Ilorin Journal of Science
Volume 10, Number 1, 2023, pp.44 – 61 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2023 Faculty of Physical Sciences, University of Ilorin
https://doi.org/10.54908/iljs.2023.10.01.004

ILORIN
JOURNAL OF SCEINCE

ILJS-23-013

Flood Prediction in Nigeria Using Ensemble Machine Learning Techniques

Oloruntoba*, K. R., Taiwo, K. and Agbogun, J. B.

Department of Computer Science, Federal University Lokoja, Kogi State Nigeria.

Abstract

Flooding is the most frequent and destructive natural catastrophe that may happen anywhere in the globe. The frequency and severity of flooding events have increased worldwide in recent years due to climate change and human activity. Flooding has caused widespread death and devastation of property, farms, and vegetation in several emerging African nations, including Nigeria, and has forced the relocation of many more. Flooding has been Nigeria's most common natural disaster during the last decade. Modern machine learning methods have shown great promise for improving flood prediction. The optimum machine learning algorithm for flood prediction is a matter of debate. To reduce the harm caused by floods, finding better ways to anticipate their occurrence is crucial. In this paper, 7 machine learning algorithms (SVM, CART, KNN, GLMNET, LG, LDA and NB) were initially applied on the default dataset. The results reveal fair accuracy (over 60%) and kappa values (< 0.4). The same set of ML algorithms were again applied on the transformed dataset using boxcox transformation technique; the accuracy and kappa values improved but not significantly. Finally, Models for predicting floods were implemented using five different ensemble algorithms: Bagged CART (BAG), Random Forest (RF), Stochastic Gradient Boosting (GBM), Extreme Gradient Boost (XG Boost), and C5.0 (C50). Compared to the other three models, the performance of RF (AUC = 0.93) and BAG (AUC = 0.92) indicated superior accuracy and Kappa.

Keywords: Flood, Flood prediction, Machine Learning, Ensemble techniques, Bagging, Boosting.

1. Introduction

The term "flood" refers to the temporary overflow of water onto normally dry ground. There are few natural calamities as catastrophic as flooding. Excessive precipitation, snowfall, coastal storms, storm surges, and the failure of dams and other water management systems are all potential flooding causes. By definition, a flash flood occurs when a river overflows its natural levels, inundating the land around it for a brief period (Xie et al., 2020). As a result of several unfavourable environmental factors, including meteorological, hydrological, geomorphological, and human participation in the breakdown of flash flood protection measures, it is essential to remember that flash floods are a distinct phenomenon. A rise in the frequency and severity of worldwide flash flood dangers has been linked to continuing global climate change over the last several decades.

Corresponding Author: Oloruntoba, K. R.
Email: kolawole.oloruntoba-pgs@fulokoja.edu.ng

Devastating flash floods are caused partly by widespread human interference with natural systems, including forest ecosystems, as shown by deforestation, riverbed sedimentation, and the encroachment of human settlements and dam building on riverbeds. Recent years have seen a shift in the severity pattern of flash floods due to the progressive growth in world population, particularly in developing nations (Mosavi *et al*., 2017). Flash floods may cause significant socioeconomic losses. These damages include destroying homes and lives and critical infrastructure, including farms, factories, and communication networks. Several people are displaced, and many more are killed yearly due to flash floods. Nigeria happens to be one of the nations most vulnerable to flooding. High rainfall intensity, the propensity to create runoff, the rapidity of the rainfall-runoff process, soil characteristics and infiltration rate, a poorly maintained flow pattern of a river system, and changes in land use are all contributors to the possibility of a disastrous flash flood. The literature demonstrates that floods have the highest fatalities among all natural catastrophes (Panahi *et al*., 2021). Maspo *et al.* (2020) reported that flooding is a major frequently occurring natural catastrophe with serious consequences on lives, infrastructure property and the surroundings. While stopping flooding is difficult, one can reduce its impact through more accurate predictions. This fact was corroborated by Mosaffa *et al.* (2022). Flooding has also caused irreversible harm to the ecosystem, property, human life, and infrastructure like bridges, buildings, roads, and many more (Egbinola *et al.*, 2017).

Predicting the likelihood of future flash floods based on the frequency of flash floods is a crucial part of flood risk assessment. As a result, many types of flash flood statistics, including discharge, rainfall, and runoff, have been used to quantify the recurrence of flash floods in the past (Xia *et al.*, 2017). Devastating flash flood damage necessitates various structural and non-structural methods for long-term mitigation and prevention. Floods are one of the most devastating natural catastrophes, not just in Nigeria but also in many other countries of the globe. Recent floods have significantly impacted damage to human life, property, infrastructure, and the economy and social fabric. Thus, creating flood forecasting models that can provide precise maps of potentially vulnerable locations is crucial, allowing for better measures to reduce and respond to flood risks (Mosavi *et al.*, 2018). Thus, cutting-edge technologies are crucial for short- and long-term flood forecasting. Hydrological event forecasting traditionally relies on physically based models (Mosavi *et al*., 2018). According to Akinyokun *et al*. (2020), several communities continue to experience the catastrophic effects of floods resulting from climate change, acute rainfall, rapid increase in population, and industrialisation, among others.

2. Background and Related Works

Basically, the techniques for flood prediction can be broadly categorized into three, namely, physical, statistical and data driven. The physical method combines hydraulic and computer hydrological models. It has a distinct physical basis. However, a significant amount of information about the river basin, which is typically scarce, must be deposited. According to Mosavi *et al.* (2018), the statistical method has limited performance capabilities and is typically not utilised to predict floods because it does not adequately expose the nonlinear underlying elements that are crucial to flooding processes.

The use of physical and statistical techniques for flood prediction has several limitations, such as their susceptibility to ambiguous and subjective interpretation. Additionally, they don't provide quantitative flood predictions, have a low level of prediction capability, and are inaccurate. Modern data-driven models like machine learning are used as a result of the limitations of the physical and statistical models that were previously addressed.

Studies have revealed a gap in the short-term prediction capability of physical models (Mosavi *et al*., 2018). Machine learning methods for flood forecasting have emerged as a response to the limitations of physically based and statistical models. The following are some of the benefits that may be gained by using Machine Learning for flood prediction: Faster development with fewer inputs; more straightforward implementation with low computation cost; faster training, validation, testing, and evaluation; relatively less complexity; and the ability to numerically formulate the flood's non-linearity based on historical data alone, without knowledge of the physical processes.

Forecasters have attempted to anticipate floods in several ways, each with advantages and disadvantages and varying degrees of success. It's not like there's a model that everyone agrees on. The precision, speed, and data-distribution assumptions of available models vary widely. Most hydrological event forecasts have been made using physical models (Zhao et al., 2014). Yet, this often requires in-depth knowledge and skill regarding hydrological aspects, which may be complex and demanding. Research has shown that specific physical models cannot provide predictions soon (Costabile and Macchione, 2015).

Several machine learning methods have been employed for flood modelling. These include long short-term memory (Li *et al.*, 2021), linear models, fuzzy logic, artificial neural network, multi-layer perceptron, Naïve Bayes, and decision trees (Pham *et al.*, 2021). Machine learning methods for flood forecasting have emerged as a response to the limitations of physically based and statistical models. Ardabili *et al.* (2019) wrote that conventional machine learning algorithms continuously advance and evolve quickly by introducing novel learning algorithms using hybridisation and ensemble techniques. The hydrological strategy, which uses hydrological and hydraulic modelling, was the conventional one in the past. According to Tehrany *et al.* (2019), the qualitative model considers the influencing elements and their qualities while modelling. It uses expert knowledge and qualitative methodologies to associate independent variables with flood incidence based on numerical expressions. The frequency ratio (FR), logistic regression (LR), and the index of entropy (IOE) are only a few of the standard statistical methods.

Nevertheless, Tehrany *et al.* (2019) pointed out that statistical techniques depend significantly on linearity assumptions, and flooding does not fit that description. Statistical methods such as statistical correlations using the gauge to gauge, gauge discharge data, multiple coaxial correlations using gauge, rainfall, and antecedent precipitation index (API) data are used by Nimet and NiHsa. Nimet and NiHsa are the two sister agencies in Nigeria responsible for flood forecasting in Nigeria to provide the flood forecast. Hydrological event forecasts have been based on physically based models (Zhao *et al.,* 2014).

Accurately predicting future floods with the available forecasting methods is challenging. Existing flood forecasting methods do not offer a way to assess uncertainty in inputs and models. Short-notice flood forecasting is paramount for decision-making and disseminating flood alerts and warnings. While floods occur yearly, operational flood forecasting has made modest advances in recent years.

In order to create machine learning prediction model, historical records of flooding are used in conjunction with data sourced from several rain gauges. The dataset typically includes rainfall and water levels obtained from ground rain gauges or using remote sensing technologies.

Isaac *et al.* (2021) reported that findings indicate that during the past decade, hybrid models have been used extensively by machine learning researchers than individual models. This is because the models' strengths and shortcomings complement one another. Statistical methods, physical models, and soft computing techniques are typically combined to create hybrid models.

According to Panahi *et al.* (2020), there is no general agreement on which machine learning method is best for flood prediction. Hence new techniques, usually a hybrid of different algorithms, are often explored. Researchers in machine learning models for flood prediction have used more hybrid models than stand-alone models in the last decade. Hybrid models complement each other in terms of strengths and weaknesses.

3. METHODS

The framework/process flow for the study is shown in Figure 1. It consists essentially of five steps, namely: data collection and data pre-processing, the definition of the training set (data splitting and training), application of machine learning algorithms on untransformed features, application of algorithms transformed with Box- cox and advance to better performance using ensemble algorithms.

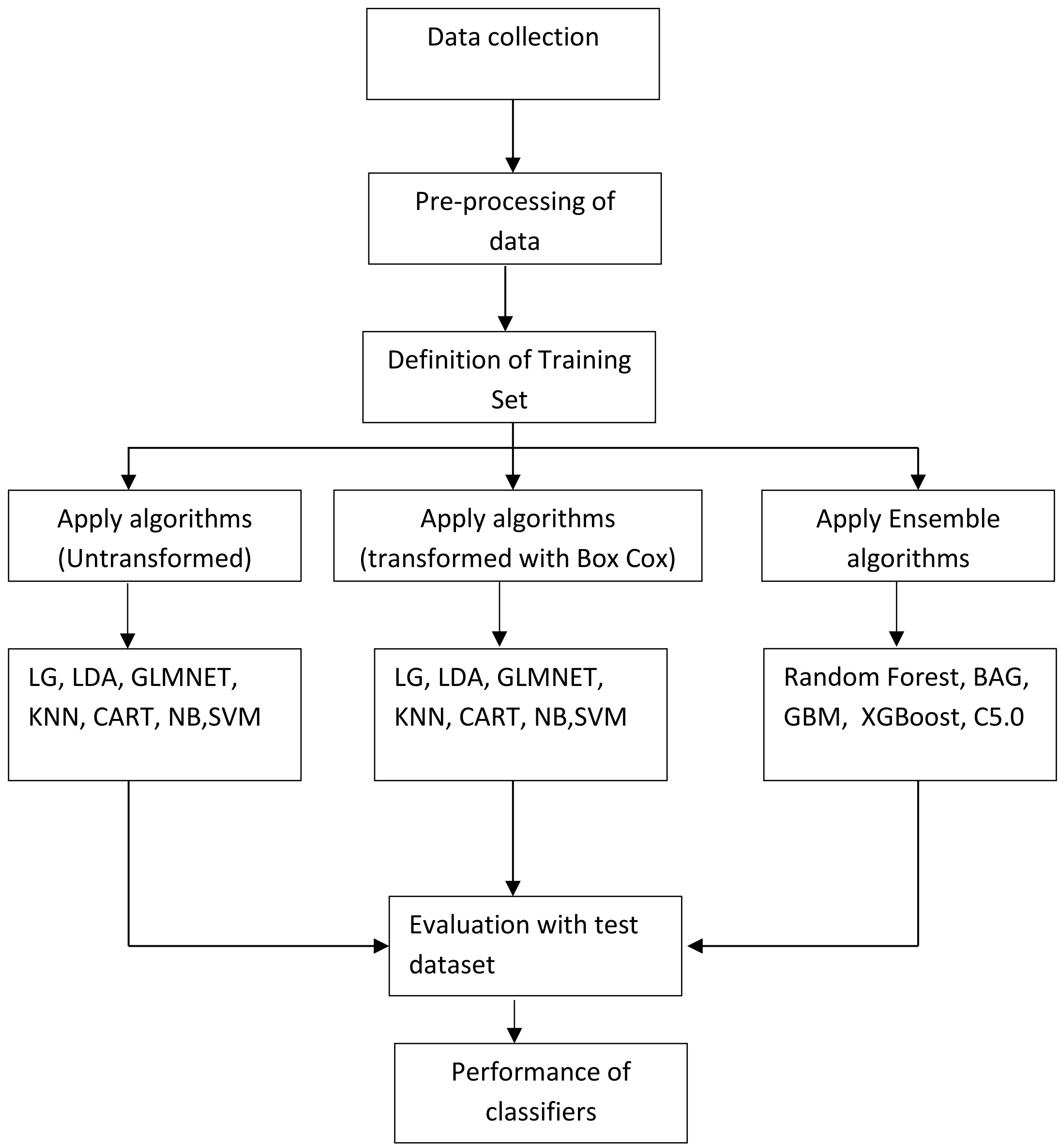

Figure 1: Workflow of the Flood Prediction Model

3.1. Data collection

Information used in the analysis was obtained from the Nigerian Bureau of Statistics (NBS) website. The bulletin was analysed for the following details: precipitation, maximum and minimum temperatures, average radiation and evaporation, relative humidity, and flood reports. Data were included from 2012 to 2018. The size of the informational file is 21 kilobytes. The variables are listed in the columns, while the yearly data for the 36 states and the FCT are shown in the rows. In this study, we used annual rainfall, mean radiation, mean evaporation, relative humidity, minimum and maximum temperature, and the number of reported flood cases as the dependent variable and the number of reported flood cases as the outcome variable.

3.2 Data Pre-processing

The variables required for this analysis have been retrieved from the original dataset, and they are as follows: annual rainfall, mean radiation, mean evaporation, relative humidity, minimum and maximum temperature, and reported flood incidents from 2012-2018. A longitudinal sample was obtained from 36 states and the Federal Capital Territory (FCT), Abuja. After being extracted, the dataset was entered into an Excel spreadsheet and saved using a comma-delimited filename (.csv). The information is sent to R Studio and placed in the "Ensemble data" data frame. When dealing with missing data in R, the mean feature was used to fill in the gaps.

3.3 Data Splitting and Training

Separating the dataset is the last step in the pre-processing phase of the data. The information is going to be divided into training and testing sets. The model will be trained on 80% of the data and verified using the other 20%. Both classification and regression are viable applications for machine learning algorithms. Since our outcome variable is binary, we used classification machine learning techniques in this investigation (0 or 1). Accuracy, Kappa, McNemar's Test, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, Prevalence, Detection Rate, Detection Prevalence, and Balanced accuracy are some of the performance metrics employed.

As can be seen in figures 3 and 4 below, the dataset are not normally distributed, Most ML models perform better if the data is normally distributed. A box-cox transformation on the dataset is employed to make the prediction better. The analyses of the results of ML algorithms of untransformed dataset and transformed dataset are given in sections 4.4 and 4.5 respectively.

3.4 System Implementation

All the tests were run on a 1.8 GHz Intel Quad-Core i5-82500U with 8 GB of Memory and 64-bit Windows 10 Home Edition. We used 10-fold cross-validation with three repetitions to divide the datasets. The classifiers' efficacy is evaluated here using a 10-fold cross-validation method. Now, we break up the training dataset into 10 equal-sized subgroups and put each of those subsets through the classifier trained on the other nine. The computational cost of doing cross-validation is minimised by performing the procedure 10 times in a ten-fold cross-validation, which is one of its many benefits. In addition, because each data point is only tested once and used for training ten and a half times in other validation methods, 10-fold cross-validation produces less bias.

3.5 Implementation Tools

The research used the statistical capabilities of the R-project software [R version 4.0.5 (2021-03-31)]. The R Core Team and the R Foundation for Statistical Computing maintain R, a free software environment for programming, statistical computation, and graphics. Data miners and statisticians often use the R programming language for various applications, including creating statistical tools and examining large datasets. To complete the R-project, the CART library was included as a package. Classification And Regression Training (CART) is a collection of tools meant to simplify the development of prediction models. Data partitioning, feature selection for pre-processing, resampling for model tuning, assessment of variable relevance, and other features are all included in the package. To import the dataset into R-Studio for pre-processing and analysis, it was extracted from the portable document format (pdf) and then put into Microsoft Excel as comma-separated values (CSV).

4. Results and Discussion

The present study's experimental analyses were carried out using a laptop equipped with an Intel® coreTM i7-4340M CPU @ 2.90GHz (4 CPUs) 8GB RAM, and the Python programming language, together with the Scikit-Learn, Matplotlib, pandas Pycaret, and seaborn libraries. In the following sections, we describe the findings from our experiments.

4.1 Unimodal Data presentation and Visualizations

The research data is presented in Table 1. The whole data set is shown in Figures1-4. Figure1 shows the predictive variables of the flood. Figure 2 presents flood incidence in Nigeria, and Figures 3 and 4, present histogram and density plots for the independent variables, respectively.

Table 1: Descriptive Statistics of Predictor Variables over the Study Period

	Rainfall (mm)	MinTemp (°C)	MaxTemp (°C)	MeanRad (kW/m^2)	RelHum (%)	MeanEvap (mm)
Mean	1347.25	22.25	33.73	19.90	63.74	4.78
Std. Deviation	903.79	2.02	2.13	1.91	14.14	0.46
Minimum	109.00	15.40	27.40	15.+70	34.40	3.70
Maximu m	10719.00	27.00	41.10	24.00	85.50	5.90
Kurtosis	43.18	1.61	1.59	-0.84	-0.92	-0.59
Skewness	4.42	-1.09	0.32	0.21	-0.23	0.08

Table 1presents the descriptive statistics of predictor variables over the study period. It shows that rainfall has a high deviation from the mean value. Also, relative humidity has high variability, so we can expect extreme rainfall values more frequently, which can cause hazards. In addition, the table also indicates that the distribution is highly skewed for rainfall, and min temperatures are approximately symmetric for the other predictor variables.

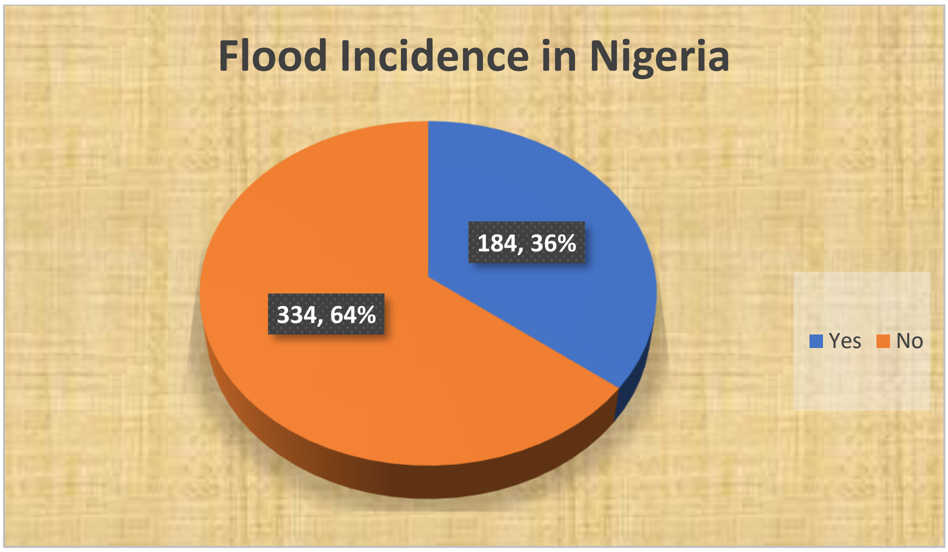

Figure 2: Pie-chart showing flood incidence in Nigeria over the study period.

The pie chart shows the flood incidence in Nigeria over the study period. It reported 184 (35.5%) cases of flood and 334 (64.5%) cases of no flood in different states of the country.

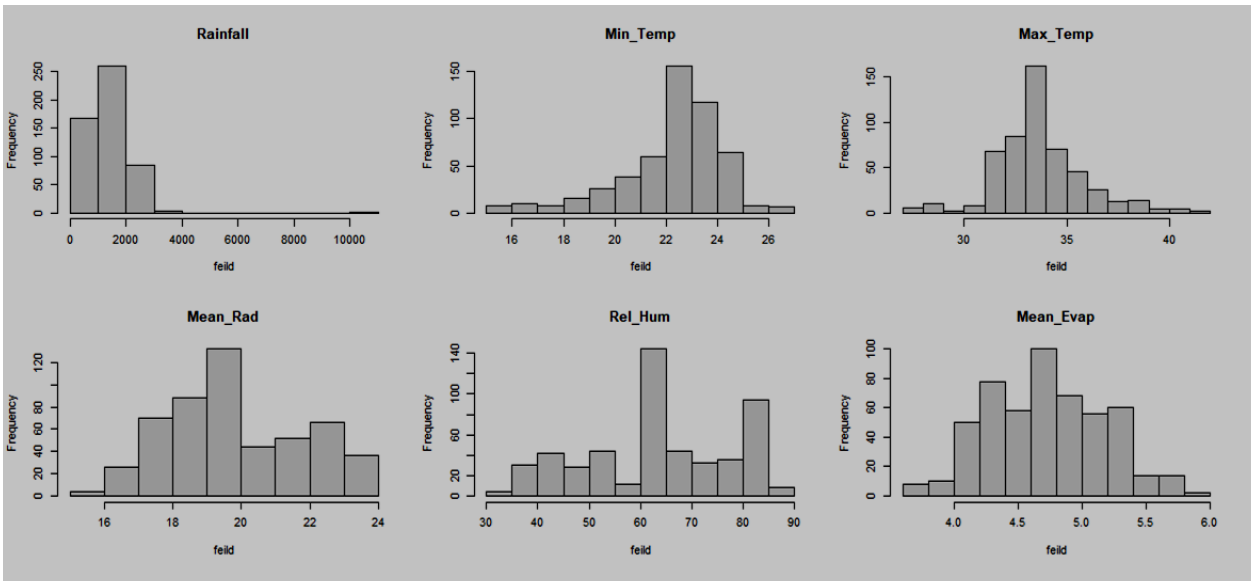

Figure 3: Histogram Plots for Each Independent Variable

Figure 3 above shows the histogram plots for each predictor. The histogram plots show virtually all the distributions have bimodal shapes, typically indicating deviation from normality. We employed the density plots to get a smoother look at the distribution.

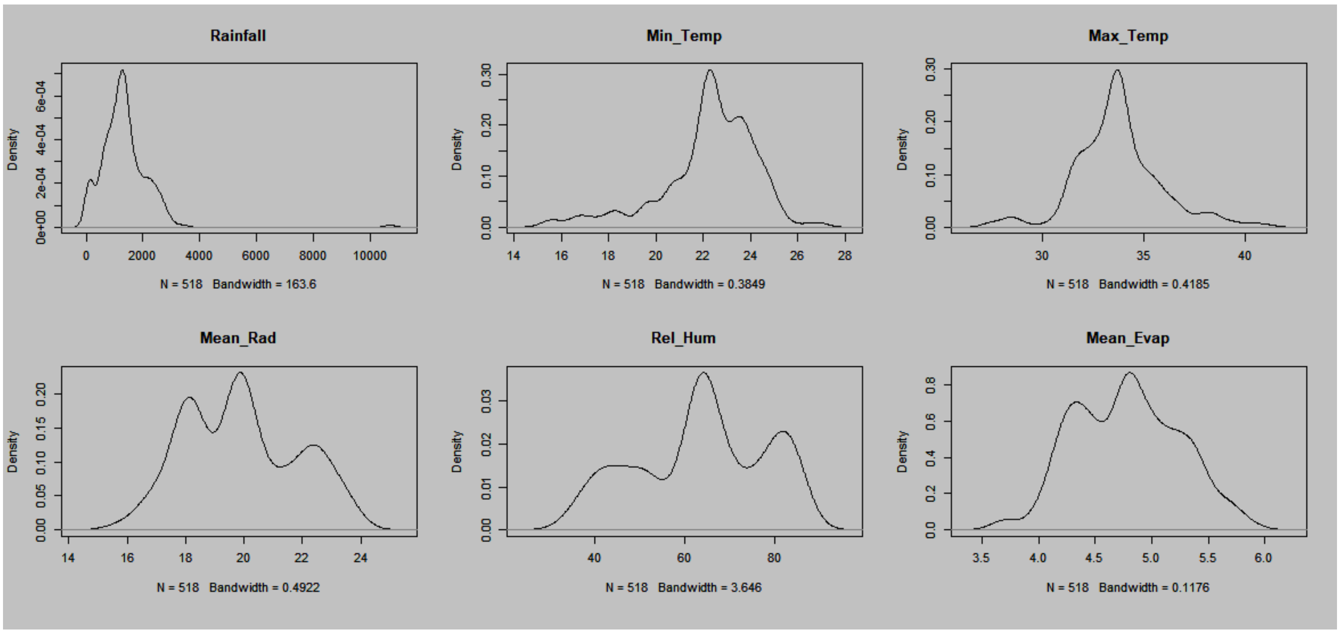

Figure 4: Density plots for each independent variable

Figure 4 shows the density plots for each of the independent variables. It shows that they all have multi-modal behaviours.

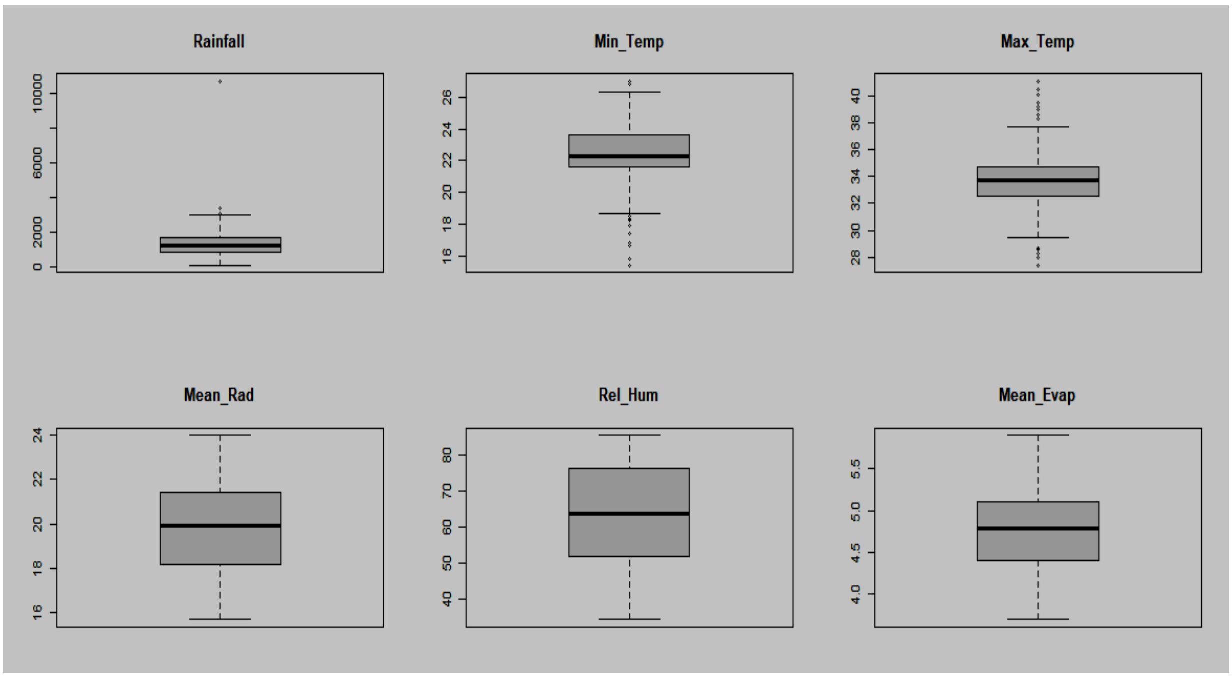

Figure 5: Boxplots for the Predictor Variables

Figure 5 shows the boxplots showing the distribution of the predictor variables. It shows that rainfall, min temperature and maximum temperatures have outliers which shows the deviation of the dataset from normality.

4.2 Multi-modal Data Presentation and Visualisations

The research data of intercorrelation coefficients between independent variables is shown in Table2. The other data set representing scattered matrix plots by flooding is shown in Figure 6 below.

Table 2: Intercorrelation coefficients between independent variables

	Rainfall	MinTemp	MaxTemp	MeanRad	RelHum	MeanEvap
Rainfall	1					
MinTemp	.236**	1				
MaxTemp	-.121**	.349**	1			
MeanRad	-.480**	-.438**	.320**	1		
RelHum	.463**	.462**	-.300**	-.840**	1	
MeanEvap	-.474**	-.318**	.400**	.866**	-.747**	1
**. Correlation is significant at the 0.01 level (2-tailed).						

Table 2 shows the intercorrelation coefficients between independent predictors. It shows that most of the variables have weak relationships. However, there were high correlations for Mean Radiation and Relative Humidity, Mean radiation and Mean evaporation, and Relative humidity and mean evaporation.

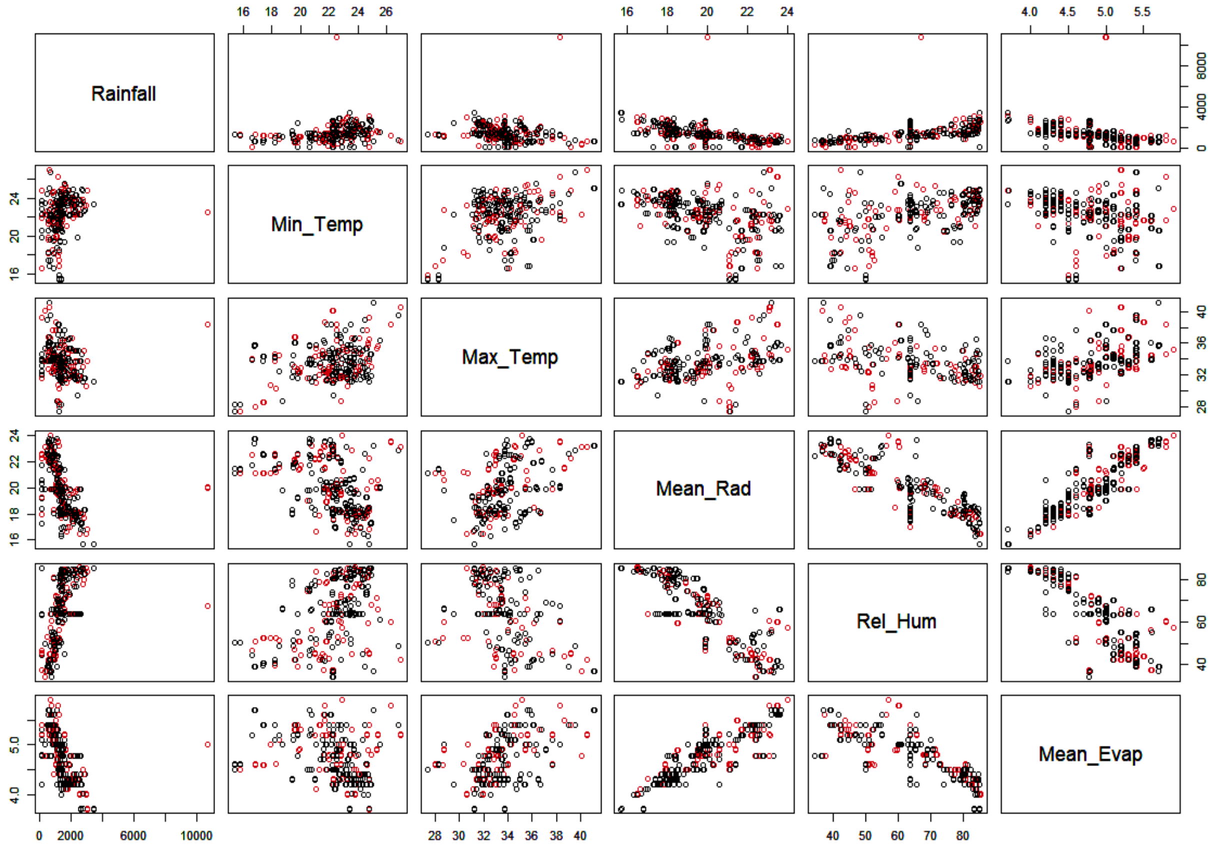

Figure 6: Scattered Matrix Plot by Flooding

Figure 6 shows the scattered matrix plot by flooding. It offers specific positive and negative linear correlations between the predictors, while rainfall didn't establish any relationships with other independent variables.

4.3 Evaluate Algorithms: Baseline

There is no prior knowledge of the performance of the different machine algorithms on the dataset. So, a spot-check on other methods was considered. We commenced this check by looking at linear and non-linear algorithms:

1. Linear Algorithms: Logistic Regression (LG), Linear Discriminate Analysis (LDA) and Regularized Logistic Regression (GLMNET).
2. Non-Linear Algorithms: k-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Naive Bayes (NB) and Support Vector Machines with Radial Basis Functions (SVM).

We have a good amount of data, so we used10-fold cross validation with three repeats. This is an excellent standard test harness configuration. The dataset's outcome variable suggests we are dealing with a binary classification problem. We used Accuracy, Kappa Metrics, and ROC to select the best Machine Learning algorithms. In creating our fitting models, we used the default parameters without transformation, introduced a box-cox change and advanced to better performance using the Ensemble algorithms. For each algorithm, the random number seed is reset before training to ensure that each algorithm is evaluated on the same data splits.

4.4 Untransformed Analysis of Machine Learning Algorithms

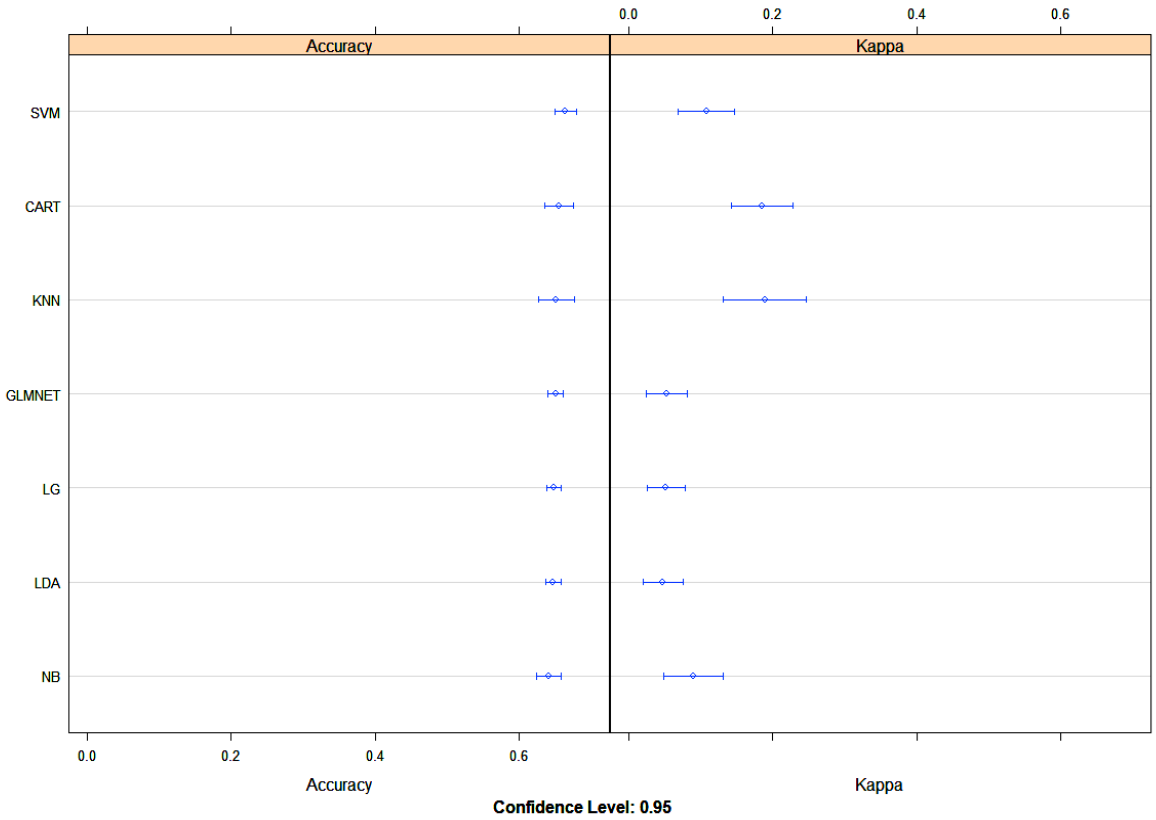

Figure 7: Performance chart of the Untransformed Analysis of Machine Learning Algorithms

We can see fair accuracy across the board. All algorithms have a mean accuracy above 60%, well above the baseline of 34.5% if we just predicted flood. This implies that the problem is learnable. We can see that KNN, CART, and SVM had the highest accuracy on the problem.

4.5 Transformed Analysis of Machine Learning Algorithms

From the unimodal visualisation, we saw that our predictors had skewed distributions. Hence, a transformation must be applied to adjust and normalise these distributions. Therefore, we used a transformation favouring positive input attributes, as in our case. The Box-Cox transformation was applied.

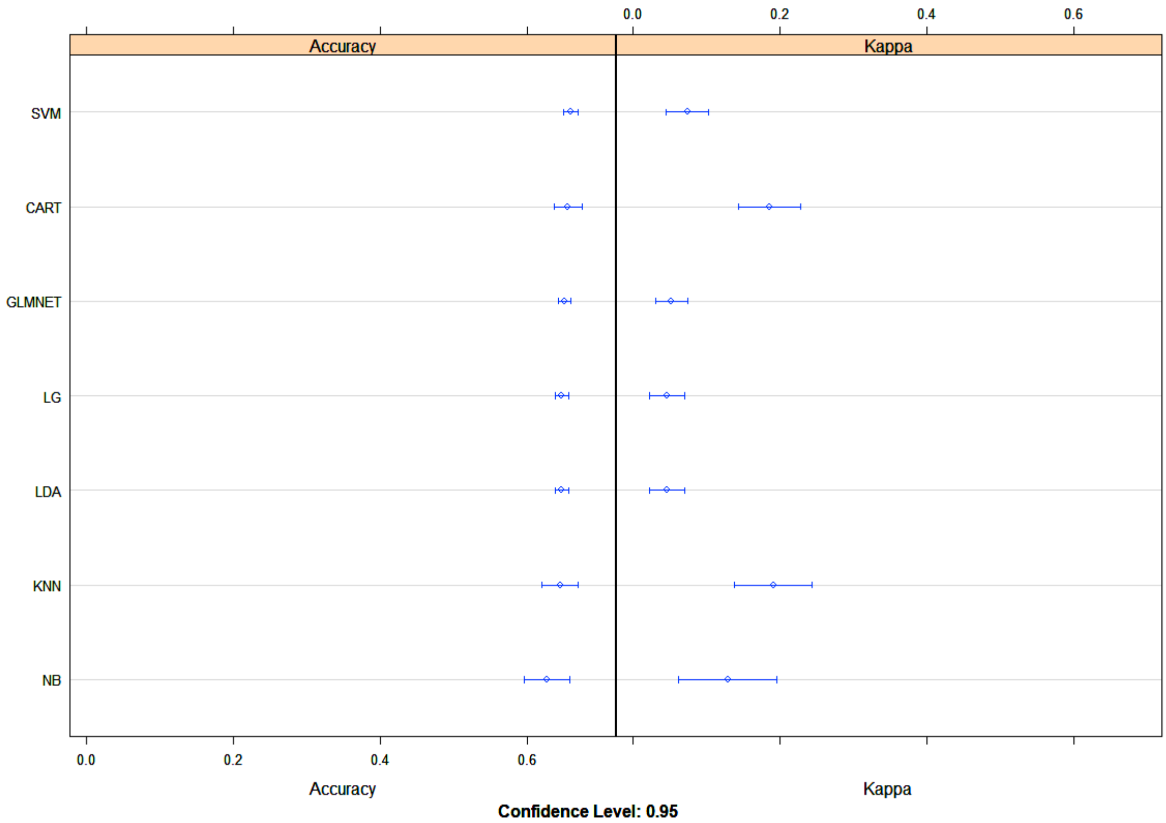

Figure 8: The performance chart of different algorithms on applied on the transformed data set

Figure 8 shows the performance of the different machine learning algorithms on the application of transformation. The performance did not improve significantly compared to the initial results; hence, the need for higher performance algorithms such as Ensemble algorithms.

4.6 Ensemble Methods

In this paper, five (5) ensemble algorithms were used, and they are classified below

- Bagging: Bagged CART (BAG) and Random Forest (RF).
- Boosting: Stochastic Gradient Boosting (GBM), XG Boost and C5.0 (C50).

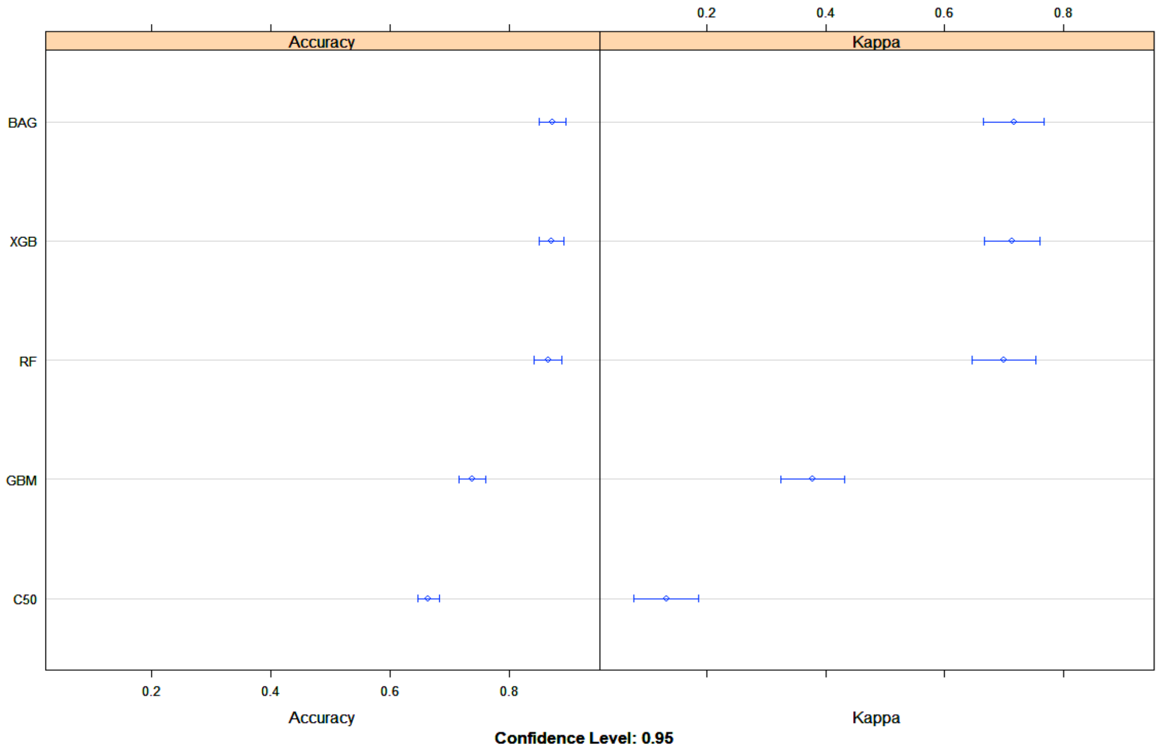

Figure 9: Ensemble Algorithms performance chart

Figure 9 shows the performance of the different ensemble algorithms. Random forest and BAG were the algorithms with the highest accuracy, and Kappa indicated better performance than the other three models (i.e. GBM, XGB and C5.0).

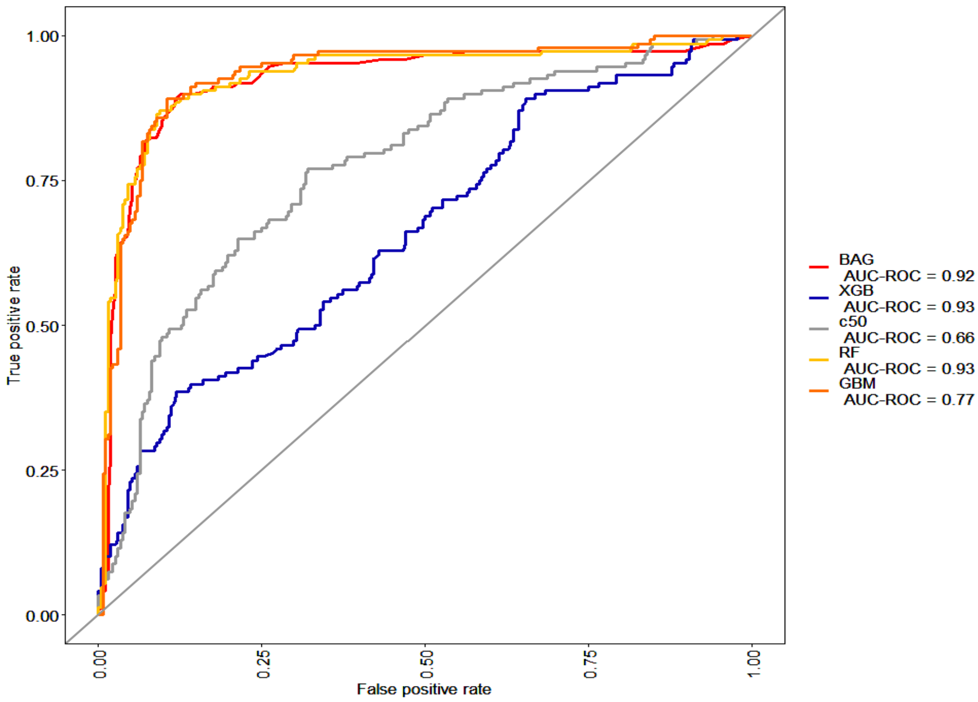

Figure 10: ROC Curve Assessing the Performance of the Different Ensemble Algorithms.

Figure 10 shows the ROC assessing the performance of the different Ensemble algorithms. RF and XGB gave the best Area Under the Curve (AUC), followed by BAG. This shows that RG and XGB performed better than other models trained in this study (i.e. GBM, XGB Boost and C5.0).

Table 4: Comparison of ensemble model performances

	Bagging		Boosting		
	RF	**BAG**	**C5.0**	**XGB**	**GBM**
Accuracy	0.91	0.89	0.68	0.91	0.78
No Information Rate	0.65	0.65	0.65	0.65	0.65
P-Value [Acc> NIR]	0.00	0.00	0.30	0.00	0.00
Kappa	0.80	0.75	0.20	0.80	0.50
Mcnemar's Test P-Value	0.05	0.02	0.00	0.05	0.06
Sensitivity	0.98	0.98	0.89	0.98	0.91
Specificity	0.78	0.72	0.28	0.78	0.56
PosPred Value	0.89	0.87	0.69	0.89	0.79
Neg Pred Value	0.97	0.96	0.59	0.97	0.77
Prevalence	0.65	0.65	0.65	0.65	0.65
Detection Rate	0.64	0.64	0.58	0.64	0.59
Detection Prevalence	0.72	0.74	0.83	0.72	0.75
Balanced Accuracy	0.88	0.85	0.59	0.88	0.73

Table 4 shows the Random Forest and the XGboost algorithms' performance based on the test dataset with an accuracy of 91%, a sensitivity of 98%, and a specificity of 78%.

5. Conclusion and further studies

In this study, we have gone through the process of predicting floods in Nigeria using ensemble machine-learning methods. The study employed linear and non-linear machine learning models to ascertain performance in the classification problem and further advanced the models using some ensemble algorithms. Comparing the models' performances showed that the ensemble algorithms performed better than the conventional machine learning algorithms**.** The random forest and BAG performed best in the training datasets from the different ensemble models based on their higher accuracy and Kappa. In contrast, the Random forest and the XGboost algorithms performed better on testing with the test dataset based on their accuracy, specificity, and sensitivity.

There is a need for NIMET and other agencies in Nigeria to improve their online data portal to be made readily available for researchers. This will make researchers faster and assist with better weather predictions and disaster predictions related to weather like floods. One way to improve their online system is to capture a daily log of meteorological changes. Daily capturing will go a long way to improve our machine learning algorithms with better performance since logging data generates a more significant mass of datasets. Nevertheless, further study is required to incorporate and operationalise the high-performance algorithms Random forest, BAG, and XGboost as early flood warning systems. Several factors need to be thought about here. The first benefit is that a stochastic input may be utilised to estimate the probabilistic distribution across flood quantities, given the relatively short run time. Second, less severe than historical occurrences but still causing floods, precipitation projections should be used to assess the models further. If such precipitation occurrences are considered, the trained algorithms' tendency to overreact to very few swings may impair their effectiveness. The outcome variable in this study was a binary categorical variable which further analysis can adopt a count variable that follows a Poisson distribution. This time instead of predicting the incidence of the flood, it will be predicting the number of storms that occur in those states per year. The dataset in this study is more like time series data. Hence further work can look at machine learning algorithms that work with time series data better to capture the role of years in the model.

References

Akinyokun, O. C., Inyang, U. G., & Akpan, E. E. (2020). A Hybrid Machine Learning Approach for Flood Risk Assessment and Classification. International Journal of Computational Intelligence and Applications. Vol. 19, No. 02, 2050012

Ardabili S, Mosavi A, Várkonyi-Kóczy A.R. (2019). Advances in machine learning modeling reviewing hybrid and ensemble methods. International Conference on Global Research and Education 2019 18 2020 (pp. 215-227). Springer International Publishing.

Egbinola, C.N., Olaniran, H.D., and Amanambu A.C. (2017). Flood management in cities of developing countries: the example of Ibadan, Nigeria, Journal of Flood Risk Management, 10(4), 546-554.

Huo W, Li Z, Zhang K, Wang J, Yao C. GA-PIC: An improved Green-Ampt rainfall-runoff model with a physically based infiltration distribution curve for semi-arid basins. Journal of Hydrology. 2020 Jul 1;586:124900.

Isaac K.F., Owusu N. B., Samuel B., Faiza U.B., Promise A., Nicodemus S.A., Peter N., Adebayo F. A., Benjamin A. W. and Vivian A.A. (2021) Enhancing flood Prediction using Ensemble and Deep Learning Techniques. 2021 22nd International Arab Conference on Information Technology (ACIT) 978-1-6654-1995-6/21/S31 DOI:10.1109/ACIT53391.2021.9677084

Li W., Kiaghadi, A. & Dawson, C. (2021). Exploring the best sequence LSTM modelling

architecture for flood prediction. Neural Computing & applications. 33, 5571-5580. https://doi.org/10.1007/s00521-020-05334-3

Maspo, N. A., Harun, A. N. B., Goto, M., Cheros, F., Haron, N. A., &Nawi, M. N. M. (2020, June). Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. In *IOP Conference Series: Earth and Environmental Science, 479, 012038.* IOP Publishing.

Mosaffa, H., Sadeghi, M., Mallakpour, I., Jahromi, M. N., &Pourghasemi, H. R. (2022). Application of machine learning algorithms in hydrology. In: Pourghasemi et al. (eds), *Computers in Earth and Environmental Sciences. Artificial Intelligence and Advanced Technologies in Hazards and Risk Management* (pp. 585-591).

Mosavi A., Oztturk P. Chau K., (2018). Flood Prediction using Machine Learning models: literature review. Water(Switzerland), Vol. 10, no. 11, pp 1-40. DOI: 10.3390/w1011536

Mosavi, A., Bathla, Y., &Varkonyi-Koczy, A. (2017). Predicting the Future Using Web Knowledge: State of the Art Survey. In Recent Advances in Technology Research and Education (pp. 341-349). Springer: Cham, Switzerland.

Panahi, M., Dodangeh, E., Rezaie, F., Khosravi, K., Van Le, H., Lee, M. J., ... & Pham, B. T. (2021). Flood spatial prediction modeling using a hybrid of meta-optimisation and support vector regression modelling. Catena, 199 105114.https://doi.org/1016/j.catena.2020.105114

Pham, B. T., Jaafari, A., Van Phong, T., Yen, H. P. H., Tuyen, T. T., Van Luong, V., & Foong, L. K. (2021). Improved flood susceptibility mapping using a best first decision tree integrated with ensemble learning techniques. *Geoscience Frontiers*, *12*(3), 101105.https://doi.org/10.1016/j.gsf.2020.11.003.

Shahabi H., Shirzadi A. Ronoud S. Asadi S., Pham B.T., Mansouripour F, Geertsema M, Clague J.J. (2021) Flash flood succeptibilityb mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. Geoscience Frontiers

Tehrany, M.S. Jones, S.,Shabani, F. (2019). Identifying the essential Flood Conditioning

Factors for Flood Prone Area Mapping using Machine Learning Techniques. Catena 2019, 175, 174-192

Xia, X.; Liang, Q.; Ming, X.; Hou, J. (2017). An efficient and stable hydrodynamic model with novel source term discretisation schemes for overland flow and flood simulations. *Water Resources. Research,53*, 3730–3759.