

Application of Machine Learning for Flood Prediction and Evaluation in Southern Nigeria

Emeka Bright Ogbuene^{1*}, Chukwumeche Ambrose Eze¹, Obianuju Getrude Aloh², Andrew Monday Oroke³, Damian Onuora Udegbonam¹, Josiah Chukwuemeka Ogbuka¹, Fred Emeka Achoru¹, Vivian Amarachi Ozorme¹, Obianuju Anwara¹, Ikechukwu Chukwunonyelum¹, Anthonia Nneka Nebo¹, Obiageli Jacinta Okolo¹

¹Centre for Environmental Management and Control (CEMAC), University of Nigeria, Enugu, Nigeria

²Department of Geography, Enugu State University of Science and Technology, ESUT, Enugu, Nigeria

³School of Civil Engineering, Newcastle University, Newcastle upon Tyne, UK

Email: *ogbuene.emeka@unn.edu.ng

How to cite this paper: Ogbuene, E.B., Eze, C.A., Aloh, O.G., Oroke, A.M., Udegbonam, D.O., Ogbuka, J.C., Achoru, F.E., Ozorme, V.A., Anwara, O., Chukwunonye, I., Nebo, A.N. and Okolo, O.J. (2024) Application of Machine Learning for Flood Prediction and Evaluation in Southern Nigeria. *Atmospheric and Climate Sciences*, **14**, 299-316.

<https://doi.org/10.4236/acs.2024.143019>

Received: March 20, 2024

Accepted: June 30, 2024

Published: July 3, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study explored the application of machine learning techniques for flood prediction and analysis in southern Nigeria. Machine learning is an artificial intelligence technique that uses computer-based instructions to analyze and transform data into useful information to enable systems to make predictions. Traditional methods of flood prediction and analysis often fall short of providing accurate and timely information for effective disaster management. More so, numerical forecasting of flood disasters in the 19th century is not very accurate due to its inability to simplify complex atmospheric dynamics into simple equations. Here, we used Machine learning (ML) techniques including Random Forest (RF), Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), and Neural Networks (NN) to model the complex physical processes that cause floods. The dataset contains 59 cases with the goal feature “Event-Type”, including 39 cases of floods and 20 cases of flood/rainstorms. Based on comparison of assessment metrics from models created using historical records, the result shows that NB performed better than all other techniques, followed by RF. The developed model can be used to predict the frequency of flood incidents. The majority of flood scenarios demonstrate that the event poses a significant risk to people’s lives. Therefore, each of the emergency response elements requires adequate knowledge of the flood incidences, continuous early warning service and accurate prediction model. This study can expand knowledge and research on flood predictive modeling in vulnerable areas to inform effective and sustainable contingency planning, policy, and management actions on flood disaster incidents, especially in other technologically underdeveloped settings.

Keywords

Machine Learning, Flood, Prediction, Evaluation, Southern Nigeria

1. Introduction

Southern Nigeria region faces several challenges relating to accurate predictions and analysis of flood scenarios. Floods are a recurring natural disaster in the region, causing significant damage to infrastructure, loss of lives, and disruption of livelihoods. These challenges include the complex nature of weather patterns, inadequate historical data, limited resources for monitoring and early warning systems, and the need for localized predictions due to variations in terrain and land use. Traditional methods of flood prediction and analysis often fall short of providing accurate and timely information for effective disaster management [1]. It has been reported that numerical forecasting of flood disasters in the 19th century lacked accuracy due to its inability to simplify complex atmospheric dynamics into simple equations [2]. Although, the nonlinear modeling capability of Artificial Neural Networks (ANNs) has been used in developing nonlinear predictive models for weather analysis with the ANN approach [3] [4], it has shown limited effectiveness in accuracy and timeliness. The critical challenge in flood disasters in the south-south of Nigeria includes poor attention to flood modeling and assessing vulnerability to flooding. Therefore, there is a need for novelty in knowledge on machine learning (ML) model building of flood prediction. Machine learning (ML) offers a promising approach to address this challenge by leveraging historical data, weather patterns, topographical information, and other relevant factors to develop predictive models for flood occurrences. The application of machine learning for flood prediction and analysis in Southern Nigeria has become an increasingly important area of research due to the region's vulnerability to flooding. The review paper of [5] introduces the most promising prediction methods for both long-term and short-term floods. Furthermore, the major trends in improving the quality of the flood prediction models are investigated. Among them, hybridization, data decomposition, algorithm ensemble, and model optimization are reported as the most effective strategies for the improvement of ML methods. The report of [6] gives insight into the mechanism of the Non-linear (NARX) and Support Vector Machine (SVM) machine learning algorithm from the perspective of flood estimation. Furthermore, to evaluate the link between flood incidence and the fifteen (15) explanatory variables, which include climatic, topographic, land use and proximity information, [7] used artificial neural network (ANN) and logistic regression (LR) models were trained and tested to develop a flood susceptibility map.

However, much research on the application of ML techniques is reviewed works that do not encompass most of the ML algorithms in one study. Hence, the current study seeks to apply five ML algorithms such as SVM, Random For-

est (RF), Logistic Regression (LR), Naïve Bayes (NB), and Artificial Neural Networks (ANN) for flood prediction and evaluation in Nigeria's southern region.

2. Materials and Methods

The main focus of this study is the application of ML to predict and evaluate flooding based on the flood type, location, duration, begin/end location, begin/end latitude and longitude, injuries direct/indirect, death direct/indirect and property and crop damage. The proposed method uses historical information collected from 1999 to 2019, to learn the patterns and changes in various parameters' behavior in flood events and make remarks for future events. (**Figure 1**)

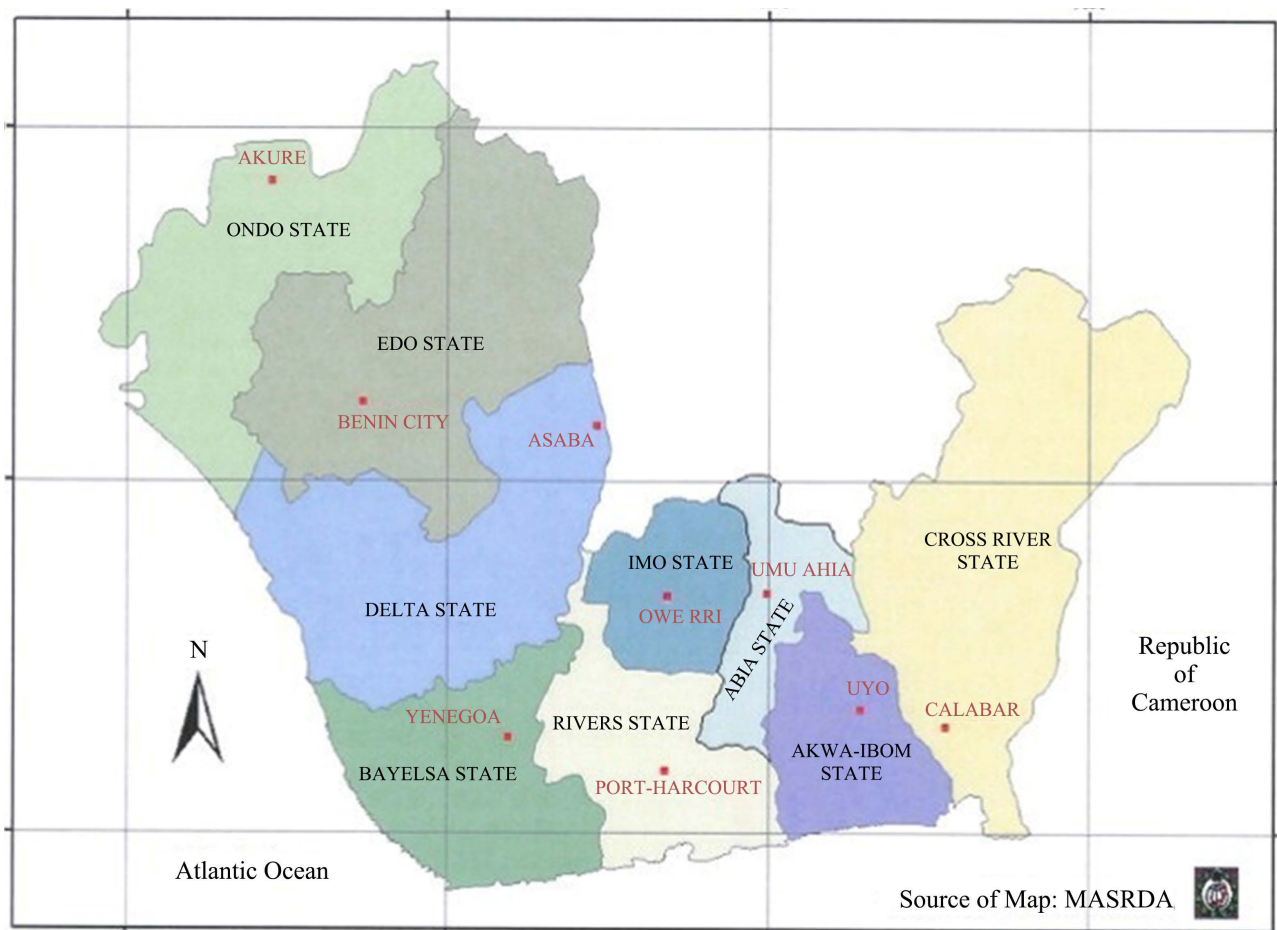


Figure 1. Map of the study area (Source: [8]).

2.1. Data Collection and Pre-Processing

One of the most important requirements for this research was a detailed historical and inclusive data set, which was acquired from the National Emergency Management Agency (NEMA), National Oceanic and Atmospheric Administration (NOAA) and the National Climatic Data Centre (NCDC) [9] [10]. The data used in this study covers the period from 1999 until 2019. The data collection sub-task is the process of identifying, extracting, and integrating log data from

the source systems into a single repository. However, preprocessing is required to reduce the size of the dataset and transform it into a sliding window representation. Feature selection, the process of identifying a set of features from the data to be used in machine learning, is only performed for initial training and evaluation of the model. Therefore, the flood data was collected from different sources such as the National Emergency Management Agency (NEMA) and other publications. The details of the different sources, data of event, event type and references are given in **Table 1**.

Table 1. Data set for flood disaster inventory.

| Period | Contents | Data Type | References |
|-----------|---|----------------------|---|
| 1999-2002 | Causes and consequences of flooding in Nigeria | Field data-Numerical | [11] |
| 2002-2004 | 8 states are under red alert 50 LGAs affected | Field data-Numerical | [12] |
| 2004-2006 | Disaster Profile-Type of hazards, location-Detailed impact on population, GDP | Field data-Numerical | [13] |
| 2006-2009 | Climate Change and Menace of Floods in Nigerian Cities: Socio-economic Implications | Field data-Numerical | [14] |
| 2009-2011 | The Devastating Effect of Flooding in Nigeria | | [15] |
| 2011-2016 | Flood risk management in Nigeria | | [16] |
| 2016-2018 | Flooding conceptual review | | [17] |
| 2018-2019 | News situation tracking-Nigeria flood disaster update in Nigeria | | https://www.premiumtimesng.com/news/headlines/331715-hunger-rainstorm-kill-11-villagers-after-forced-evacuation-by- |

2.2. Data Pre-Processing

Data transformation operations are used to convert the dataset into an appropriate structure to facilitate machine learning. However, data aggregation and feature selection are common data transformation techniques used to obtain a reduced representation of the dataset without impacting its predictive accuracy [18]. Data pre-processing is required to transform the data into a format usable by machine learning algorithms. The data sets collated were inspected for outliers and extreme values, missing data and redundant information via a bespoke MATLAB application known as a data cleaning tool. This tool removes all existing outliers and missing data and re-orders the data based on specific categories chosen for the implementation of the ML techniques and it converts the alphanumeric and alphabetic data to numeric data using one-hot encoding. The processed dataset is then divided into training and testing data sets. The training data set is used to develop the model whereas the testing data set is used to quantify the accuracy of the model built. A larger portion of data is separated for training and the remaining is used for testing and validation to ensure accuracy of the classification model built and software performance. **Figure 2** shows an

overview of the overall analytical process employed in this study. The raw data collected is fed to the MATLAB data cleaning tool for data cleaning, normalization, aggregation, and other pre-processing steps. The output data is divided into testing and training data and passed through the ML/data mining application, the patterns are extracted, and the model is built, followed by analysis to verify its quality.

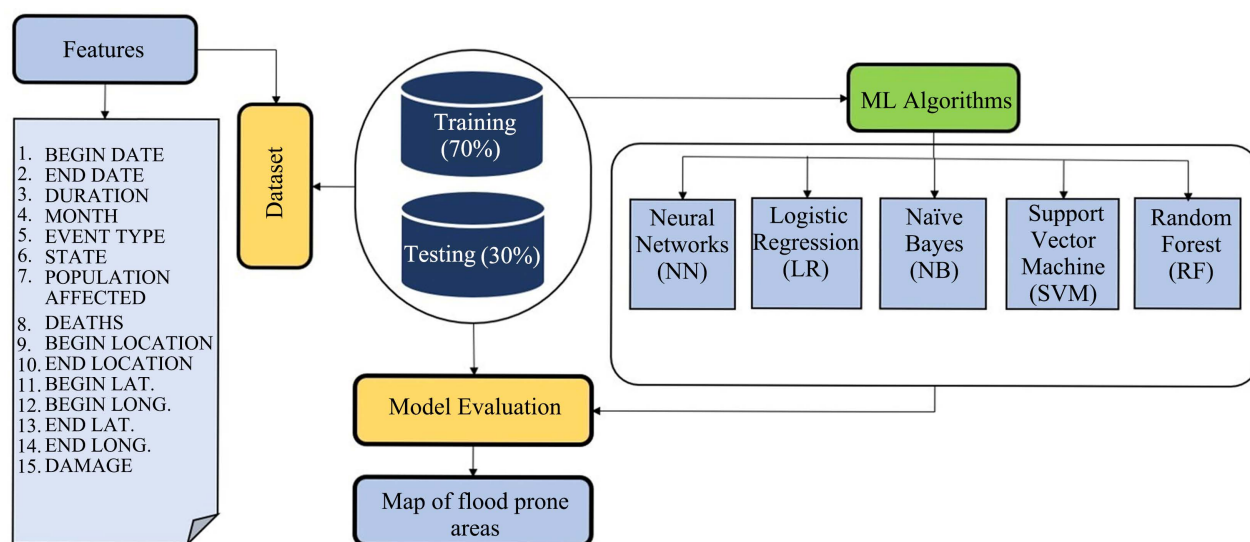


Figure 2. Schematics of ML methodology.

2.3. Machine Learning Techniques

This study focused on supervised ML to learn from historical data, find clustered data, and build a classification model for future events. This type of ML works particularly best when used in combination with historical data (results included). For this purpose, several data mining tools such as orange canvas have been deployed. The reason for using the two software is to test more ML techniques with various training and testing dataset sizes. This software is user-friendly and can be easily accessible. The data will be divided into two parts. The first will be used for training and generating the model, and the second will be used for testing and verification. Several models were developed using different ML techniques to be able to measure and compare their performance and accuracy and choose the best. These techniques included Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and Logistic Regression (LR). The class for the model in all cases was set as “event type”, which included flood, flood/rainstorm, and flood/Windstorm. The independent attributes in all models were: location (community), state, population affected, injuries direct, injuries indirect, death direct, death indirect, property damage and crop damage.

2.4. Orange Data Mining Software

Orange data mining software was originally developed by scientists at the Uni-

versity of Ljubljana in 1997 using the Python, Cython, C++ and C programming languages. The software's graphical environment and interfaces have been developed using the Python and Qt3 libraries [19]. It opens commonly used dataset extensions such as txt, basket, CSV, arff. or Excel spreadsheet format. The method allowed input of climatic data such as rainfall variables (rainfall amount, intensity, duration, magnitude), it may also involve relative humidity, percentage relative humidity among others. The data could be uploaded and processed in the Orange Canvas software. This enables accurate prediction of like flood hazard over a long run (see **Table 2**).

Table 2. Nature of flood data input and affected population across the Southern Nigerian State.

| 1 | Begin date | End date | Duration | Duration month | Event type | State | Population affected | Begin location | End location | Begin lat. |
|----|------------|------------|----------|--|----------------------|----------------|---------------------|----------------|--------------|------------|
| 2 | 3/4/2001 | 3/30/2001 | 26 | March | Flood/ Rain storm | Edo | 820 | Esan west | Esan central | 6.66166 |
| 3 | 3/7/2012 | 3/23/2012 | 16 | March | Flood/ Rain storm | Edo | 0 | Lkpoba | Okha | 6.16445 |
| 4 | 3/9/ 1999 | 3/22/1999 | 13 | March | Flood/ Rain storm | Delta | 425,839 | Ugheli | Effrun | 5.48956 |
| 5 | 4/11/2001 | 4/30/2001 | 19 | April | Flood | Bayelsa | 0 | Patani | Patani | 5.22885 |
| 6 | 3/15/1999 | 3/21/1999 | 6 | March | Flood | Bayelsa | 0 | Yenagoa | Patani | 4.92675 |
| 7 | 3/6/2001 | 3/23/2001 | 17 | March | Flood/ Rain storm | Akwa lbom | 4000 | Lkom | Va la | 5.95666 |
| 8 | 3/8/2006 | 7/23/2006 | 135 | March, April, May, June, July | Flood/ Rain storm | Rivers | 350 | Opobo | Nkoro | 4.50607 |
| 9 | 3/1/2012 | 7/15/2012 | 127 | March, April, May, June, July | Flood/ Rain storm | Rivers | 500 | Ahoda | Mbiama | 5.08333 |
| 10 | 3/2/2013 | 7/25/2013 | 137 | March, April, May, June, July | Flood/ Rain storm | Rivers | 430 | Ahoda | Mbiama | 5.08333 |
| 11 | 3/6/2017 | 7/28/ 2017 | 140 | March, April, May, June, July | Flood/ Rain storm | Rivers | 301 | Ahoda | Mbiama | 5.08333 |
| 12 | 3/13/2017 | 3/27/2017 | 14 | March | Flood/ Rain storm | Cross river | 25000 | Yala | Akamkpa | 6.58916 |
| 13 | 4/10/1999 | 9/7/1999 | 177 | April, May, June, July, August, September | Flood | Akwa lbom | 1 | Lkom | Vala | 5.95666 |
| 14 | 4/13/1999 | 9/16/1999 | 183 | April, May, June, July, August, September | Flood | Delta | 1 | Ugheli | Warri | 5.48956 |

Continued

| | | | | | | | | | | |
|----|-----------|-----------|-----|--|----------------------|----------------|---|---------------|---------------|---------|
| 15 | 4/18/1999 | 9/8/1999 | 170 | April, May, June, July, August, September | Flood | Bayelsa | 1 | Yenagoa | Patani | 4.92675 |
| 16 | 4/11/1999 | 9/23/1999 | 192 | April, May, June, July, August, September | Flood | Edo | 1 | Oredo | Egor | 6.23581 |
| 17 | 6/2/2004 | 7/8/2004 | 36 | June, July | Flood | Edo | 0 | Ostacocentral | Ostacocentral | 9.07775 |
| 18 | 6/10/2004 | 6/13/2004 | 3 | June | Flood/ Rain storm | Rivers | 0 | Opobo | Nkoro | 4.50607 |
| 19 | 8/26/2004 | 9/2/2004 | 6 | August, September | Flood | Delta | 0 | Ugheli | Sapele | 5.48956 |
| 20 | 2/16/2005 | 3/3/2005 | 17 | February, March | Flood/ Rain storm | Cross river | 0 | Lkom | Va la | 5.95666 |
| 21 | 7/5/2005 | 8/2/2005 | 27 | July, August | Flood | Edo | 0 | Oredo | Egor | 6.23581 |
| 22 | 9/24/2018 | 9/26/2018 | 2 | September | Flood | Bayelsa | 0 | Yenagoa | Patani | 4.92675 |
| 23 | 9/24/2018 | 9/26/2018 | 2 | September | Flood | Delta | 0 | Ugheli | Warr i | 5.48956 |

Table 2 shows the flood data and its behavior over the period of study, it describe flood beginning, end and disaster recorded across the study areas. It also shows event type and state affected with population estimate within the beginning of location and end. The result is robust enough and could be reliable.

2.5. ML Flood Prediction Model Evaluation

The system has been trained with several different combinations; however, the final system uses one based on the selected attributes, which was an output of the classifier attribute evaluation from an ML tool. All ML models developed were validated using evaluation criteria, *i.e.*, confusion matrix [20], Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [21]. These metrics are used for summarizing and assessing the quality of the ML model. A confusion matrix summarizes the classifier performance concerning the test data. It is a two-dimensional matrix, indexed in one dimension by the actual class of an object and in the other by the class that the classifier allocates, and the cells represent: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) identified in a classification. Multiple measures of accuracy are derived from the confusion matrix *i.e.*, specificity (SP), sensitivity (SS), positive estimated value (PPV) and negative estimated value (NPV). These are calculated as follows [22]:

$$SP = \frac{TN}{TN + FP} \quad (1)$$

$$SS = \frac{TP}{TP + FN} \quad (2)$$

$$PPV = \frac{TP}{TP + FP} \quad (3)$$

$$NPV = \frac{TN}{TN + FN} \quad (4)$$

The MAE is the mean of the absolute value of the error per instance over all samples in the test data. Each estimation error is the difference between the true value and the estimated value for the sample. MAE is calculated as follows [21]:

$$MAE = \frac{\sum_{i=1}^n |y_{\text{estimate},i} - y_{\text{actual},i}|}{n} \quad (5)$$

where, y_{actual} is the true value for the test sample i and $y_{\text{estimate},i}$ is the estimated or predicted value for the test sample i and n is the number of test samples.

The RMSE of a model for test data is the square root of the mean of the squared estimation errors over all samples in the test data. The estimation error is the difference between the true value and the estimated value for a sample. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |y_{\text{estimate},i} - y_{\text{actual},i}|^2}{n}} \quad (6)$$

Equations (1)-(6) were employed to validate the model, this step is also known as the model evaluation.

3. Results

There were 66 samples in the original data set. A MATLAB data cleaning application was used to remove outliers and filter the data, resulting in 59 occurrences with 14 features that could be used for learning. Afterwards, the data was split into two sections: a larger (75%), designated for training purposes, and a smaller (25%), designated for testing.

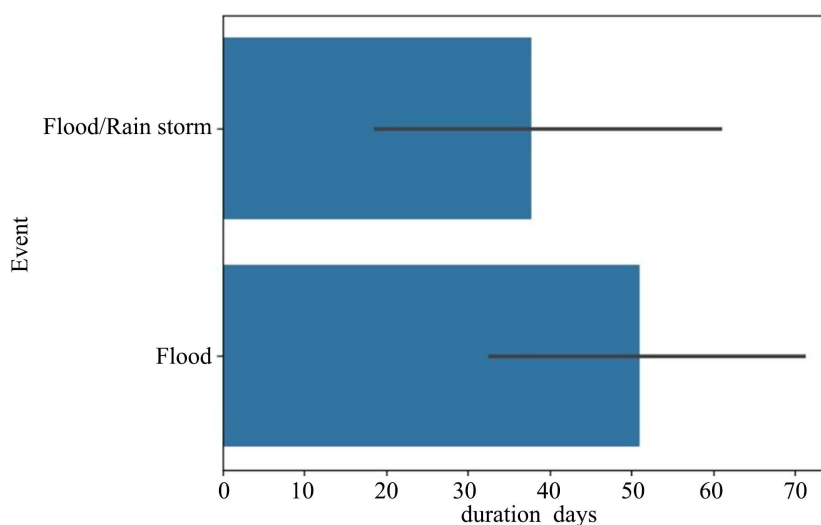
3.1. Descriptive Statistics of Flood Dataset

Descriptive statistics are used to summarize and describe the features of a dataset, providing insights into its central tendency, variability, and distribution. These statistics include measures such as mean, median, mode, standard deviations, and minimum and maximum values. **Table 3** presents the summary of the south-south flood historical dataset. It was observed that the maximum duration of the flood event is 192 days and a minimum of 2 days. Between the periods of 1999-2019, a total of 59 deaths were recorded.

In addition, a plot of the flood events which reflects the two classes “Flood”, and “Flood/Rainstorm” indicates that “Flood” occurred more than “Flood/Rainstorm” in terms of duration (**Figure 3**).

Table 3. Summary of features of the flood dataset from 1999-2019.

| | duration_days | affected_population | deaths | begin_lat | begin_long | end_lat | end_long |
|--------------|---------------|---------------------|-----------|-----------|------------|-----------|-----------|
| count | 59.000000 | 59.000000 | 59.000000 | 59.000000 | 59.000000 | 59.000000 | 59.000000 |
| mean | 46.525424 | 20836.779661 | 1.050847 | 5.570503 | 6.577166 | 6.022912 | 6.786401 |
| std | 59.207804 | 63583.33191 | 2.402755 | 0.839237 | 0.974734 | 1.280885 | 1.107064 |
| min | 2.000000 | 0.000000 | 0.000000 | 4.506070 | 5.551140 | 4.506070 | 5.575470 |
| 25% | 9.000000 | 0.500000 | 0.000000 | 4.926750 | 6.004070 | 5.062380 | 6.060160 |
| 50% | 19.000000 | 500.000000 | 0.000000 | 5.489560 | 6.191390 | 5.517370 | 6.267640 |
| 75% | 45.500000 | 4800.000000 | 0.000000 | 6.060555 | 6.650000 | 6.589160 | 7.871400 |
| max | 192.000000 | 425839.000000 | 12.000000 | 9.077750 | 8.706500 | 9.077750 | 8.677460 |

**Figure 3.** Summary of flood events within Nigeria's south-south zone from 1999-2019.

3.2. Flood Data Model Testing and Training

The 59 cases in the Orange Software test data have the target feature “Event-Type,” including 39 flood cases and 20 cases of flood combined with rainstorms. To determine which machine learning technique performs best, five various types of techniques are tested and evaluated in Orange Canvas software. The methods NN, LR, RF, NB, and SVM are tested. **Figures 6-10**, which show the model training and testing procedure implemented in Orange, provide an overview of the process. To create the classification models, the training data is first run through various classification techniques (NN, LR, RF, and NB). The models are then tested on the test data. The study revealed that RF and SVM outperformed all other methods in terms of the percentage of classifications on the test data. The evaluation results and confusion matrix for the various ML models based on the provided test set are displayed in **Figures 6-10**. The NB model categorized 8 out of 39 as flood and 18 out of 20 as flood/Rainstorm based on the confusion matrix (**Figure 4**).

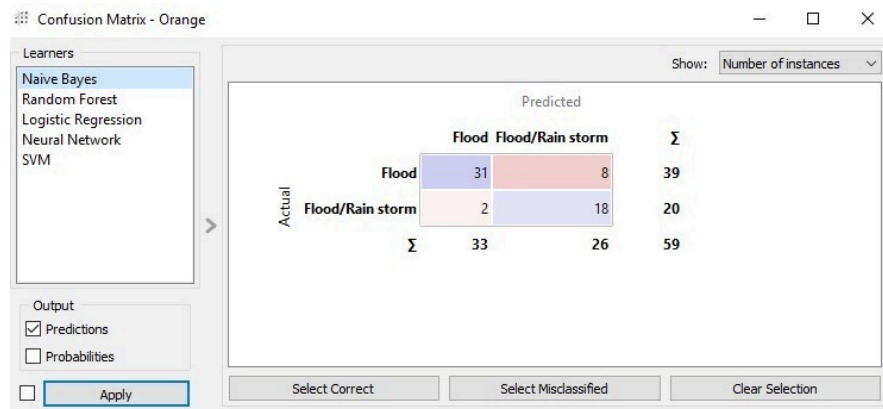


Figure 4. Confusion matrix for Naive Bayes classification.

In the RF model, it was classified 35 out of 39 instances as Flood, and 18 out of 20 as Flood/Rainstorm. The correctly classified instances in total were 59 (100%) (Figure 5).

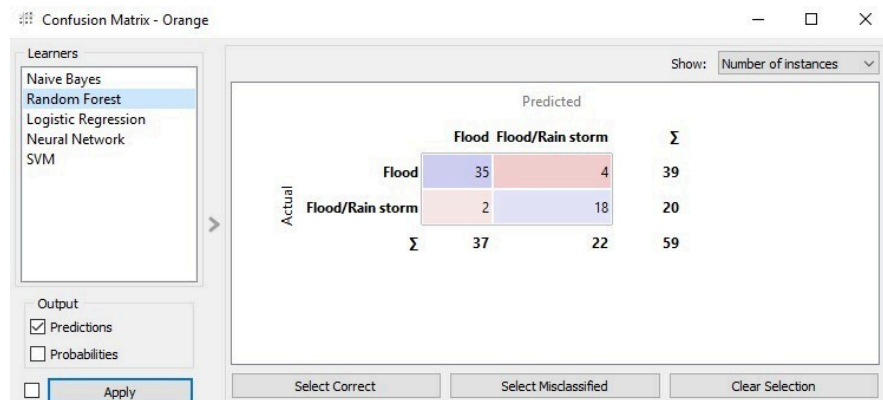


Figure 5. Confusion matrix for random forest classification.

In the LR model, it was classified that 39 out of 59 instances as Flood and 20 out of 20 as Flood/Rainstorm are correct. The correctly classified instances in total were 59 (100%) (Figure 6).

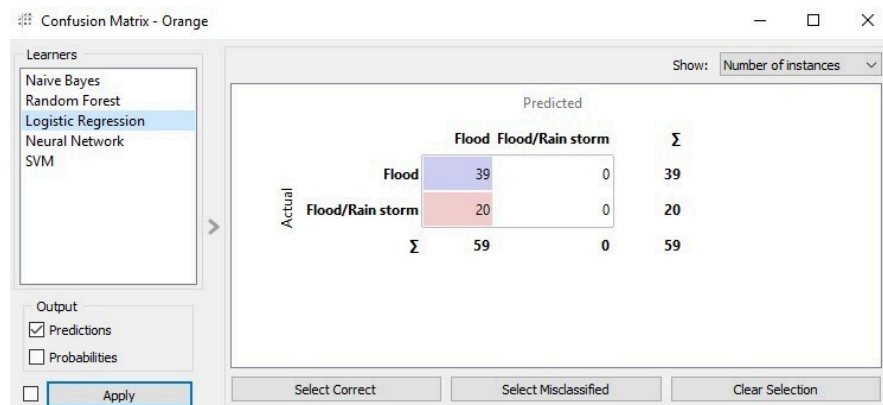


Figure 6. Confusion matrix for logistic regression.

In the NN model, it was classified that 39 out of 39 instances as Flood and 20 out of 20 as Flood/Rainstorm are correct (**Figure 7**).

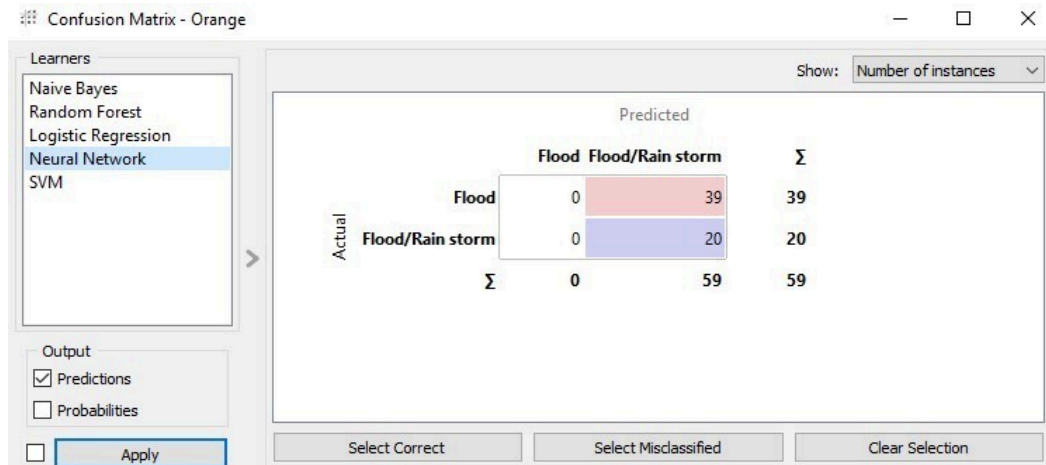


Figure 7. Confusion matrix for neural networks.

In the SVM model it was classified 39 out of 39 instances as Flood, 19 out of 20 were Flood/Rainstorm, and the correctly classified instances in total were 59 (100%) (**Figure 8**).

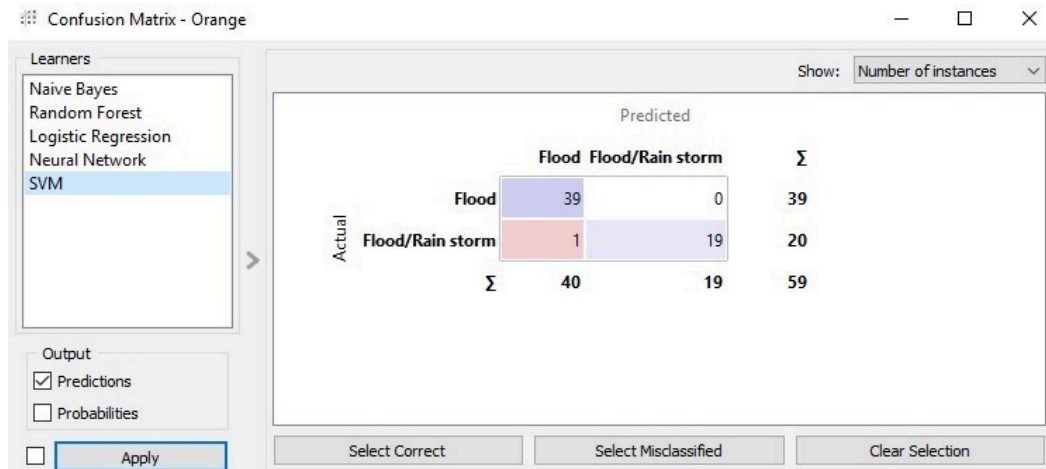


Figure 8. Confusion matrix for SVM classification.

The findings indicated that the south-south zone, particularly south-south settlements, is prone to flooding and rainstorms. The flood events at the starting and finishing locations are depicted in **Figure 9**. Despite this, the print maps' beginning and ending locations for the flood event do not significantly differ from one another (**Figure 9**).

The study area's maximum population affected by a flood occurrence is 50,000, as seen in **Figure 10**.

Figure 11 demonstrates that, in the south-south region, the range of deaths brought on by flood events is 0 to 5.

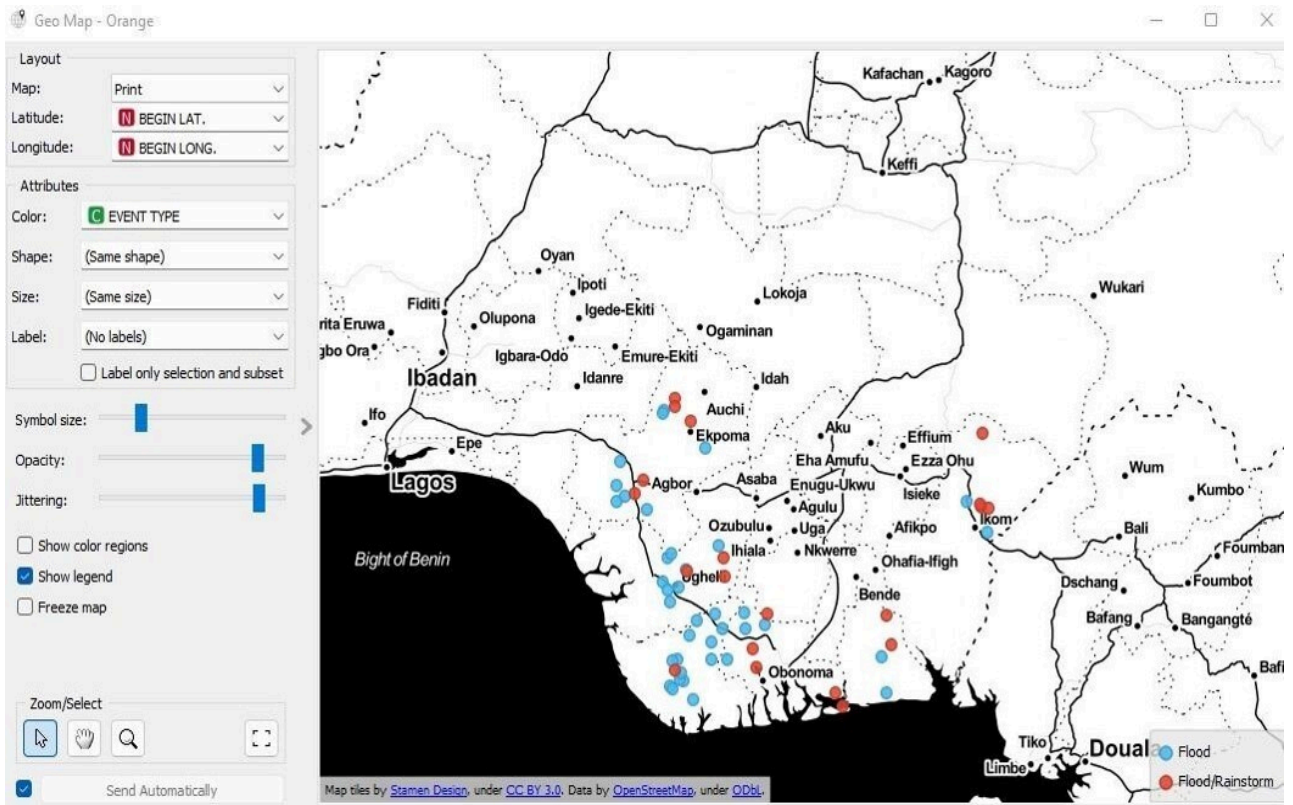


Figure 9. Flood patterns at different locations at the beginning locations.

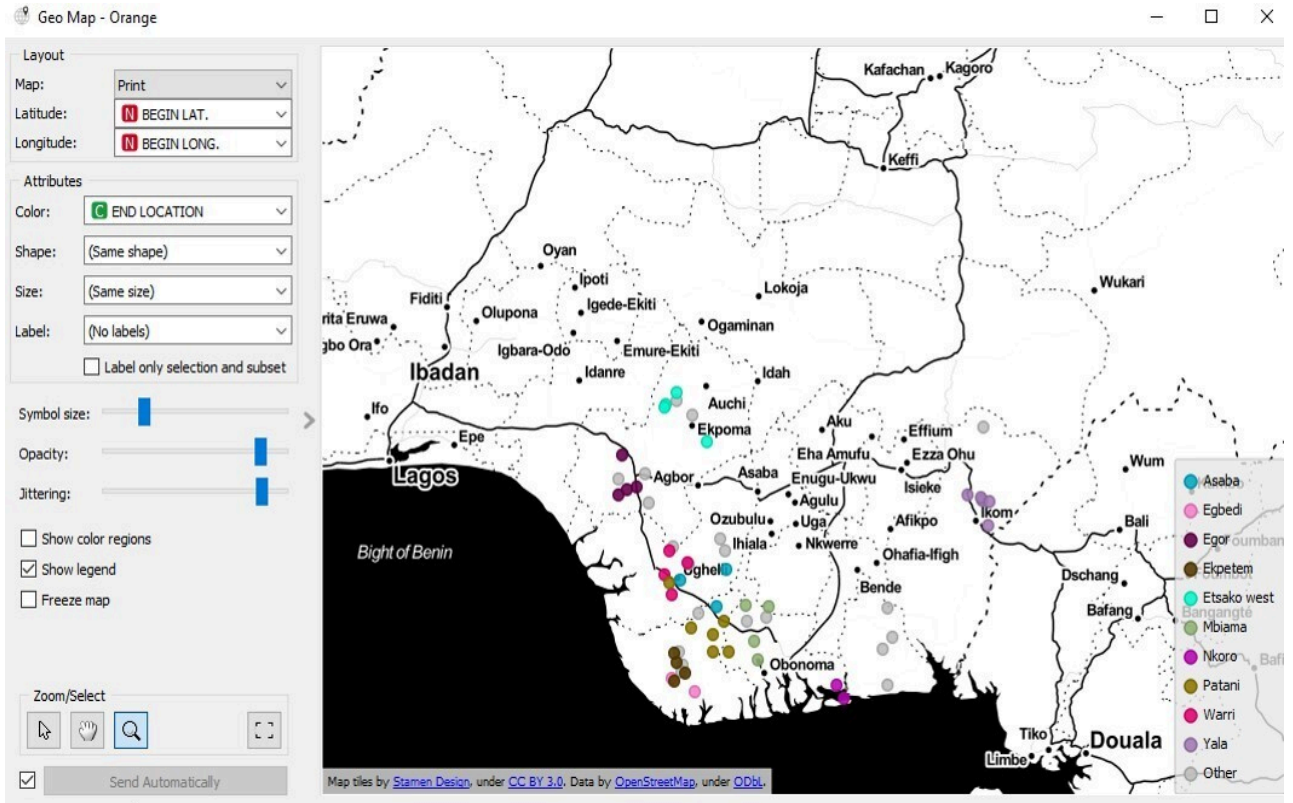


Figure 10. The impact of flood events at various towns within Nigeria's south-south.



Figure 11. A topographic map showing the number of deaths caused by flood events.

3.3. Model Performance Evaluation

The comparison of evaluation metrics from models constructed with both software tools and different test data sets shows that NB outperforms all other strategies, followed by RF. Note that the created model can be used to estimate the number of flooding incidents. As a result, the machine learning approach utilized in this work can provide insight into the patterns and frequency of flooding episodes, as well as the impact on population and property damage projected over a given period. **Table 4** summarizes the model performance characteristics for the various machine learning techniques utilized.

Table 4. ML model performance evaluation.

| ML model | AUC | AC | F1 | Precision | Recall |
|----------|-------|-------|-------|-----------|--------|
| NB | 0.906 | 0.831 | 0.835 | 0.856 | 0.831 |
| RF | 0.971 | 0.898 | 0.899 | 0.903 | 0.898 |
| LR | 0.596 | 0.661 | 0.526 | 0.437 | 0.661 |
| NN | 0.500 | 0.339 | 0.172 | 0.115 | 0.339 |
| SVM | 0.000 | 0.983 | 0.983 | 0.983 | 0.983 |

F1 is a simple metric that involves the overall recall and precision of the model, while AUC is the area under the ROC curve, which is determined at thresholds between the True Positive Rate and the False Positive Rate. According to **Figure 12**, NB (precision = 0.856) and RF (precision = 0.903) had the most accurate classifications of the flood event.

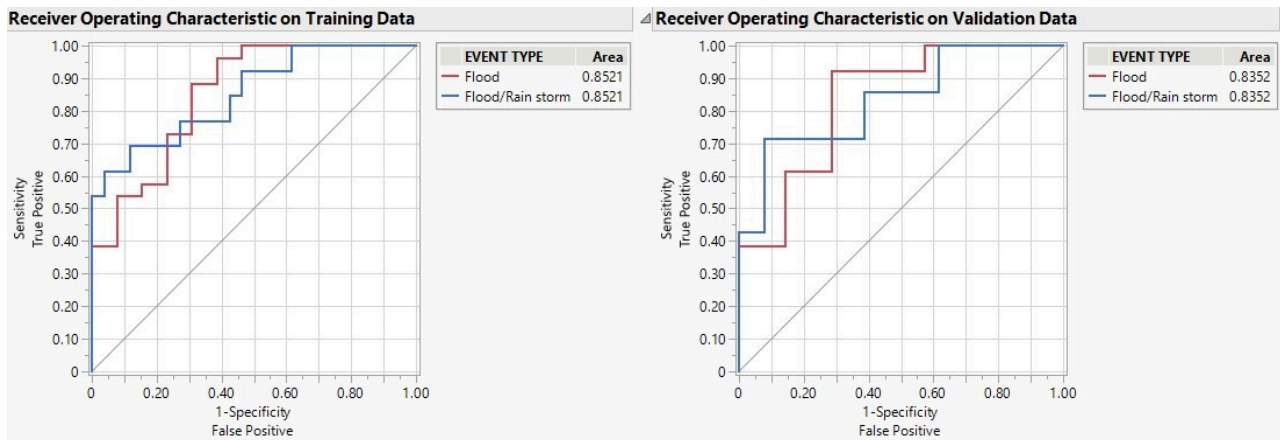


Figure 12. Sensitivity analysis for NN showing the ROC curve.

3.4. Discussion of Findings

This study uses thirty years' worth of historical flood data—which is extremely sparse because NEMA, the national disaster management agency, does not provide access to or availability of its data—to identify the types of floods that are most likely to occur in the future. Using MATLAB, the data was filtered to eliminate outliers, fill in missing values, arrange the data, and more. The machine was trained using 59 filtered instances (spanning the years 1990 to 2020) as the output while the remaining 25% of cases were used for testing. It is well recognized that Nigeria's south-south region is particularly susceptible to the effects of climate change because of its location, climate, vegetation, soils, economic structure, population density, energy needs, and agricultural practices. The location, size, and distinctive terrain of south-south Nigeria result in a range of climates, from the tropical hinterland climate to the tropical maritime climate, which is typified by the rainforest along the country's southern and coastal regions.

The location, duration, and effects of flood disasters on property and human life vary widely. To determine the kind of flood and its effects, it is necessary to consider several variables, including the location, duration, and geographic coordinates of the affected area. Assessing the flood event's intensity and impact can also be aided by knowing the number of deaths, property and crop damages, and direct and indirect injuries that resulted from it. Nonetheless, the International Flood Event Classification System (IFECS), which divides floods into three categories—minor, moderate, and major—can serve as a basis for classifying floods. The length of the flood event, the extent of the impacted area, and the depth of the inundation all play major roles in determining this classification.

The results, which are consistent with [23] study, showed that floods and flood/Rainstorms are frequent in the southern region, especially in south-south settlements. Flood incidents are shown in **Figure 4** for the initial location (communities), and **Figure 5** for the final location. As the print maps (**Figure 4** and **Figure 5**) show, there isn't much of a difference in the flood event between the beginning and ending locations. It was discovered throughout the study period that the flood had a significant effect on the destruction of farms and residences in the northern region, but the impact on homes (destruction of livable houses) was bigger in the southern location. About 400,000 people were most affected in 2000, as **Figure 6** demonstrates. Similar effects of flooding were also found by [24] [25] [26]. On the other hand, direct repercussions include harm and deaths brought on by the flood itself, such as hypothermia, drowning, and injury from falling objects. The wider repercussions of the flood on human life, such as the interruption of necessary services, the loss of a means of subsistence, and mental health problems, are referred to as indirect impacts. Damage to property and crops are additional crucial aspect to consider when assessing the flood's effects. These losses may be direct—caused by the flood's immediate effects on buildings and farmland—or indirect—resulting from the aftermath of the incident.

Even so, there have been several significant factors that have influenced the development and evaluation of flood disaster models over time. For example, enhanced data collection and storage capabilities have made it possible to provide more precise and detailed model inputs, which has led to better simulations of flood events. The employment of increasingly intricate and sophisticated models has been made possible by developments in computer technology, producing simulations that are more precise and in-depth. Thus, assessing the precision and dependability of flood disaster models has required contrasting model simulations with actual flood occurrences. The comparison of evaluation metrics from models built using software tools and different test data sets reveals that NB beats all other strategies, followed by RF. Note that the generated model can be used to estimate the number of flooding episodes. As a result, the machine learning approach used in this study can provide insight into the patterns and frequency of flooding events, as well as the expected impact on people and property damage over time. The findings of this study are consistent with the reports of [27] and [28]. Although Rajab *et al.* rely on historical climate information. However, [7] found that prediction using machine-learning algorithms is useful since it can use data from several sources and categorize and regress it into flood and non-flood categories. Although the authors utilized Non-linear (NARX) and Support Vector Machine (SVM) machine learning techniques, they did not specify the best algorithm.

4. Conclusion

Machine learning techniques offer significant potential for enhancing flood prediction and analysis capabilities in Southern Nigeria. This study identifies and

describes a robust evaluation of ML techniques for flood classification based on location, flood duration, begin/end location (name of the community), begin/end latitude and longitude, injuries direct/indirect, death direct/indirect, and houses, schools, farmlands, and crop damage. Extensive historical data was filtered and used for training and testing purposes. Several models were created and compared utilizing assessment criteria such as RMSE, MAE, and confusion matrix. The evaluation metrics from the models constructed show that the NB technique beats other techniques in terms of RMSE, MAE, and confusion matrix (accuracy rate of 78%), followed by RF (accuracy rate of 90.12%). By improving the accuracy and timeliness of flood forecasts, and better understanding the factors influencing flood events, these techniques can help mitigate the adverse effects of flooding in the region. However, challenges such as data availability, expertise requirements, and ethical considerations must be addressed to fully realize the potential benefits of machine learning for flood prediction and analysis in Southern Nigeria.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Gong, Y., Zhang, Y., Lan, S. and Wang, H.A. (2016) Comparative Study of Artificial Neural Networks, Support Vector Machines and Adaptive Neuro Fuzzy Inference System for Forecasting Groundwater Levels near Lake Okeechobee, Florida. *Water Resources Management*, **30**, 375-391. <https://doi.org/10.1007/s11269-015-1167-8>
- [2] Lynch, C.A. (2008) The Institutional Challenges of Cyberinfrastructure and E-Research. *EDUCAUSE Review*, **46**, 74-88.
- [3] Bose, I. and Mahapatra, R.K. (2001). Business Data Mining—A Machine Learning Perspective. *Information & Management*, **39**, 211-225. [https://doi.org/10.1016/S0378-7206\(01\)00091-X](https://doi.org/10.1016/S0378-7206(01)00091-X)
- [4] Hoai, M., Lan, Z.-Z. and De la Torre, F. (2011) Joint Segmentation and Classification of Human Actions in Video. *Conference on Computer Vision and Pattern Recognition* 2011, **2011**, 3265-3272. <https://doi.org/10.1109/CVPR.2011.5995470>
- [5] Mavhura, E., Manyena, S.B., Collins, A.E. and Manatsa, D. (2013) Indigenous Knowledge, Coping Strategies and Resilience to Floods in Muzarabani, Zimbabwe. *International Journal of Disaster Risk Reduction*, **5**, 38-48. <https://doi.org/10.1016/j.ijdrr.2013.07.001>
- [6] Mosavi, A., Ozturk, P. and Chau, K.W. (2018) Flood Prediction Using Machine Learning Models: Literature Review. *Water*, **10**, Article 1536. <https://doi.org/10.3390/w10111536>
- [7] Zehra, N. (2020) Prediction Analysis of Floods Using Machine Learning Algorithms (NARX & SVM). *International Journal of Sciences: Basic and Applied Research*, **49**, 24-34.
- [8] Olawoyin, R., Nieto, A., Grayson, R.L., Hardisty, F. and Oyewole, S. (2013) Application of Artificial Neural Network (ANN)-Self-Organizing Map (SOM) for the Categorization of Water, Soil and Sediment Quality in Petrochemical Regions. *Expert*

- Systems with Applications*, **40**, 3634-3648.
<https://doi.org/10.1016/j.eswa.2012.12.069>
- [9] Ighile, E.H., Shirakawa, H. and Tanikawa, H. (2022) A Study on the Application of GIS and Machine Learning to Predict Flood Areas in Nigeria. *Sustainability*, **14**, Article 5039. <https://doi.org/10.3390/su14095039>
- [10] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemman, P. and Witten, I.H. (2009) The WEKA Data Mining Software: An Update. *Special Interest Group on Knowledge Discovery in Data*, **11**, 10-18. <https://doi.org/10.1145/1656274.1656278>
- [11] NOAA (2020) 2019 NOAA Science Report.
- [12] Magami, I.M., Yahaya, S. and Mohammed, K. (2014) Causes and Consequences of Flooding in Nigeria: A Review Causes and Consequences of Flooding in Nigeria: A Review. *Biological and Environmental Sciences Journal for the Tropics*, **11**, 154-162.
- [13] NEMA (2018) National Emergency Management Agency 12 States Affected 4 States Are Declared under National Disaster 8 States Are under Red Alert 50 LGAs Affected. Flood Data 2002-2004, 2-4.
- [14] NEMA (2006) Disaster Risk Reduction and Prevention Country Name: Nigeria. Flood Data 2004-2006, 1-45.
- [15] Adeoye, N.O., Ayanlade, A. and Babatimehin, O. (2009) Climate Change and Menace of Floods in Nigerian Cities: Socio-Economic Implications. *Advances in Natural and Applied Sciences*, **3**, 369-377.
- [16] Etuonovbe, A.K. (2011) The Devastating Effect of Flooding in Nigeria. http://www.fig.net/pub/fig2011/papers/ts06j/ts06j_etuonovbe_5002.pdf
- [17] Oladokun, V.O. and Proverbs, D. (2016) Flood Risk Management in Nigeria: A Review of the Challenges and Opportunities. *International Journal of Safety and Security Engineering*, **6**, 485-497. <https://doi.org/10.2495/SAFE-V6-N3-485-497>
- [18] Cirella, G.T. and Iyalomhe, F.O. (2018) Flooding Conceptual Review: Sustainability-Focalized Best Practices in Nigeria. *Applied Sciences*, **8**, Article 1558. <https://doi.org/10.3390/app8091558>
- [19] Han, J., Pei, J. and Tong, H. (2022) Data Mining: Concepts and Techniques. 3rd Edition, Morgan Kaufmann.
- [20] Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A. and Štajdohar, M. (2013) Orange: Data Mining Toolbox in Python. *The Journal of Machine Learning Research*, **14**, 2349-2353.
- [21] Liu, F., Xu, F. and Yang, S.A. (2017) Flood Forecasting Model Based on Deep Learning Algorithm via Integrating Stacked Autoencoders with BP Neural Network. *Proceedings of the IEEE International Conference on Multimedia Big Data*, Laguna Hills, CA, 19-21 April 2017, 58-61. <https://doi.org/10.1109/BigMM.2017.29>
- [22] Sammut, C. and Webb, G.I. (2011) Encyclopedia of Machine Learning and Data Mining. Springer.
- [23] Njoku, J. (2012) 2012 Year of Flood Fury: A Disaster Foretold, but Ignored? Vanguard Newspaper. <http://www.vanguardngr.com>
- [24] Adegbola, A.A. and Jolayemi, J.K. (2012) Historical Rainfall-Runoff Modeling of River Ogunpa, Ibadan, Nigeria. *Indian Journal of Science and Technology*, **5**, 1-4. <https://doi.org/10.17485/ijst/2012/v5i5.5>
- [25] Odunuga, S., Adegun, O., Raji, S.A. and Udofia, S. (2015) Changes in Flood Risk in Lower Niger-Benue Catchments. *Proceedings of the International Association of Hydrological Sciences*, **370**, 97-102. <https://doi.org/10.5194/piahs-370-97-2015>

- [26] Anunobi, A.I. (2014) Informal Riverine Settlements and Flood Risk Management: A Study of Lokoja, Nigeria. *Journal of Environment and Earth Science*, **4**, 35-43.
- [27] Saravi, S., Kalawsky, R., Joannou, D., Casado, M.R., Fu, G. and Meng, F. (2019) Use of Artificial Intelligence to Improve Resilience and Preparedness against Adverse Flood Events. *Water*, **11**, Article 973. <https://doi.org/10.3390/w11050973>
- [28] Rajab, A., Farman, H., Islam, N., Syed, D., Elmagzoub, M.A., Shaikh, A., Akram, M., and Alrizq, M. (2023) Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh. *Water*, **15**, Article 3970. <https://doi.org/10.3390/w15223970>