Journal of Environment and Earth Science
ISSN 2224-3216 (Paper) ISSN 2225-0948 (Online)
Vol.12, No.10, 2022

Skills Assessment of Selected Supervised Machine Learning Algorithms in Predicting Seasonal Rainfall over Bauchi in Nigeria

Adefisan Elijah Adesanya[1,2,*], Balogun Richard Ayodeji[1], and Audu Molema[1,3]
[1]Department of Meteorology and Climate Science, Federal University of Technology, Akure, Nigeria
[2]African Centre of Meteorological Applications for Development, Niamey, Niger.
[3]Central Forecasting Office, Nigerian Meteorological Agency, Abuja, Nigeria.

Abstract
An attempt is made to use four selected machine learning algorithms (MLAs) to predict the seasonal and monthly amount of rainfall over a Savana station in Nigeria. The four MLAs are the artificial neural network (ANN), Random Forest model (RFM), K-Nearest Neighbor (KNN), and kernel basis Support Vector Machine (SVM). Monthly mean rainfall and monthly mean air temperature data from June to October over a period of 34 years (1986-2019) were used and seventeen atmospheric variables are used to develop the model during training period. The period is divided into two, the training (1986 - 2013) and testing (2014 - 2019) periods. The results show that SVM and ANN better reproduce both monthly and annual rainfall amount over the study area by accessing their skills during training period and also having lowest RMSE and MAE during testing period. SVM is the most suitable among the four MLAs. Though, some show better results for specific month(s), the SVM and ANN summary yield 84% and 82% respectively of good forecasts for seasonal rainfall amount over Bauchi. The web interface was developed using R (ShinyR Package) programming has a very interactive and good graphical user interface (GUI) for user with little or no computer knowledge. It is recommended that the two MLAs can be used to predict monthly and seasonal rainfall over Savana climatic zone of West Africa using the seventeen input variables and hence other variables can be selected for forecasting other rainfall properties like onset, cessation and length of rainy season over West Africa sub-region. The results also show the importance and weight of each of the seventeen input variables has in reproducing the dependent variable and hence be useful in choosing which input variable can be used in further studying the dynamics of West African rain producing systems.
Keywords: Machine Learning, rainfall amount, training period, error analysis.
DOI: 10.7176/JEES/12-10-06
Publication date: October 31st 2022

1.0 Introduction

Rainfall is a natural climatic parameter whose prediction is highly challenging and demanding most especially over the tropical region of the world due to the wide range of scales both in space and time, non-linearity and complex nature of the interactions of different atmospheric systems responsible for producing it. It is one of the most complex and difficult elements of the hydrological cycle to understand and model (Sivakumar, 1988). The complexity of the atmospheric processes that generate rainfall makes the quantitative forecasting of rainfall an extremely difficult tasks (Hung *et al*., 2008). Accurate forecasting of rainfall has been one of the most important issues in hydrological, weather and climate research activities because good forecast of rainfall helps to: prevent losses of property; human lives; plants and animals; support hydro-electric power generation; plan for farm operations and activities; and other socio-economic activities. Constructing an algorithm for accurate prediction of rainfall over the Savana zone of Nigeria commonly referred to as "food basket" is a very welcoming approach (Omotosho, 1990; Omotosho *et al*., 2000; and Adefisan and Abatan, 2015).

Machine learning algorithms such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF) and K-Nearest Neighbor (KNN) models are intelligent techniques, and has recently become an important alternative tool to conventional methods such as regression methods in modeling of non-linear functions such as rainfall. The challenge posed by the non-linear nature of rainfall has been argued against the traditional methods which use independent variables that are highly correlated with each other. Traditional methods cannot determine, which independent variables best predict(s) the dependent variable without duplicating characteristics (Paswan and Begum, 2013; and Amoo and Dzwairo, 2016). Thus, the advent of digital computer neural simulation has made data-driven techniques a good substitute for forecasting in time series, which is useful for rainfall prediction. Machine learning algorithms are mostly suited to problems, where a more traditional regression model cannot fit a solution (Mishra *et al*., 2016; Nicoleta *et al*., 2019; and Tian *et al*., 2021). The Machine learning algorithms uses adaptive weight functions when approximating non-linear functions of their inputs during training.

In this research, training and comparison of several machine learning methods for forecasting seasonal and monthly rainfall over Bauchi is performed. This paper also evaluates the predictive skills of the selected models using some statistical measures. All the methods are coupled with two data-preprocessing techniques. For the modeling of the rainfall, a novel hybrid multi-model method is proposed. The constituent models of the hybrid method are the artificial neural network, Random Forest, K-Nearest Neighbor, and kernel basis Support Vector Machine (SVM). The hybrid method generates sub-models first from each of the above methods with different parameter settings. The forecast, using the out of samples, is done by a weighted combination (Timmermann, 2006) of the final selected models. For evaluation of this hybrid method, we have constructed all these methods with their respective optimal parameters and applied to test sample forecasting.

2.0 Data and Methodology

2.1 *Description of the study area*

Bauchi, in the North-East geopolitical zone of Nigeria, is located within the coordinates of 10.64°N and 10.08°E as presented in Figure 1. It is bordered by Kano and Jigawa to the north, Taraba and Plateau to the south, Gombe and Yobe to the east, and Kaduna to the West. It has an elevation of 616 m. Average annual temperature is 29.3°C, maximum temperature is between 37.6ºC and 39.2ºC in April, which is the hottest month of the year (https://www.weather-atlas.com/en/nigeria/bauchi-climate; and Odiana *et al.* 2015), whereas, August is the coldest month, with an average high temperature of 29.6°C and an average low temperature of 21.4°C (https://www.weather-atlas.com/en/nigeria/bauchi-climate). Rainfall is lowest in January, with an average of 0 mm. Highest rainfall amount is in August, averaging 287 mm. Up until August, 2014, Bauchi was served by Bauchi Airport, located in-town. Scheduled airline service was then transferred to the newly constructed Sir Abubakar Tafawa Balewa International Airport, 23 kilometers north of Bauchi, near the village of Durum.

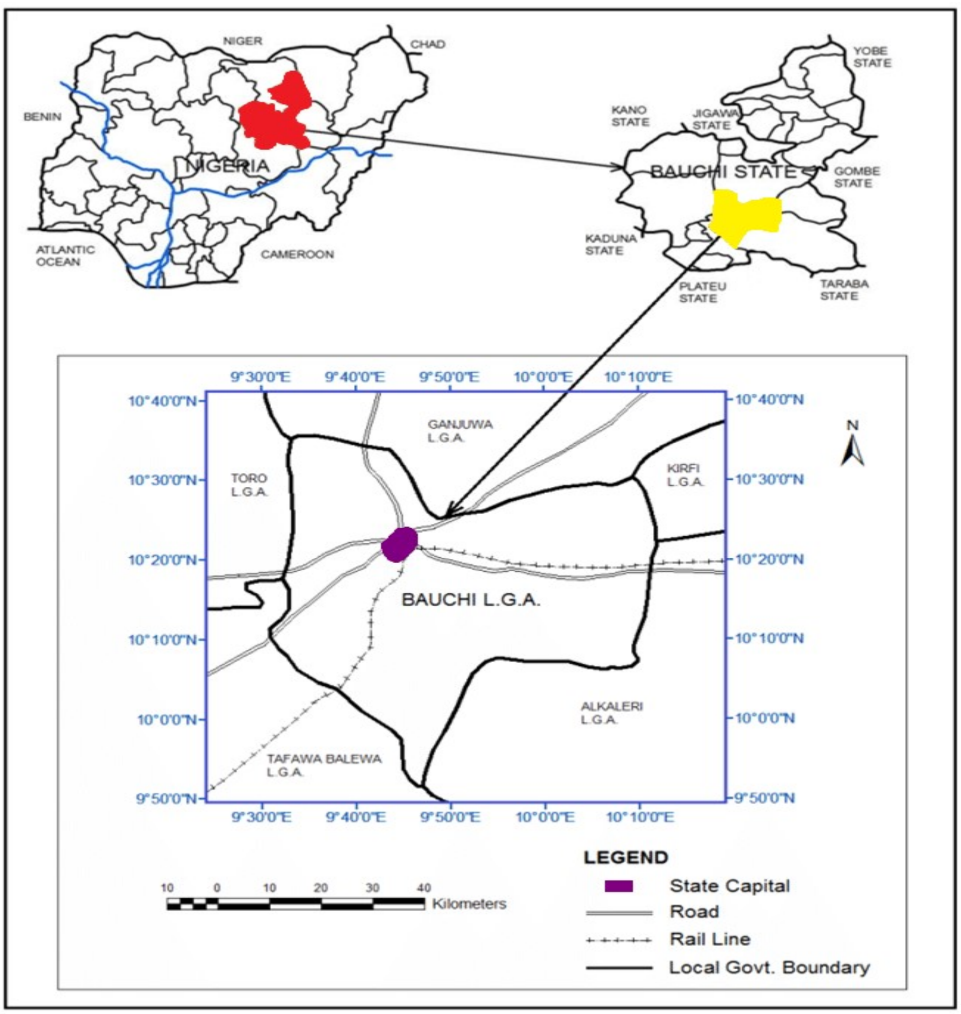

Figure 1: Showing the study area.

2.2 *Ground Observation data*

Total monthly rainfall and mean monthly air temperature data from June to October over a period of 34 years (1986-2019), were obtained from the archive of Nigerian Meteorological Agency (NiMet) Bauchi. The data was then subjected to error pruning and checks.

2.3 *Era-Interim Reanalysis Data*

The European Center for Medium-Range Weather Forecast (ECMWF) Reanalysis (ERA-Interim: Dee *et al.* 2011) with resolution of 0.125° x 0.125° were used for the present study. The reanalysis merges observations and model data across the globe using data assimilation principle. The monthly means of ERA-Interim Sea Surface Temperature (SST) and U-wind at two different pressure levels (850hpa, and 750hpa), from January to May for a period of 34years (1986-2019), over Bauchi were sourced from the archive of ECMWF and used for this study.

2.4 *NOAA El-Nino Indices Data*

The monthly climate indices used are: (1) southern Oscillation Index (SOI) which is a standardized sea level pressure differences between Tahiti (149.23°W, 17.78°S) and Darwin (130.83°E, 12.45°S) for the period of 1984 to 2019; (2) Nino1+2 and its anomaly is an index used to monitor SST over tropical pacific, it corresponds to region of coastal South America (0 - 10°S, 90°W - 80°W); (3) Nino3 for the region (5°N - 5°S, 150°W - 90°W); (4) Nino3.4 for the region (5°N - 5°S, 170°W - 120°W); (5) Nino4 for the region (5°N - 5°S, 160°E - 150°W); and (6) Oceanic Nino Index (ONI) and Multivariate El-Nino Southern Oscillation Index (MEI). These climate indices are used to represent the ENSO phenomena. Monthly means of these variables for 34 years from 1986 to 2019 to align with the observed monthly rainfall data were obtained from Climate Prediction Center (CPC) (http://www.cpc.ncep.noaa.gov/data/indices).

2.5 *Model Building Process of Machine Learning Algorithms*

The model building process consists of four sequential steps: (1) selection of the input and the output data for the supervised learning; (2) normalization of the input (Predictors) and the output data (Seasonal Rainfall)

$$x' = \frac{(x-a)}{(b-a)} \tag{1}$$

$where\ a = minimum\ value\ and\ b = maximum\ value$;

(3) training of the normalized data (1986 to 2013); and (4) testing the goodness of fit of the model using a set of test data (2014 to 2017), different from those employed in training the model were used to assess the level of skill that the model is likely to achieve in real time prediction; and (5) comparing the predicted output with the observed data reserved for evaluation.

2.6 *Model Comparison*

In this research study, a comparative analysis of the RFM, ANN model, SVM model and KNN was carried out using an open-source software called Rstudio, the ANN model was trained using the neuralnet package while the RFM model was trained using the Random Forest package in Rstudio, SVM and KNN model were carried out using e1071, caret and FNN package. The model comparison was carried out using the following criteria:

(a) Mean square error; given as:

$$MSE = \frac{\sum_{l=1}^{n}(yi-\hat{y}i)2}{n} \tag{2}$$

where: $y_i = Observed\ Rainfall(mm)$;
$\hat{y}_i = Predicted\ Rainfall(mm)$; and
$n = number\ of\ row.$

(b) Root Mean Square Error (RMSE) is given as:

$$RMSE = \sqrt{\frac{\sum_{l=1}^{n}(yi-\hat{y}i)2}{n}} \tag{3}$$

where: $y_i = Observed\ Rainfall(mm)$;
$\hat{y}_i = Predicted\ Rainfall(mm)$; and

$n = number\ of\ row.$

(c) Prediction Error (PE), given as:

$$\frac{(|y_{Predicted} - y_{Observed}|)}{(y_{Observed})} \tag{4}$$

The predictive model is identified as a good one if the PE is sufficiently small i.e., close to 0

(d) Correlation Coefficient (r)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (5)$$

where $x = Observed\ or\ actual\ rainfall\ and\ \ y = Predicted\ seasonal\ rainfall.$

3.0 Results and Discussion

3.1 *Training and Validation of Some Selected Machine Learning Algorithms*

Annual rainfall amount during the rainfall season of June to October in Bauchi was used as dependent variable with 17 parameters as inputs. The parameters are air Temperature, SST, U-wind at surface, 750hpa, 800hpa, 1000hpa, relative humidity, and specific humidity from January to May (JFMAM). The data was divided into a training data set and test data test for cross validation/model performance for each selected machine learning algorithm (SVM, ANN, RFM and KNN). Both input and dependent data must be normalized to avoid over-fitting of model results. Normalization of data is done mathematically below;

$$x' = \frac{(x-a)}{(b-a)} \quad (6)$$

$where\ a = minimum\ value\ and\ b = maximum\ value$

Presented in Figure (2a to 2b) is the mean (1986-2017) daily unscaled and scaled rainfall amount over Bauchi from June to October. While the unscaled rainfall amount ranges from 0.0mm (no rainfall day) to about 740mm (highest rainfall day), the scaled ranges from 0.0 to 1.0. Days with no rainfall is still scaled to 0.0 and the day with highest rainfall amount scaled to 1.0 and all other values in between these two extreme values range from 0.01 to 0.99.

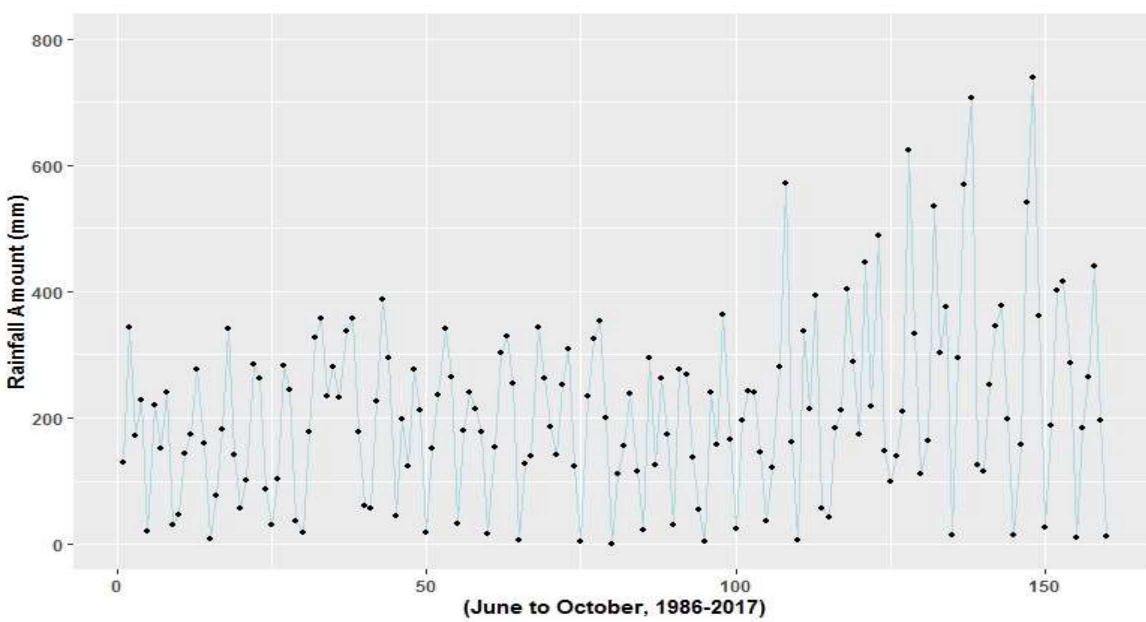

Figure 2(a): Unscaled Rainfall Amount (mm) over Bauchi.

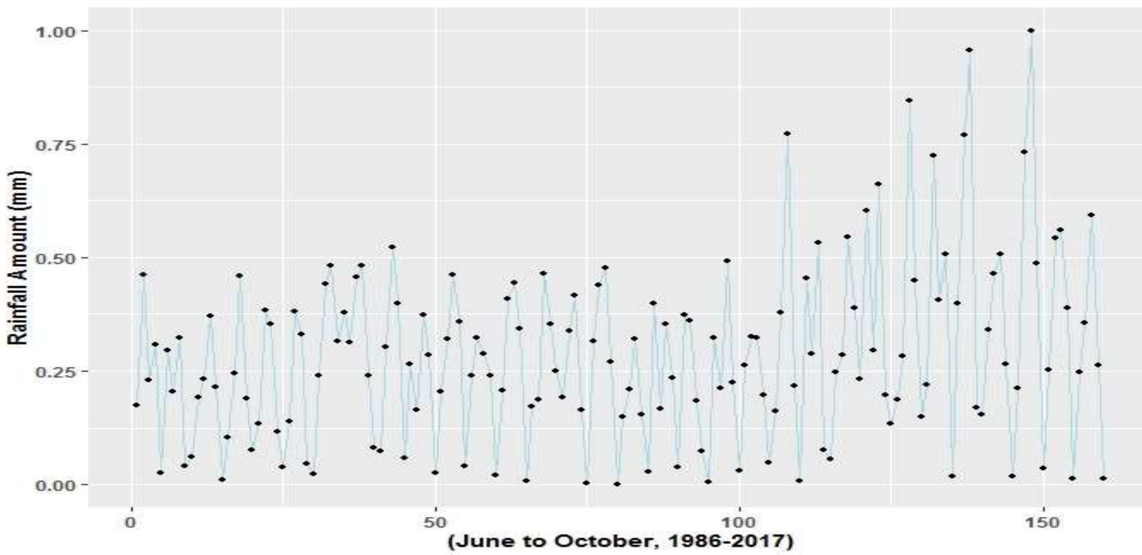

Figure 2(b): Scaled Rainfall Amount (mm) over Bauchi.

3.2 *Training Seasonal Rainfall on Artificial Neural Network (Back-propagation Algorithm)*.

Apart from the input variables, the number of hidden layers also determines the performance of the model as shown in (Figure 3). Both input and dependent data must be normalized to avoid over-fitting of model results. In Figure 3, the one hidden layer is responsible for mapping a nonlinear relationship between the seventeen (17) inputs and the output (rainfall amount). The significance of these values between these input parameters and the 1 hidden layer, accounts for capturing nonlinear and complex underlying characteristics of rainfall amount with a high degree of accuracy. However, this computation cannot deal with uncertainties. The training performance after 1000 epoch's iteration had a minimum mean square error of 0.0158 as shown in (Figure 4 and Table 1). As observed in Figure 4, the training was done using the normalized data [0, 1] from 1986 to 2013. Figure 4 shows the performance of ANN Model during training in simulating seasonal rainfall as compared with the observed seasonal rainfall from 1986 to 2013. Figure 5 shows the predicted seasonal rainfall during the test period (June to October, 2014-2017), generated from artificial neural network algorithm. This was done to further test statistically the performance of the predicted seasonal rainfall with the observed during June to October, 2014-2017.

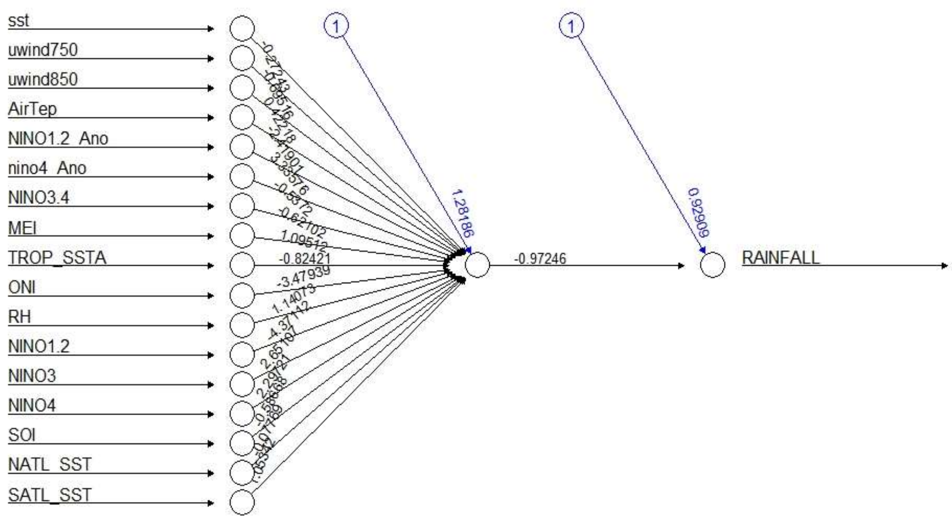

Figure 3: ANN model architecture for the seventeen (17) input parameters, 1 hidden layer and the final output (Rainfall) parameter.

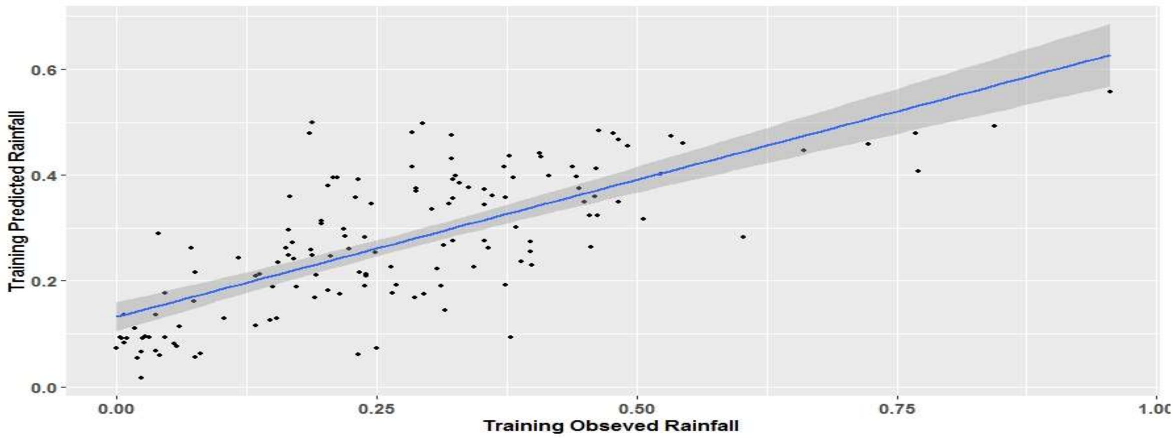

Figure 4: Training Performance of ANN from (1986-2013).

Table 1: Results of ANN training model.

Model Architectures	Epochs	MSE Training
17-1-1	1000	0.0158

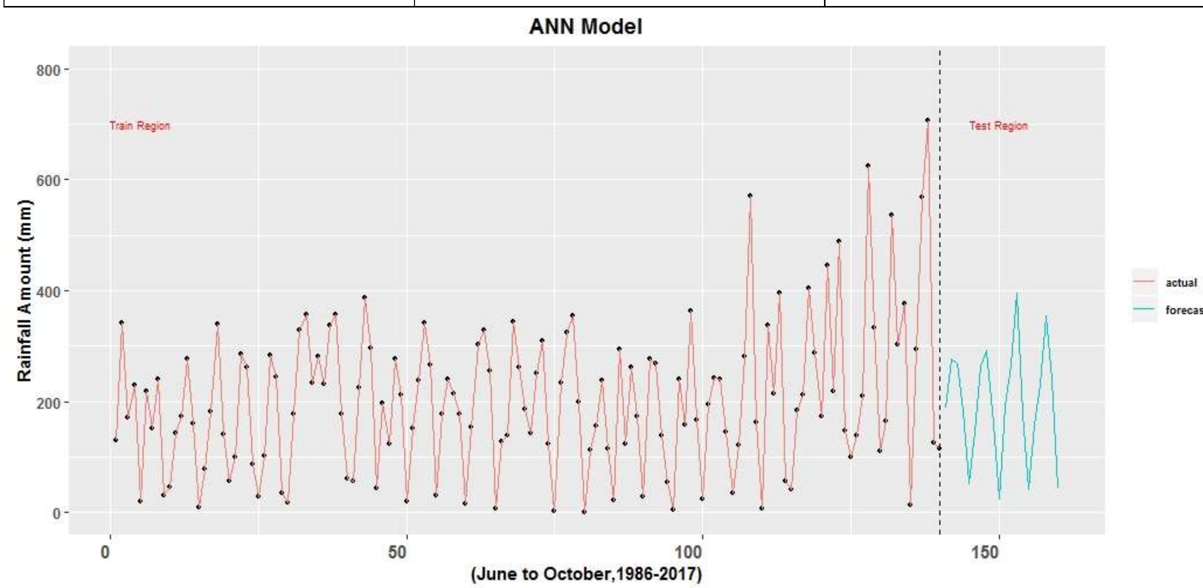

Figure 5: ANN Model Forecast for Test Period (2014-2017).

3.3 *Training Seasonal Rainfall on Random Forest Algorithm*

The Random Forest model algorithm was trained on the training dataset (1986 to 2013) in order to predict the test data (2014 to 2017) after cross validation. The model training parameters during the building process is shown in (Table 2). Figure 6 shows the RF observed training error, the RF trained model reached its maximum learning iteration at 450 trees with an error of 0.0245. In building Random Forest model, internal estimates were used to measure variable importance, as shown in Figure 7. These estimates answer or address the question: "What is the relative importance of predictor variables that contribute to Seasonal Rainfall predictions?" The variable importance list shows how much each variable improves the model's predictive capabilities (i.e., node purity or goodness-of-fit). Moreover, the variable importance list can serve as a guide for parameter selection and accuracy for modeling rainfall amount over Bauchi. The variables that mostly contributed to the reduction in prediction error as seen in Figure 7 are U component of Wind at 850hpa and 750hpa pressure levels, Nino 1+2 and Air temperature.

The model had a training performance with a mean square error of 0.0246 as depicts in Table 3, Figure 8 shows the performance of RF Model during training in reproducing seasonal rainfall while Figure 9 shows the predicted seasonal rainfall during the test period (June to October, 2014-2017), generated from the Random Forest algorithm.

Table 2: Random Forest Training Parameters

Random Forest-Type	Number of Tress	Number of split	Mean Squared residuals	% Variance explained
Regression	450	13	0.0249	24.76

Random Forest (formula =f, data = train, mtry = 13, ntree = 450)

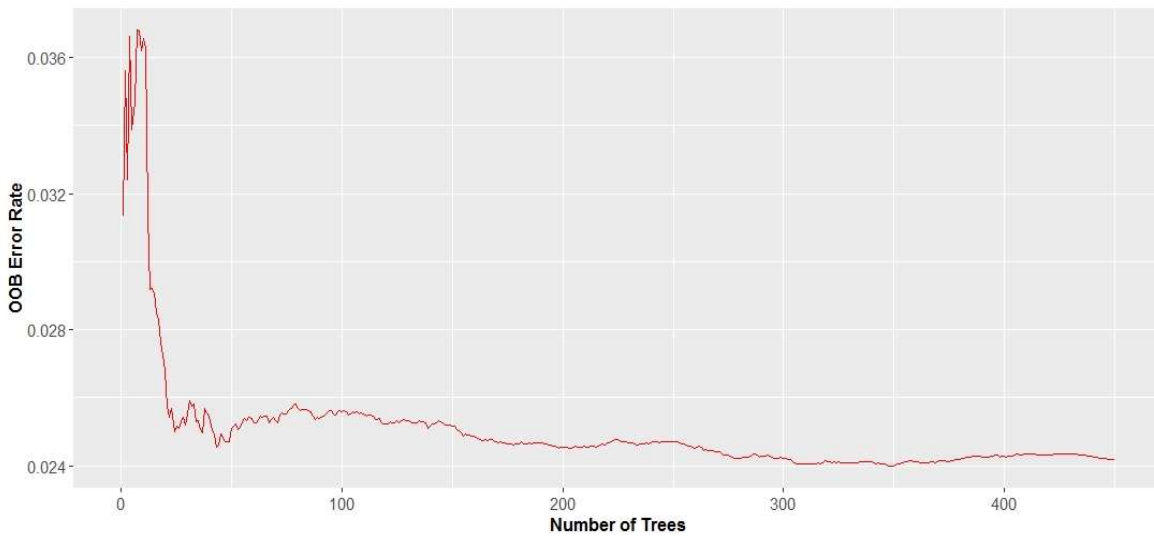

Figure 6: Random Forest model training error and number of trees.

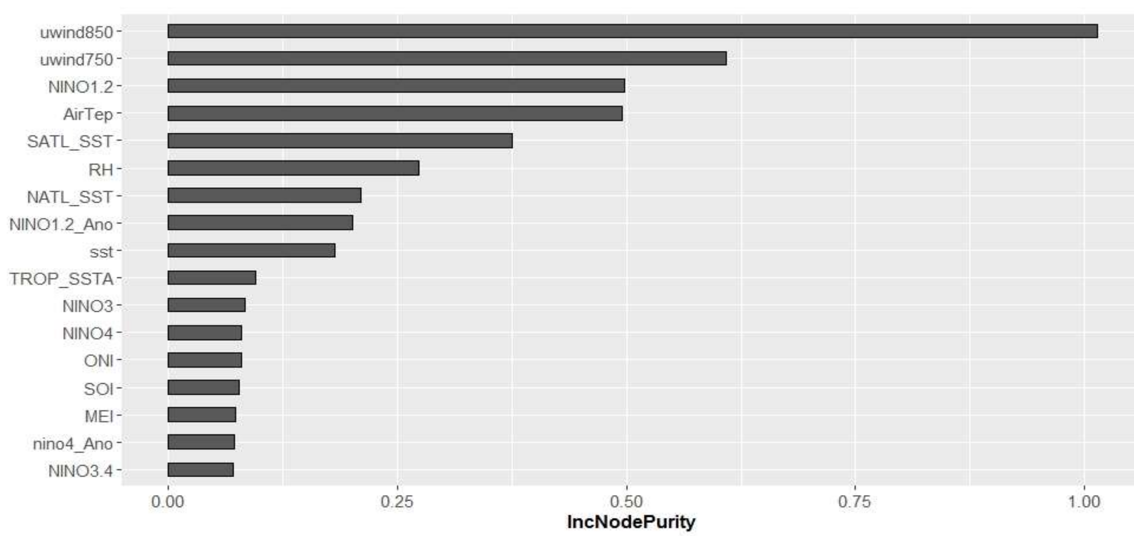

Figure 7: Relative importance of predictors in the built RFM.

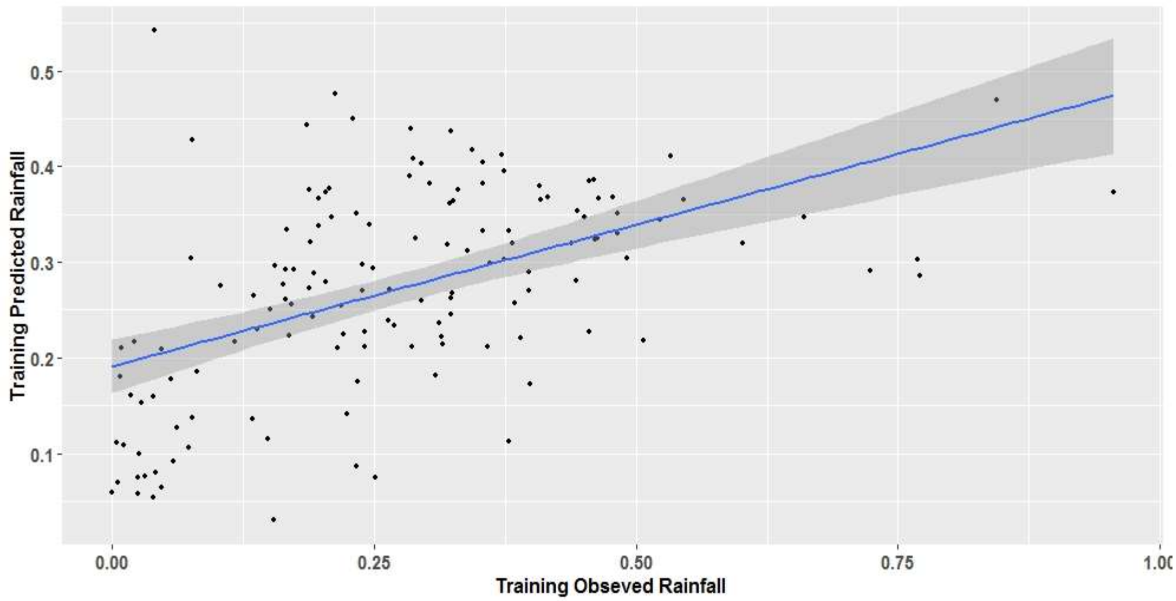

Figure 8: Training Performance of RFM from 1986 to 2013.

Table 3: Results of training Random Forest model

Model Architecture	Tress	MSE Training
Tree-Based	450	0.0246

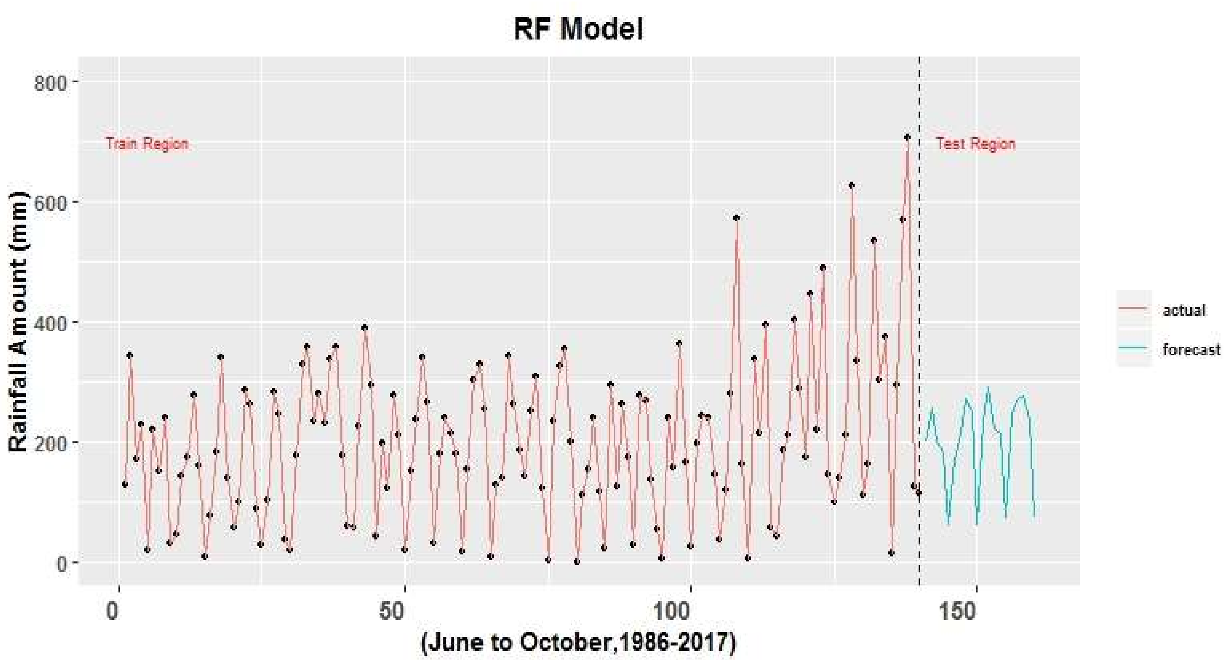

Figure 9: Random Forest Model forecast for the test period from 2014 to 2017.

3.4 *Training Seasonal Rainfall on SVM Algorithm.*

The support vector model was trained on the training dataset (1986 to 2013) using Caret and e1071 Packages in Rstudio, which is used to predict the test data (2014 to 2017) after cross validation. SVM algorithm is based on

four kernels (SVM-Linear, SVM-Radial, SVM-Sigmoid and SVM-Polynomial), during the model building process SVM-Linear performs better than other Kernels for forecasting seasonal rainfall over the study region. In selecting the best parameters for training the SVM as illustrated in Table 4, a 10-fold validation sampling was carried out, which indicated the followings: gamma=0.01; cost=5; and epsilon=0.022 (the darkest blue region is the best area of selection) as shown in (Figure 10).
The SVM model had a training performance with a mean square error of 0.0140 as depicts in Figure 11 and Table 5. Figure 12 shows the predicted seasonal rainfall during the test period (June to October, 2014-2017), generated from SVM algorithm.

Table 4: Support Vector Machine training parameters.

SVM-Type	SVM-Kernel	Cost	Gamma	Epsilon	Support Vectors
Regression	Linear	5	0.01	0.02	119

svm(formula = f, data = train, kernel = "linear", gamma = 0.1, type = "eps-regression")

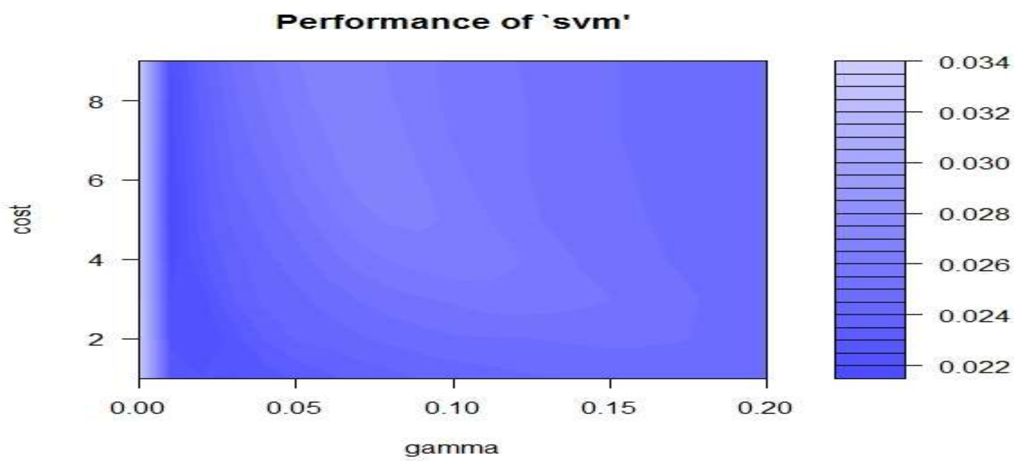

Figure 10: Selection Performance of SVM-Linear Parameters.

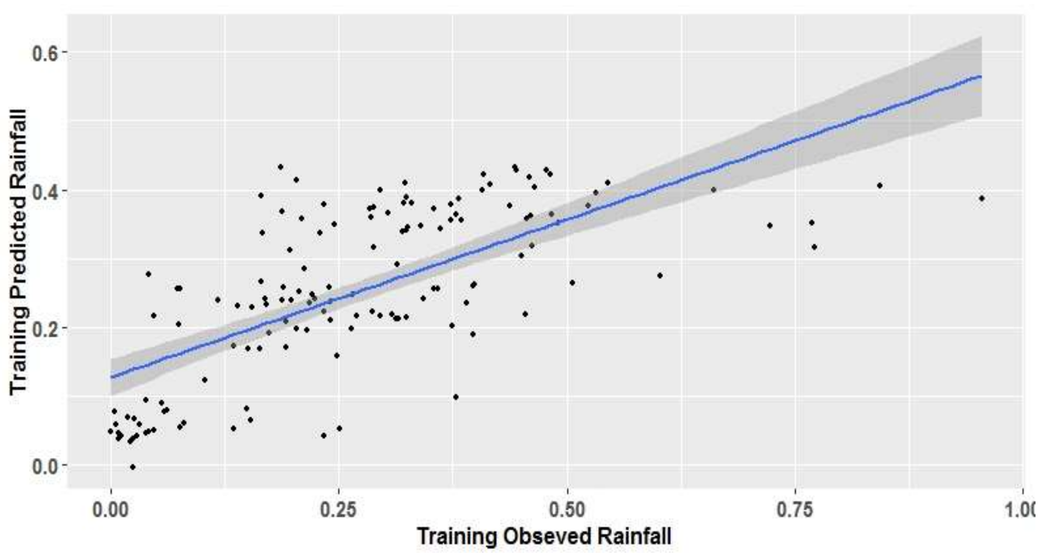

Figure 11: Training performance of SVM model from 1986 to 2013.

Table 5: Results of training SVM model.

Model Architecture	Support Vectors	MSE Training
SVM-Linear	119	0.0140

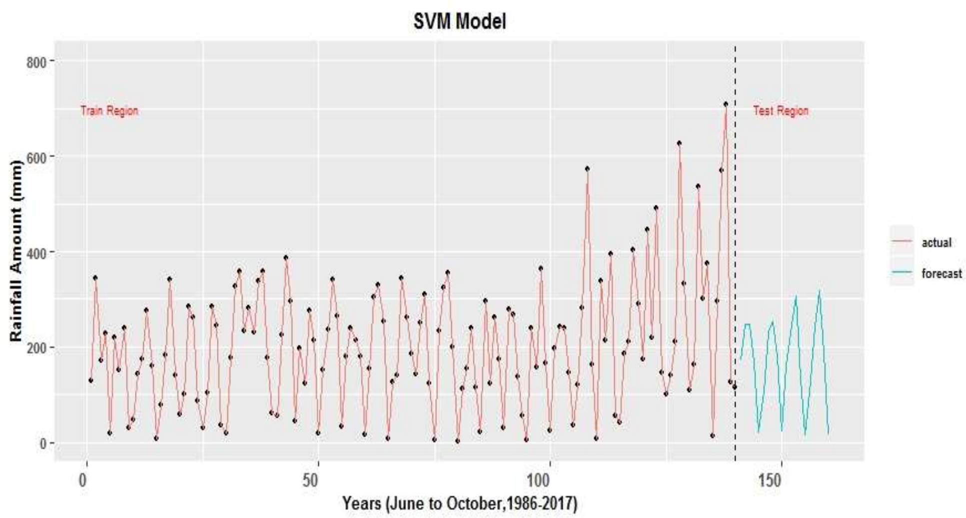

Figure 12: SVM Model Forecast for Test Period (2014-2017).

3.5 *Training Seasonal Rainfall on K-Nearest Neighbor Algorithm*

Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used to select the optimal/best K-Nearest Neighbor model using the smallest value, which was at k=9 during the training process, after re-sampling, cross-validation (10 fold, repeated 3 times) was done as illustrated and shown in Table 6 and Figure 13 from 1986 to 2013. Figure 14 shows the predicted seasonal rainfall during the test period (June to October, 2014-2017), generated from KNN algorithm.

Table 6 K-Nearest Neighbor selection criteria.

K (Neighbors)	RMSE	MAE
5	0.1520	0.1180
7	0.1505	0.1181
9*	0.1477	0.1156

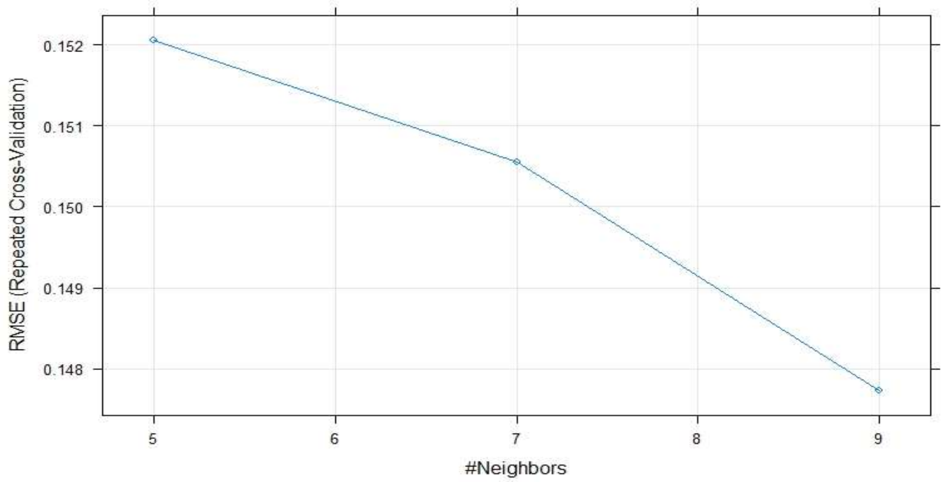

Figure 13: Best Selection of K (Neighbours) after Cross-Validation from (1986-2013).

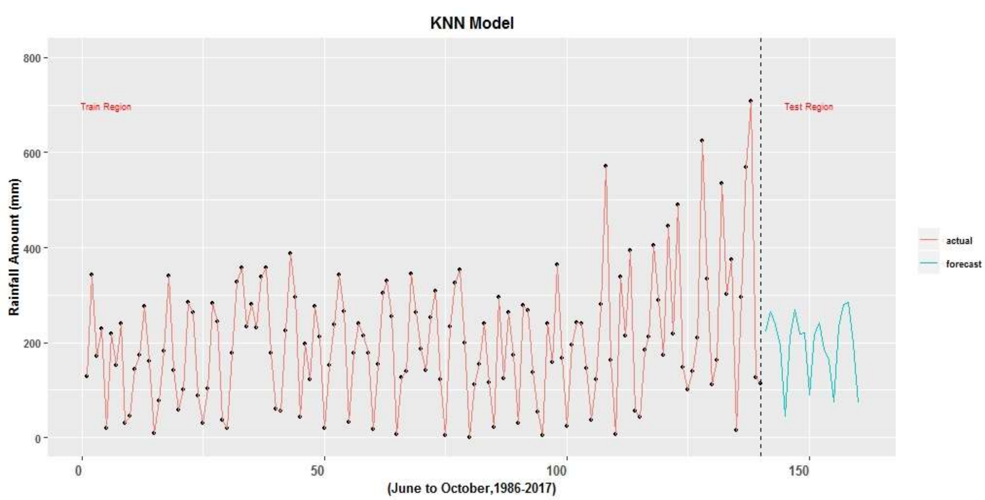

Figure 14: KNN Model Forecast for Test Period (2014-2017).

3.6 *Skill Evaluation of Machine Learning Algorithms in Predicting Seasonal Rainfall*

The skills performances of the selected machine learning models (ANN, SVM, RFM and KNN) have been evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). This was to further evaluate the performances of the selected models with the validation data over Bauchi. The results of testing performances evaluation for Bauchi station are presented in Table 7. Deduction from Table 7 indicated that Support Vector Machine (SVM) had the least MAE=96.0mm, and RMSE=141.40mm.

In Figure 15, in terms of correlation coefficient between the observed rainfall and predicted rainfall amount, Support Vector Machine (SVM) model has the highest correlation coefficient (0.84) compared to ANN, RFM and KNN models whose correlation coefficients are (0.82, 0.75 and 0.70) respectively. The analysis of the model accuracy, shows that SVM model outperformed the other three ANN, RFM and KNN in terms of MAE, RMSE and correlation coefficient (r) in prediction of seasonal rainfall for the test data (June to October, 2014-2017).

Table 7: Evaluation of Selected Machine Learning Model Using Error Analysis.

Model	MAE	RMSE	R (Cor. Coeff.)
SVM	96.0mm	141.40mm	0.84
ANN	106.73mm	155.99mm	0.82
RFM	113.18mm	157.43mm	0.75
KNN	121.13mm	169.60mm	0.7

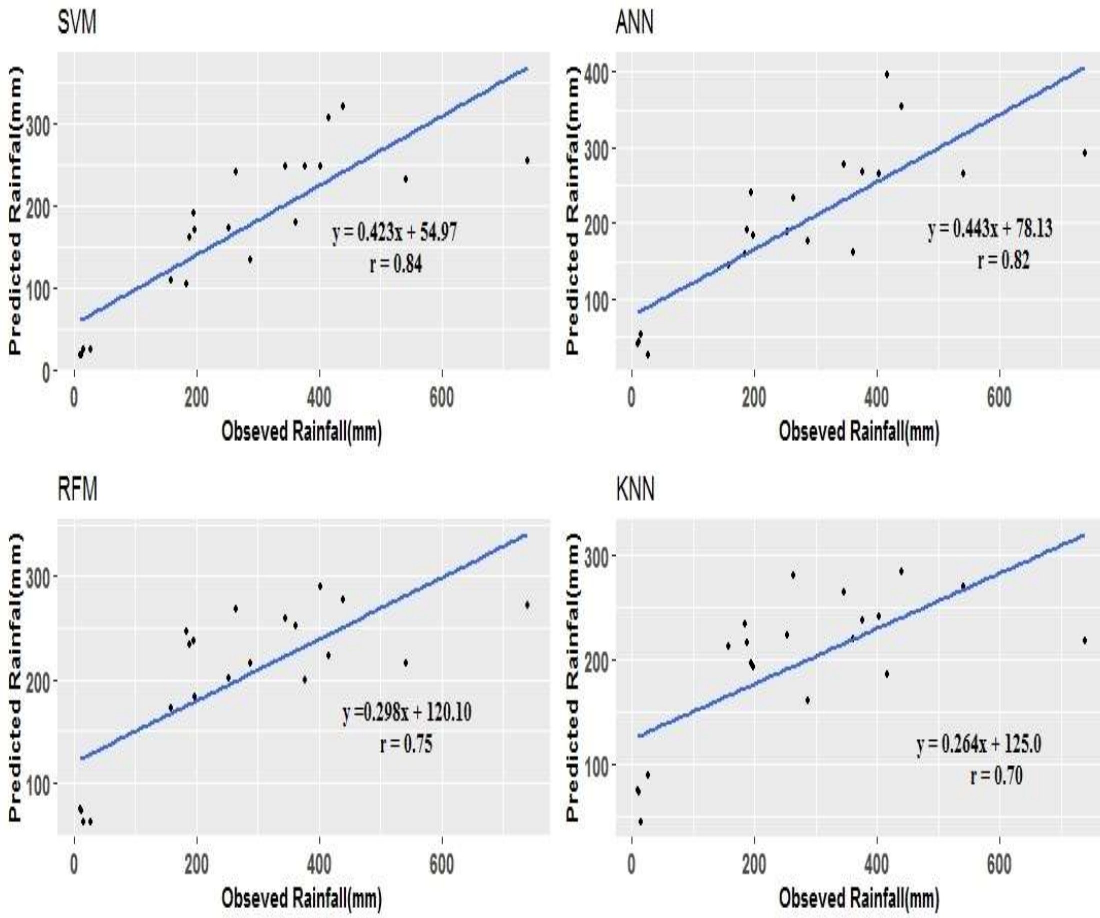

Figure 15: Comparison of observed and predicted rainfall amount (mm) for the four selected machine learning models.

3.7 *Comparison of Selected Machine Learning Models Predicted Outputs*

Figures 16, 17, 18 and 19 show the monthly predicted and observed rainfall amounts for Bauchi for the periods of June to October of 2014, 2015, 2016 and 2017 respectively. Tables 8 and 9 show predicted and observed monthly rainfall amounts and prediction bias errors of the four selected models. In Figure 16, in terms of prediction bias error between the observed rainfall and predicted rainfall amount, KNN performed better than other models in the month of June with a minimum bias error of 29.8mm In July and August, ANN performed better with minimum error of 67.2mm and 108.4mm respectively, in September KNN performed better with minimum error of 3.8mm, while monthly rainfall amounts predicted in October was very close to the observed in 2014 for all models, with SVM model having the lowest bias error of 8.0mm.

In 2015 as shown in Figure 17, ANN had the minimum bias error of 12.9mm between the predicted and observed rainfall amounts in June. In July, August and September, all models performed poorly as bias errors are large. In

October, SVM model performed better than other models with minimum bias error of 2.2mm. In terms of prediction, using the bias error between the observed rainfall and predicted rainfall amount, ANN perform better than other models in the month of June and August 2016 with minimum bias error of 3.5mm and 19.7mm respectively as depicted in Figure 18. In July and September 2016, Random Forest model (RFM) performed better than other models with minimum bias error of 112.4 and 72.4mm. In October 2016, SVM model performed better than other models (ANN, RFM and KNN) with minimum bias error of 6.5mm.

In 2017 as shown in Figure 19, ANN had the minimum bias error of 25.8mm and 85.2mm between the predicted and observed rainfall amounts in the month of June and August respectively. In the month of July, RFM performed better than other models (SVM, ANN and KNN) with minimum bias error of 4.4mm. In September, KNN was seen to have a better performance with minimum error of 2.1mm. In October 2017, SVM model performed better than other models with minimum bias error of 5.2mm.

Generally, during the test period, Artificial Neural Network (ANN) model performed better in June except for 2014 and in the month of August which is the peak of the rainy season throughout the four years, while in October which is the retreat month of rainfall, Support Vector Machine (SVM) performed better than other models throughout the four years (2014 to 2017). It can therefore be concluded that ANN and SVM are good for predicting monthly seasonal rainfall amounts over Bauchi but their efficiency or performance decreases with degree of wetness of the year or months considered.

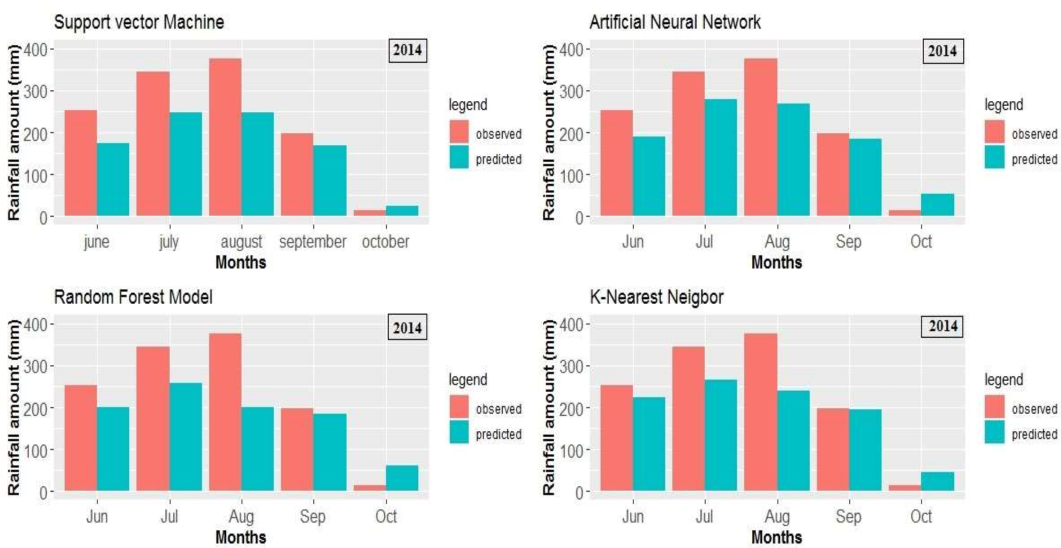

Figure 16: Monthly Predicted and Observed Rainfall Amounts for 2014.

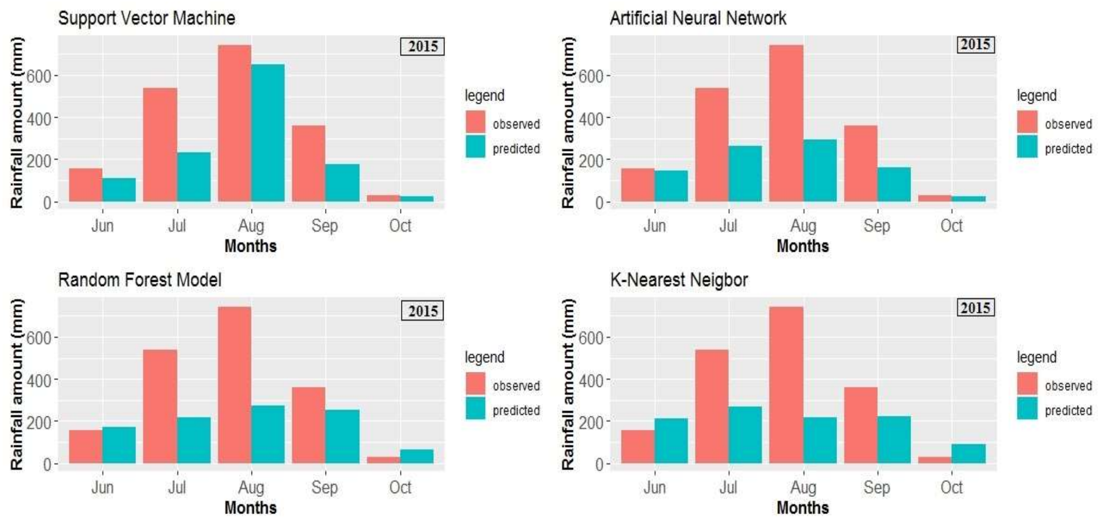

Figure 17: Monthly Predicted and Observed Rainfall Amounts for 2015.

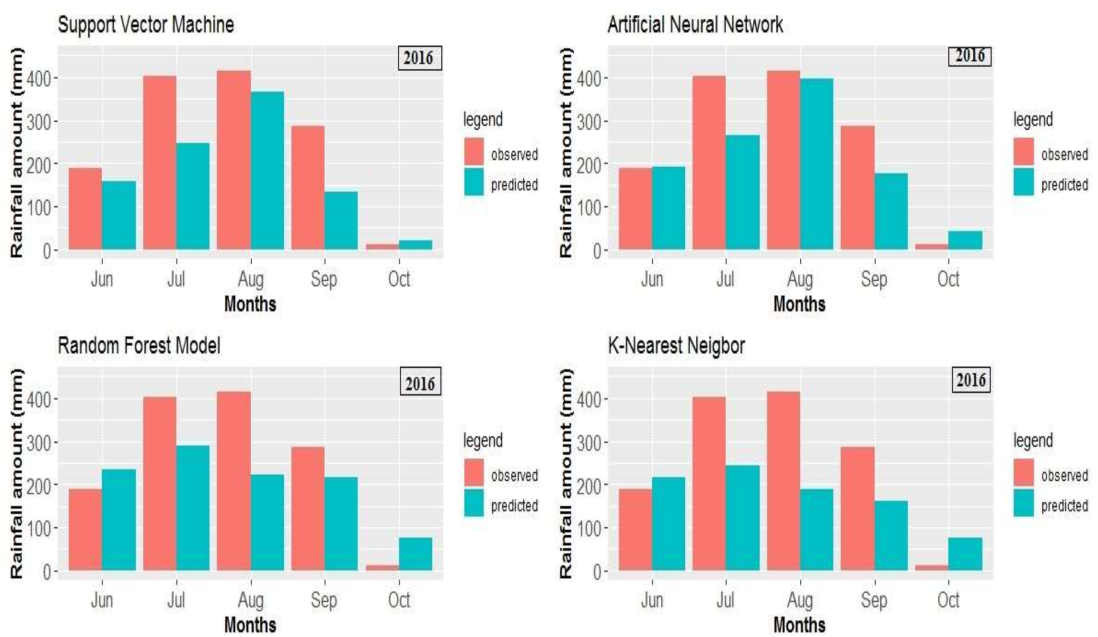

Figure 18: Monthly Predicted and Observed Rainfall Amounts for 2016.

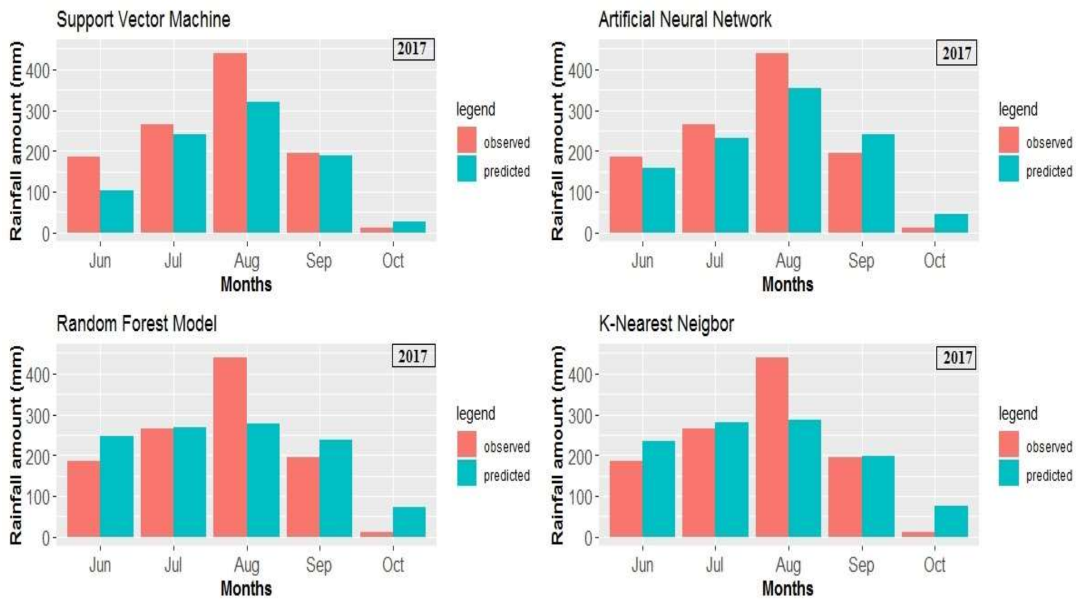

Figure 19: Monthly Predicted and Observed Rainfall Amounts for 2017.

Table 8: Predicted and Observed Monthly Rainfall Amounts

Year	Month	Observed Rainfall (mm)	SVM Predicted Rainfall (mm)	ANN Predicted Rainfall (mm)	RFM Predicted rainfall (mm)	KNN Predicted Rainfall (mm)
2014	June	253.0	172.8	188.8	200.6	233.2
	July	345.0	247.7	277.8	258.5	264.8
	August	376.6	247.8	268.1	199.3	238.1
	September	197.6	169.7	183.3	183.8	193.8
	October	15.0	23.2	52.7	61.5	45.1
2015	June	157.8	107.9	144.8	172.1	212.7
	July	540.7	231.8	264.1	215.3	269.4
	August	739.7	252.6	292.0	271.0	217.7
	September	361.2	178.9	161.1	252.1	220.3
	October	27.1	24.8	25.6	62.1	89.6
2016	June	188.5	159.3	192.0	233.6	216.1
	July	402.6	246.4	265.3	290.2	242.4
	August	415.5	366.3	395.7	223.0	187.2
	September	287.6	134.1	177.4	215.2	161.9
	October	10.0	16.5	42.2	74.5	76.2
2017	June	184.2	102.5	158.4	246.6	234.3
	July	263.7	240.7	232.6	268.1	281.4
	August	439.5	318.8	354.3	276.0	285.2
	September	195.0	189.6	241.3	237.8	197.2
	October	11.6	16.8	44.0	73.2	74.1

Table 9: Monthly Prediction Bias Error for the four selected models

Year	Month	SVM Predicted Rainfall (mm)	ANN Predicted Rainfall (mm)	RFM Predicted rainfall (mm)	KNN Predicted Rainfall (mm)
2014	June	80.2	64.2	52.4	29.8
	July	97.2	67.2	86.5	80.2
	August	128.8	108.4	177.3	138.5
	September	27.9	14.3	13.8	3.8
	October	8.0	37.4	64.3	29.9
2015	June	49.9	12.9	14.3	54.9
	July	308.8	276.6	252.3	271.3
	August	487.1	447.7	468.7	522.0
	September	182.3	200.1	109.1	140.9
	October	2.2	1.5	35.0	62.5
2016	June	29.2	3.5	45.1	27.6
	July	156.2	137.3	112.4	160.2
	August	109.2	19.7	192.5	228.3
	September	153.5	110.2	72.4	125.7
	October	6.5	32.2	64.4	66.2
2017	June	81.7	25.8	62.4	50.1
	July	22.9	31.1	4.4	17.7
	August	120.7	85.2	163.5	154.3
	September	5.4	46.3	42.8	2.1
	October	5.2	32.4	61.6	62.5

3.8 *Building a Machine Learning Web Application for Predicting Rainfall over Bauchi*

The best performed model from our study was support vector machine, its algorithm was used to forecast rainfall from June to October, 2018 to 2019 in Figure 20. The web interface was developed using R (ShinyR Package) programming available with the author on request. The beautiful graphical user interface (GUI) for home screen and the plots are respectively shown in Figures 21 (a) and (b).

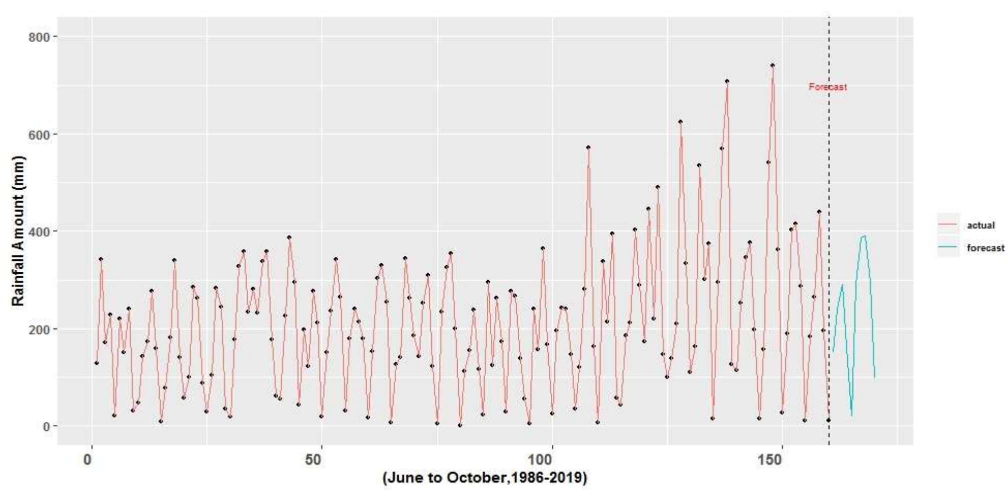

Figure 20: Rainfall Amount (mm) Forecast from June to October, 2018-2019

Journal of Environment and Earth Science www.iiste.org
ISSN 2224-3216 (Paper) ISSN 2225-0948 (Online)
Vol.12, No.10, 2022

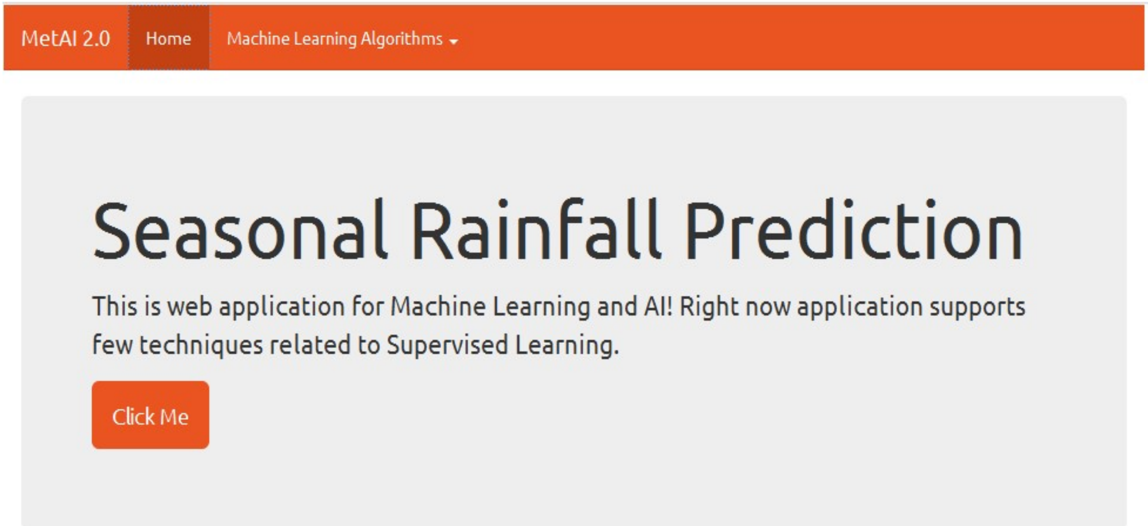

Figure 21(a): The Home screen of the Web Interface for predicting rainfall.

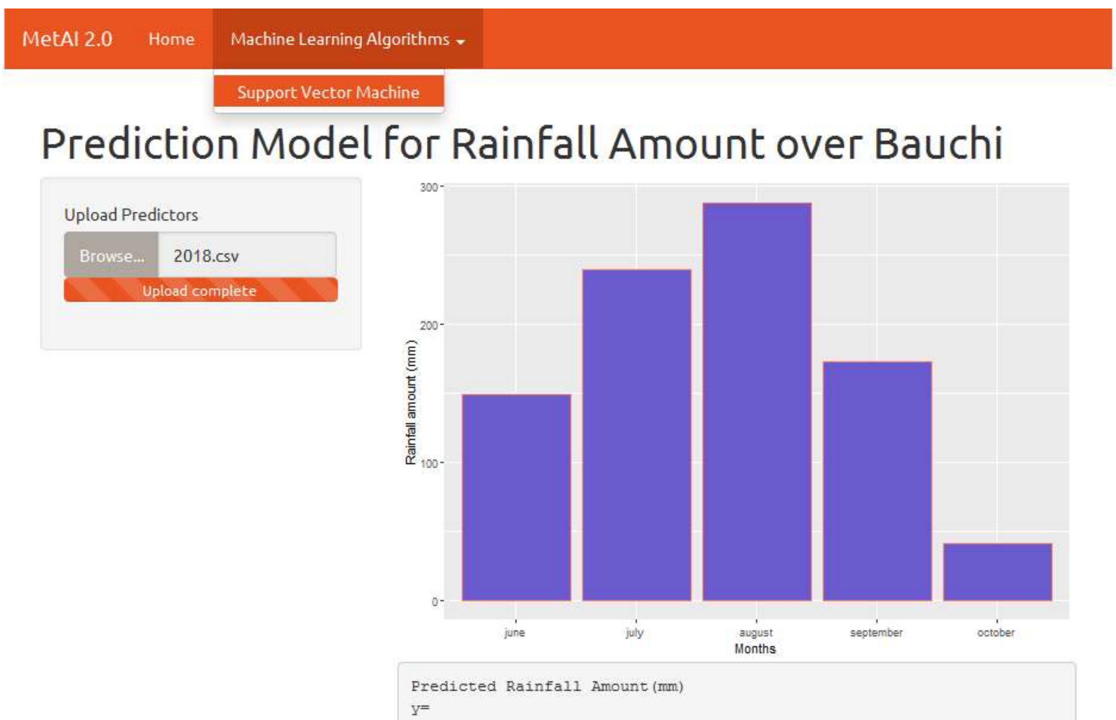

Figure 21b: Shiny-R Web Interface for Prediction Model.

4.0 Conclusions and Recommendations

4.1. *Conclusions*

The modeling of seasonal rainfall amount over Bauchi by evaluating the skills of some selected Machine Learning Models (SVM, ANN, RFM and K-NN) were performed. Our results indicated that Support Vector Machine (SVM) and Artificial Neural Network (ANN) are good methods of optimization, since errors observed in the comparison of observed and model output are minimal. The SVM and ANN summary yielded 84% and 82% respectively of good forecasts for seasonal rainfall amount over Bauchi. This showed that the trained network is reliable and fit to be used for the subsequent quantitative prediction of rainfall. Therefore, it can be concluded that SVM and ANN model with seventeen (17) input parameters considered in this study will perform well in predicting seasonal and monthly rainfall amount over Bauchi. By extension, they will perform very well over West Africa if historical data is available but its efficiency or performance decreases with degree of wetness of the year or months considered. The results from this study will provide information that will aid accurate seasonal rainfall prediction using the required parameters. In other words, agriculturist, water resources managers, power generation expert and other related sectors can adopt any of these as a reliable forecast tool.

4.2. Recommendations

The result of this study has shown that SVM and ANN models which are capable of modeling complex non-linear problems, has the ability of predicting seasonal and monthly rainfall hence, it is suitable for predicting rainfall amount. The challenge of sparse meteorological data has been an issue that reduces the representativeness of a system and can therefore have significant effect on the conclusion that can be drawn. This should be looked into by related institutions as it will bring a lasting solution to errors in the analysis. Also, before any model can be used with confidence, adequate validation or assessments of the magnitude of the error that may result from their use should be performed.

Acknowledgements

The authors are grateful to the Nigerian Meteorological Agency for sponsoring the third author for the program in FUTA and releasing the observed rainfall and temperature data used for the study.

References

Adefisan E. A. and Abatan A. A. 2015. Agroclimatic zoning of Nigeria based on rainfall characteristics and index of drought proneness. *Journal of Environment and Earth Science*, **5**(12): 115 – 127.

Amoo O. T. and Dzwairo B. (2016): Trend analysis and artificial neural networks forecasting for rainfall prediction. *Environmental Economics*, **74**(4-1), 149 – 160. doi:10.21511/ee.07(4-1).2016.07

Hung N. Q., Babel M. S., Weesakul S. and Tripathi, N. K. (2008): An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences*, **13**(8):1413-1425.

Mishra S., Mishra D. and Santra G. H. (2016): Applications of machine learning techniques in agricultural crop production: A review Paper. *Indian Journal of Science and Technology*, **9**(38) DOI: 10.17485/ijst/2016/v9i38/95032

Nicoleta T., Stavros A. and Manos R. (2019): Data-Driven Decision Making in Precision Agriculture: The rise of big data in agricultural systems. *Journal of Agricultural and Food Information*, **20**(4): 344 – 380. https://doi.org/10.1080/10496505.2019.1638264

Omotosho J. ‘Bayo (1990): Onset of Thunderstorms and precipitation over Northern Nigeria, *International Journal of Climatology*, **10**, 849 – 860.

Omotosho J. 'Bayo, Balogun A. A., Ogunjobi K. O. (2000): Predicting monthly and seasonal rainfall, onset and cessation of the rainy season in West Africa using only surface data. *Inter. J. Clim.* **20**:865-880.

Paswan R. P. and Begum S. A. (2013): Regression and Neural Networks Models for Prediction of Crop Production, *International of Scientific and Engineering Research*, **4**(9): 98 – 108.

Sivakumar M. V. K. (1988): Predicting Rainy Season Potential from the Onset of Rains in Southern Sahelian and Sudanian Climatic Zones of West Africa. *Agricultural and Forest Meteorology*, **42**:295 – 305.

Tian H., Wang P., Tansey K., Zhang J., Zhang S. and Hongmei L. (2021): An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agricultural and Forest Meteorology*, **310** https://doi.org/10.1016/j.agrformet.2021.108629.

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, **1**, 135-196.