNN-LSTM gave a better accuracy than the LSTM model. The reason for the differences in accuracy obtained may be the ability of CNN-LSTM and RNN-LSTM to extract complex features in the time series dataset. Thus, improving the accuracy over LSTM. Table 16 depicts the comparison between the proposed model and other machine learning models while Table 17 depicts the comparison between the proposed model and other deep learning models.
v. Due to the robustness of the proposed model, significant accuracy was obtained compared to other machine learning models and models with LSTM used in this research work. The proposed model was compared with RNN-LSTM (Nishu and Anshu, 2021), CNN-LSTM (Sharma et al., 2020), and LSTM (Jiang et al., 2018). Experimental results reveal that the proposed model gave better accuracy compared to other existing deep learning techniques. Fig. 14 depicts the performance comparison of all the models and the proposed model.

6. Conclusions

To accurately predict cocoa yield, the research proposed an ensemble approach consisting of a convolutional neural network, recurrent neural network, and Long Short Term Memory. The CNN was used to extract the climatic features while the RNN was used to

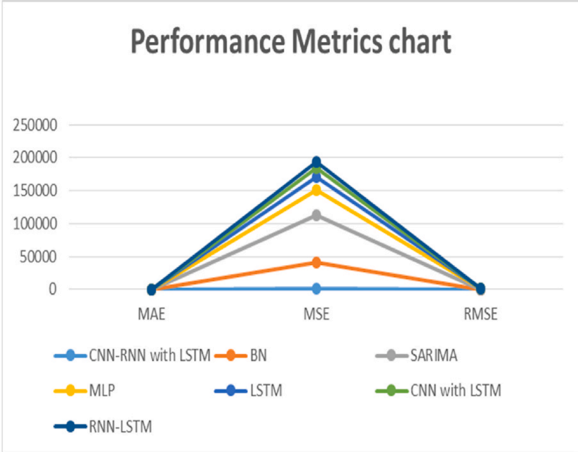

Fig. 14. Graphical Representation of the performance of the proposed techniques, other machine learning and deep learning.

extract yield features. The output of both models was concatenated by the fully connected layer and then send into the LSTM which finally does the prediction. The following conclusions were reached in the research work. (a) models with LSTM is more suitable for predicting yield when it comes to time series data compared to statistical model (b) CNN can be used to extract climatic features and RNN to extract yield features which can enhance the robustness and accuracy of the overall model. (c) Finally, the model is expected to provide a reasonable reference for yield prediction.

One of the major limitations of the research work is the accessibility to the country's climatic dataset. The price charge from the Nimet is too exorbitant, we were only able to pay for the climatic dataset of Southwest, Nigeria. In order to fully validate the effectiveness of our proposed model, more dataset from the other regions of Nigeria will have been a tremendous breakthrough for the research. Another limitation of the research is biasness of the models which is observed from the predicted and actual values. This is suspected to be from the dataset used because the dataset covered only the southwest Nigeria. Some of the features used from the climatic dataset e.g. temperature and rainfall vary to a large extend in southwest and other part of the country especially the northern part of the country. Further research can extend the scope of the dataset to curb the biasness and also effectively evaluate our model. Also, to improve the accuracy derived more advanced network architecture will be considered in the subsequent research.

Author contribution statement

Olofintuyi Sunday Samuel; Deji Olanike: Conceived and designed the experiments; Performed the experiments; Analysed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Olajubu Emmanuel Ajayi: Performed the experiments; Analysed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability statement

The data that has been used is confidential.

Declaration of interest's statement

The authors declare no conflict of interest

References

[1] Salej Sood, Vinay Bhardwaj, S.K. Kaushik, Sanjeev Sharma, Prediction based on estimated breeding values using genealogy for tuber yield and late blight resistance in auto-tetraploid potato (Solanum tuberosum L.), Heliyon 6 (Issue 11) (2020), e05624.
[2] A.C. Tyagi, Towards a second green revolution, Irrigat. Drain. 65 (4) (2016) 388–389, https://doi.org/10.1002/ird.2076.
[3] S.S. Olofintuyi, Cyber situation awareness perception model for computer network, Int. J. Adv. Comput. Sci. Appl. 12 (1) (2021) 392–397.
[4] Xingang Li, Hao Geng, Liqiang Zhang, Shuwen Peng, Xin Qi, Jianxi Huang, Xuecao Li, Suhong Liu, Yuebin Wang, Improving maize yield prediction at the county level from 2002 to 2015 in China using a novel deep learning approach, Comput. Electron. Agric. 202 (2022).
[5] Jieming Chou, Yuan Xu, Wenjie Dong, Xian Tian, Hong Xu, Zheng Wang, Comprehensive climate factor characteristics and quantitative analysis of their impacts on grain yields in China's grain-producing areas, Heliyon 5 (12) (2019), e02846.
[6] Girma Asefa Bogale, Mengistu Mengesha Maja, Gebre Hadgu Gebreyohannes, Modelling the impacts of climate change on faba bean (Vicia faba L.) production in Welmera area, central Ethiopia, Heliyon 7 (10) (2021), e08176.
[7] S. Khaki, L. Wang, S.V. Archontoulis, A CNN-RNN framework for crop yield prediction, Front. Plant Sci. 10 (2020) 1750, https://doi.org/10.3389/fpls.2019.01750.

[8] B. Nishu, S. Anshu, Deep learning based wheat crop yield prediction model in Punjab region of North India, Appl. Artif. Intell. 35 (15) (2021) 1304–1328, https://doi.org/10.1080/08839514.2021.1976091.
[9] Anjan Roy, Arun B. Inamdar, Multi-temporal Land Use Land Cover (LULC) change analysis of a dry semi-arid river basin in western India following a robust multi-sensor satellite image calibration strategy, Heliyon 5 (2019) (2019), e01478.
[10] Clement Nyamekye, Samuel Anim Ofosu, Richard Arthur, Osei Gabriel, Linda Boamah Appiah, Samuel Kwofie, Benjamin Ghansah, Dieter Bryniok, Evaluating the spatial and temporal variations of aquatic weeds (Biomass) on Lower Volta River using multi-sensor Landsat Images and machine learning, Heliyon 7 (5) (2021), e07080.
[11] Mumo Elijah Musyoki, John Ronoh Busienei, John Kamau Gathiaka, George Njomo Karuku, Linking farmers' risk attitudes, livelihood diversification and adoption of climate smart agriculture technologies in the Nyando basin, South-Western Kenya, Heliyon 8 (Issue 4) (2022), e09305.
[12] Yassine Bouslihim, Aicha Rochdi, Namira El Amrani Paaza, Machine learning approaches for the prediction of soil aggregate stability, Heliyon 7 (3) (2021), e06480.
[13] Rachit Srivastava, A.N. Tiwari, V.K. Giri, Solar radiation forecasting using MARS, CART, M5, and random forest model: a case study for India, Heliyon 5 (10) (2019), e02692.
[14] Johanna Karina Solano Meza, David Orjuela Yepes, Javier Rodrigo-Ilarri, Eduardo Cassiraga, Predictive analysis of urban waste generation for the city of Bogota, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks, Heliyon 5 (10) (2019), e02810.
[15] Ross Chapman, Cookc Simon, Christopher Donough, Ya Li Lim, Philip Vun Vui Ho, Koon Wai Lo, Thomas Oberthür, Using Bayesian networks to predict future yield functions with data from commercial oil palm plantations: a proof of concept analysi, Comput. Electron. Agric. 151 (2018) 338–348.
[16] S.S. Olofintuyi, Early cocoa black pod pathogen prediction with machine learning ensemble algorithm based on climatic parameters, J. Inf. Org. Sci. 46 (1) (2022) 1–12.
[17] Jose Manrique-Silupu, Jean C. Campos, Ernesto Paiva, William Ipanaqué, Thrips incidence prediction in organic banana crop with Machine learning, Heliyon 7 (12) (2021), e08575.
[18] Chee Cheong Lee, Voon Chet Koo, Tien Sze Lim, Yang Ping Lee, Haryati Abidin, A multi-layer perceptron-based approach for early detection of BSR disease in oil palm trees using hyperspectral images, Heliyon 8 (4) (2022), e09252.
[19] Chu Zheng, Jiong Yu, An end-to-end model for rice yield prediction using deep learning fusion, Comput. Electron. Agric. 174 (2020), 105471. ISSN 0168-1699.
[20] Z. Jiang, C. Liu, N.P. Hendricks, B. Ganapathysubramanian, D.J. Hayes, S. Sarkar, Predicting County Level Corn Yields Using Deep Long Short Term Memory Models, 2018 arXiv preprint arXiv:1805.12044.
[21] Vincent Roth, Tatenda Lemann, Gete Zeleke, Alemtsehay Teklay Subhatu, Tibebu Kassawmar Nigussie, Hans Hurni, Effects of climate change on water resources the upper Blue Nile Basin of Ethiopia, Heliyon 4 (2018), e00771.
[22] Ali Al-Naji, Ahmed Bashar Fakhri, Sadik Kamel Gharghan, Javaan Chahl, Soil color analysis based on a RGB camera and an artificial neural network towards smart irrigation: a pilot study, Heliyon 7 (1) (2021), e06078.
[23] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436, https://doi.org/10.1038/nature14539.
[24] M. Shahhosseini, G. Hu, S.V. Archontoulis, Forecasting corn yield with machine learning ensembles. [Methods], Front. Plant Sci. 11 (1120) (2020).
[25] M. Shahhosseini, G. Hu, I. Huber, S.V. Archontoulis, Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt, Sci. Rep. 11 (1) (2021) 1606, https://doi.org/10.1038/s41598-020-80820-1.
[26] Y. Cai, K. Moore, A. Pellegrini, A. Elhaddad, J. Lessel, C. Townsend, et al., Crop yield predictions-high resolution statistical model for intra-season forecasts applied to corn in the US, in: Paper Presented at the 2017 Fall Meeting, 2017.
[27] Q. Yang, L. Shi, J. Han, Y. Zha, P. Zhu, Deep convolutional neural networks for rice grain yield estimation at the ripening stage using uav-based remotely sensed images, Field Crop. Res. 235 (2019) 142–153, https://doi.org/10.1016/j.fcr.2019.02.022.
[28] J. You, X. Li, M. Low, D. Lobell, S. Ermon, Deep gaussian process for crop yield prediction based on remote sensing data, in: Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017, pp. 4559–4566.
[29] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, vol. 1, MIT Press, Cambridge, 2016.
[30] J.H. Jeong, J.P. Resop, N.D. Mueller, D.H. Fleisher, K. Yun, E.E. Butler, Random forests for global and regional crop yield predictions, PLoS One 11 (2016), e0156571, https://doi.org/10.1371/journal.pone.0156571.
[31] Edson Martinez, Elisângela Silva, Amaury Fabbro, A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil, Rev. Soc. Bras. Med. Trop. 44 (2011) 436–440, https://doi.org/10.1590/S0037-86822011000400007.
[32] A. González Sánchez, J. Frausto Solís, W. Ojeda Bustamante, Predictive ability of machine learning methods for massive crop yield prediction, Spanish J. Agric. Res. 12 (2) (2014) 313–328, https://doi.org/10.5424/sjar/2014122-4439, 2171-9292.
[33] S. Fukuda, W. Spreer, E. Yasunaga, K. Yuge, V. Sardsud, Müller, Random forests modelling for the estimation of mango (mangifera indica l. cv. chok anan) fruit yields under different irrigation regimes, Agric. Water Manag. 116 (2013) 142–150, https://doi.org/10.1016/j.agwat.2012.07.003.
[34] J.R. Romero, P.F. Roncallo, P.C. Akkiraju, I. Ponzoni, V.C. Echenique, J.A. Carballido, Using classification algorithms for predicting durum wheat yield in the province of buenosaires, Comput. Electron. Agric. 96 (2013) 173–179, https://doi.org/10.1016/j.compag.2013.05.006.
[35] A. Chlingaryan, S. Sukkarieh, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review, Comput. Electron. Agric. 151 (2018) 61–69, https://doi.org/10.1016/j.compag.2018.05.012.
[36] N. Gandhi, O. Petkar, L.J. Armstrong, Rice crop yield prediction using artificial neural networks, in: Technological Innovations in ICT for Agriculture and Rural Development (TIAR), IEEE, Chennai, India, 2016, pp. 105–110, https://doi.org/10.1109/TIAR.2016.7801222.
[37] K.A. Shastry, H.A. Sanjay, A. Deshmukh, A parameter based customized artificial neural network model for crop yield prediction, J. Artif. Intell. 9 (1–3) (2016) 23–32, https://doi.org/10.3923/jai.2016.23.32.
[38] X.E. Pantazi, D. Moshou, T. Alexandridis, R.L. Whetton, A.M. Mouazen, Wheat yield prediction using machine learning and advanced sensing techniques, Comput. Electron. Agric. 121 (2016) 57–65, https://doi.org/10.1016/j.compag.2015.11.018.
[39] P. Bose, N.K. Kasabov, L. Bruzzone, R.N. Hartono, Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series, IEEE Trans. Geosci. Rem. Sens. 54 (11) (2016) 6563–6573, https://doi.org/10.1109/TGRS.2016.2586602.
[40] S. Khaki, L. Wang, Crop yield prediction using deep neural networks, Front. Plant Sci. 10 (2019) 621, https://doi.org/10.3389/fpls.2019.00621.
[41] N. Kim, K.-J. Ha, N.-W. Park, J. Cho, S. Hong, Y.-W. Lee, A comparison between major artificial intelligence models for crop yield prediction: case study of the midwestern United States, 2006–2015, ISPRS Int. J. Geo-Inf. 8 (2019) 240, https://doi.org/10.3390/ijgi8050240.
[42] A.X. Wang, C. Tran, N. Desai, D. Lobell, S. Ermon, Deep transfer learning for crop yield prediction with remote sensing data, in: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, ACM, New York, United States, 2018, p. 50, https://doi.org/10.1145/3209811.3212707.
[43] S. Khaki, Z. Khalilzadeh, Classification of Crop Tolerance to Heat and Drought: A Deep Convolutional Neural Networks Approach, 2019 arXiv preprint arXiv:1906.00454.
[44] J. Fourie, J. Hsiao, A. Werner, Crop yield estimation using deep learning, in: 7th Asian-Australasian Conference on Precision Agriculture, NewZealand, 2017, https://doi.org/10.5281/zenodo.893710.
[45] S. Sharma, S. Rai, N.C. Krishnan, Wheat crop yield prediction using deep LSTM model, 2020 arXiv preprint arXiv:2011.01498.
[46] P. Sivanandhini, P.G. Scholar, J. Prakash, Crop yield prediction analysis using feed forward and recurrent neural network, Int. J. Innov. Sci. Res. Technol. 5 (5) (2020) 1092–1096.
[47] I. Oliveira, R.L. Cunha, B. Silva, M.A. Netto, A Scalable Machine Learning System for Pre-season Agriculture Yield Forecast, 2018 arXiv preprint arXiv:1806.09244.
[48] S.A. Haider, S.R. Naqvi, T. Akram, G.A. Umar, A. Shahzad, M.R. Sial, M. Kamran, M. Kamran, LSTM neural network based forecasting model for wheat production in Pakistan, Agronomy 9 (2) (2019) 72, https://doi.org/10.3390/agronomy9020072.

[49] A. Nigam, S. Garg, A. Agrawal, P. Agrawal, Crop yield prediction using machine learning algorithms, in: 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, Shimla, India, 2019, pp. 125–130, https://doi.org/10.1109/ICIIP47207.2019.8985951.
[50] M.K. Dharani, R. Thamilselvan, P. Natesan, P.C.D. Kalaivaani, S. Santhoshkumar, Review on crop prediction using deep learning techniques, J. Phys. Conf. 1767 (1) (2021), 012026 (IOP Publishing).
[51] O.S. Afolayan, Cocoa production pattern in Nigeria: the missing link in regional agro-economic Development.Analele Universităţii din Oradea, Seria Geografie 30 (1) (2020) 88–96.
[52] Alvaro David Orjuela, Andres Leonardo Jutinico, Mario Enrique Duarte, Carlos Enrique Awad García, Erika Vergara, María Angelica Palencia, Time series forecasting for tuberculosis incidence employing neural network models, Heliyon 8 (7) (2022), e09897.
[53] S.S. Olofintuyi, A predictive linear model for Outbreak of cocoa black pod disease in South West Nigeria, Achiev. J. Sci. Res. 3 (2) (2021) 19–27.