International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 8 Aug 2023, pp: 400-409 www.ijaem.net ISSN: 2395-5252

Modeling Of a Data Driven Flood Detection and Early Warning System Using Machine Learning Technique

Theophilus Ndulue
Vincent C. Chijindu, Edward C. Anoliefo, and Joy N. Eneh.

Date of Submission: 10-08-2023 Date of Acceptance: 20-08-2023

ABSTRACT
This research presents modeling of a data driven flood detection and early warning system using machine learning technique. The aim is to save lives and properties, through detection of flood, early notification and warning for immediate evacuation. To achieve this, Anam community in Anambra State, Nigeria, which is annually characterized with flood since 1966 till date was considered as the case study. Literatures were reviewed on flood detection systems and lack of reliability was identified as a research gap. To address the gap, the methodology used are hydrological flood modeling, data collection, explorative data analysis approach, data processing with multiple imputation approach, machine learning algorithms (Linear Regression (LR), Random Forest (RF) and Decision Tree (DT)), early warning system based alarm notification. The models were implemented with Python programming language. The results of the machine learning flood detection models were evaluated considering recall, accuracy, receiver operator characteristics curve and compared after tenfold cross validation. The average recall obtained for LR is 99.146%, average accuracy of flood detection is 91.87% and average ROC result which is probability of correct flood detection is 83.23%. Similarly the validation result of the LR detection model for flood early detection reported average recall of 88.42%, average accuracy of flood detection is 92.409% and average ROC result which is probability of correct flood detection is 83.44%. In the same vein, the DT reported average recall of 77.77%, average accuracy of flood detection is 75% and average ROC result which is probability of correct flood detection is 72.22%. The table 5 was used to compare the models to recommend the best which was used for the modeling of the new system. Overall the RF and LR recorded better performance than the DT and are recommended for the modeling of flood early warning system.

Keywords: flood machine learning, early warning, data driven, python, rainfall, pressure, volume

I. INTRODUCTION

Climatology has continued to gain increase research attentions over the past few decades, due to the impact it has on the society and potential intensification in time series as a result of temperature increase as represented in [1]. Intensive researches have been triggered to determine the impact of this climate change on extreme precipitation (rainfall) and relate how it affects the society [2]. Ali and Mishra [3] Revealed that climate change impacts on the sensitivity of temperature and results to high evaporation rate of water vapor from the oceans, sea and rivers, thus leading to persistent precipitation that causes flood. Before diving deep into the impacts of flood, it is vital to point out some of the major cause of climate change which results to this flood problem at the first place. According to the United Nations Environmental Programme (UNEP), climate change are caused by changes in the land patterns like deforestation, human activities like the air pollution through fossil fuel, oil and gas, industrial chemical waste, etc. which result to increase in temperatures and directly destroys the green house. All these events indirectly increase the intensity of rain, which triggers river overflow on dry land as flood. In the same vein it has been revealed that the impact of climate change leading to flood will continues to increased and create new records in different part of the world, like the case of Nigeria where over 600 people killed and more than 1.3 million people displaced due to recent flood event between April and November, 2022 [4] [5].

Over the years, methods to control the problem of flood recommends strategic geographical planning which involve approaches like adequate drainage systems [6]; forestry [7], strict administrative rules to minimize the impact of industrial wastes and environmental pollution [8], among others. However, these approaches are more of long term solutions and cannot address the problem of flood at the moment. UNEP [4] posited that annually, the current global temperature increase by 1.1^{o}C, which is not good. The implication of this evidence shows that the rate of flood will continues to increase globally in the coming years, and hence requires urgent solution to manage the problem.

Early flood detection and warning system is one of the most common flood impact mitigation measured currently in use globally for flood detection and control [9]. It is an embedded device made of sensor, controller and communication module, with the capacity to detect this hydrological flood model, process it and signal occupants for evacuation [10] [11]. The sensor is responsible for data collection, using transducers like ultrasonic sensors, infrared sensors, piezo-meter, bubler sensors, etc [12], then the controller process the data and then if flood is detected uses rule based or optimization algorithm to notify the users via the communication module.

Over time, extensive researches have been focused on improving the performance of flood detection and early warning systems, using various techniques. Mousa et al. [10] used ultrasonic and infrared sensor to solve the problem of flood detection and then applied artificial neural network for the control. Ahmad et al., [13] used machine learning to develop the flood detection model using hydrodynamic parameters and applied it for early warning system. Jahangir et al. [14] proposed the use of Geographic Information System (GIS) based on satellite imagery for spatial mapping of the flood incoming and early notification. Samikwa et al., [15] improved the instant notification performance of flood detection system using artificial neural network for real time monitoring. [16] Developed a detection model for early flood detection notification system, using artificial neural network, among many other studies.

However, despite the success of the existing systems reviewed, presently to the best of the researcher's knowledge, there is no flood detection system that is generally accepted for flood detection and early warning, due to many technical problems such as limited risk knowledge, very high implementation cost like the case of satellite based remote sensing, false alarm in monitoring and forecasting, delay response capabilities, etc. thus leading to poor acceptability and system unreliability. There is need for an improved solution which is cheap but very effective in early detection of flood and notifying for control measures via evacuation.

This research proposed to model multiple machine learning based solution for early flood detection and control, and then recommend the best model after evaluation and validation for flood detection and monitoring solution. This when achieved will provide the desired reliable system for intelligent flood detection and early warning and hence go a long way to save lives and properties that are normally affected by previous flood events.

II. LITERATURE REVIEW

Shahabi et al. [17] presented a study on flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. The work is aimed at the application of Deep Backpropagation Genetic Algorithm (DBPGA) which is based on Deep Belief Network (DBN), Back Propagation (BP) and optimized using Genetic Algorithm (GA) for generation of an accurate flood susceptibility map for the watershed area of Iran. One-R Attribute Evaluation (ORAE) technique was used for the database which compares ten conditioning factors and creation of 194 flood locations. The result of the model reported an Area under Curve (AUC) of 0.989 and precision of 0.985. [18] Applied Artificial Neural Network (ANN) for flood warning systems using data collected from rainfall runoff. In addition hydrological model of steam flow was also developed. Two radar-based stream forecast models were developed with ANN using the respective input data. The models are radar-runoff (2D) model and radar-rainfall-runoff (3R) model. [19] Researched on flood detection in urban areas using satellite imagery and machine learning. The work used sentinel 1 satellite imagery and information reports from police departments for the acquisition of data fed to Support Vector Machine (SVM), Random Forest, and Maximum Likelihood Classifier (MLC). The study further presented a Change Detection (CD) approach which is based on Otsu algorithm, iso-clustering and fuzzy rules for detection of flood in urban areas. The CD approach presented the best performance result out of all other models with 0.81 precision measures, 0.85 F1-score, and 0.87 accuracy measures.

[20] Presents a work on application of integrated aerial imagery and convolutional neural

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 8 Aug 2023, pp: 400-409 www.ijaem.net ISSN: 2395-5252

network for flood detection. The work applies Convolutional Neural Network (CNN) on images acquired from an Unmanned Aerial Vehicle (UAV) for the determination of flood information in a location. The work uses 2150 image datasets for the training of the CNN algorithm for accurate detection of flood and flood related disasters. The work is aimed at helping in disaster management and organization of response for disaster damages in localities and improve livelihood in smart cities. The result of the study shows that the CNN model adopted achieved an accuracy of 91% in flood detection. Similarly, [21] Researched on Convolutional Neural Network (CNN) based flood management system with Internet of Things (IoT) sensors and cloud data. The study presented the management of water storage facilities like lakes and dams by mounting sensor signals on them to monitor the flow and the amount of water present in the system to control the likelihood of flooding the environments. The study used sensor to collect data from dams and lake and then train a CNN. The result of this work achieves a detection accuracy of 85.53%, sensitivity of 92.11% and specificity of 78.95%.

Mane et al. [22] presented a study on early flood detection and alarming system using machine learning techniques. The work experimented on the implementation of machine learning techniques like support vector machine, k-nearest neighbor, logistic regression, and naïve bayes for the early detection of flood and alarm system which consists of website and android application for alerting the concerned masses and authorities. The system uses rainfall datasets available for training the machine learning techniques. The result of the system performance presented that the support vector machine and logistic regression has the highest performance of 79% accuracy than the others. Priscillia et al. [23] Researched on flood susceptibility assessment using ANN. The study uses some environmental factors like elevation, Topographical Witness Index (TWI), slope, curvature, land-cover, Euclidean distance from river, soil type, precipitation and Stream Power Index (SPI) alongside Sentinel-1 satellite imagery to develop a model that back-predict flood susceptibility in an area in January 2020 for historic flood event in 260 key locations. It uses Synthetic Minority Oversampling Technique (SMOTE) for the implementation of balance in both classes in the training set of the machine learning algorithms such as K-Nearest Neighbor (K-NN), Random classifier, ANN, and Support Vector Machine (SVM). The result of the study shows that the ANN algorithm has the best performing F1 score (harmonic mean of precision and recall) of 0.45, random classifier with second best with score of 0.24, SVM with 0.00 and k-NN is 0.03.

2.1 Research Gap

While these studies have all made great contribution on the detection and flood, gap existing in the reliability of the models and also inconsistencies reported in the various results on the best models. In addition the problem of false alarm was not addresses as pressure and water volume which are key elements of flood were not mentioned.

III. METHODOLOGY

The research method begins with the review of relevant literatures which considers previous works on flood detection and early warning systems. From the study, it was observed that the hydrological flood model of the existing system to the best of the researcher's knowledge did not consider pressure, which is a key variable for the modeling of flood. To this end, this study developed an improved flood hydrological model considering the identified variable and trained multiple machine learning algorithms with the data to generate detection models. The models were validated with comparative analysis and the best was used to model an early warning flood detection system for the people of Anam community.

3.1 Hydrological modeling of the problem formulation by River Niger

The hydrological model of the problem formulation used mathematical equations to describe how this flood occurs in the community, considering the three rivers, among which the Anamabra and Ezichi rivers are the sub-rivers which flows into the main river Niger channel and afterwards overflows and case flood in the community. The governing equation of the unsteady flow of the river Niger inspired from the 1-Dimensional river low model in [23]; However the model did not consider the water pressure was identified from the review as vital to solve the problem of false alarm and differentiate flood from erosion. The modeling of a single flowing river in an unsteady state condition is presented equation 1

$$\frac{\partial Q}{\partial x}+\frac{\partial Ap}{\partial t} \quad (1)$$

Where Q is the output of the water flow; x is the distance along the main stream, A is the cross surface area, t is the time and p is pressure.

Equation 1 serves as a base equation for modeling the unsteady flow of the river Niger as presented in equation 2.

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left[\frac{Q^2}{Ap}\right] + gAp\frac{\partial Z}{\partial x} + g\frac{n^2 Q|Q|}{ApR^{4/3}} \quad 2$$

Where Z is the water level, n is the manning's roughness coefficient, R is the hydraulic radius and g is gravitational force. The equation 2 is derived from Equation 1 by incorporating additional factors to represent the inflow of the Ezichi River and Anambra River into the river Niger. From equation 1, the $\partial Q/\partial t$ represents the change in water flow (Q) with respect to time (t), which is a common component also in equation 2; $\frac{\partial}{\partial x}\left[\frac{Q^2}{Ap}\right]$ represents the change in water flow velocity $\left(\frac{Q^2}{Ap}\right)$ with respect to distance (x) along the main stream. Here, $\frac{Q^2}{Ap}$ is the velocity of the water flow, and the derivative with respect to distance accounts for any changes in velocity along the river. $gAp\frac{\partial Z}{\partial x}$ incorporates the change in water level (Z) with respect to distance (x) and gravitational force (g). $g\,(n^2Q|Q|)/ApR^{\frac{4}{3}}$ accounts for the effect of Manning's roughness coefficient (n), water flow velocity (Q), cross-sectional area (A), hydraulic radius (R), and gravitational force (g) on the flow. It represents the impact of friction and resistance to flow in the river. By combining these additional terms with the base equation (Equation 1), Equation 2 accounts for the complexities of the river system, include the inflow from the Ezichi River and Anambara River, the impact of water level changes, and the effects of roughness and resistance to flow which formulates the flood model in equation 3.

$$Q_m + \sum_{l=1}^{I_m} Q_{m,i} = \frac{\partial W_m}{\partial t} m = 1,2,3,4 \ldots\ldots, k \quad 3$$

The equation 3 introduces the concept of multiple rivers connected to the river Niger and their respective channels flowing into different nodes (mth nodes). It calculates the total discharge into the Anam community by summing up the individual discharges from these connected rivers. From the equation Q_m represents the total discharge into the mth node, which is a cumulative value of the inflow discharges from the connected rivers. $\sum_{l=1}^{I_m} Q_{m,i}$ represent the sum of the inflow discharges $(Q_{m,i})$ from all the channels (ith channel) connected to the mth node. It considers the contribution of each channel to the total discharge. $\frac{\partial W_m}{\partial t}$ represent the change in the inflow discharge of the river (W_m) with respect to time (t). It accounts for any variations in the inflow rate into the river Niger. By combining these components, Equation 3 captures the total discharge into the Anam community by considering the inflow discharges from all the connected rivers and their respective channels. It takes into account the changes in the inflow rate over time to provide a comprehensive understanding of the flood problem and the overall water flow dynamics in the community.

3.2 Data collection

Data of the flood problem was collected from Nigerian Emergency Management Agency (NEMA), South East Zone; 55A independence Avenue, Enugu. The data collected was characterized by Anambra state flood problem which occurred as a result of river Niger overflow as formulated in equation 3. The data contain records of rainfall in within the location from April till December 1965 to 2021 and considering the water volume and pressure as key flood parameters.

3.3 Explorative data analysis and Data Processing

Explorative data analysis approach [25] was used for the graphical representation of the attributed. This was used to show the normal distribution of the monthly rainfall in the number of years considered for the study and also the yearly or annual rainfall recorded. In addition, data processing was used to fix error characterized with the flood data attributes collected such as missing data and issues of noise. To achieve this, multiple imputation technique which is a missing data replacement approach was adopted from [26] and used to replace the data missing from the data set. The approach used detection model to estimated and replace missing data repeatedly until a standard dataset is produced. Data extraction was then applied to drill the flood features values into a compact statistical feature vector. This was achieved utilizing the min-max feature extraction approach which [27] submitted as a good feature extraction solution for pattern recognition problem. The approach uses the minimum or maximum feature values of the attributes to determine the feature selection process and ordering [28].

3.4 Machine learning based Models

From the review of literatures, various machine learning algorithms were used for the detection of flood, but due to the inconsistencies with the results, it has been difficult for research to make decision on the best ML algorithm to adopt for the detection model development. To this end, three of the most popular ML algorithm for developing detection models which are decision tree, random forest and linear regression are considered for the study.

3.5 Linear Regression (LR)

This ML algorithm is a popular detection model which can employed the relationship between dependent and independent variable to predict the future outcome of data input. The mathematical model of the LR was presents as [74] which used $\theta_1 \theta_2 \ldots \ldots \ldots . \theta_n$ as the set of data input from the flood datasets to predict the outcome of future flood. Where n is the number of input attributes. The general model of LR is presented as;

$$h_\theta(x) = \sum_{i=0}^{m} (\theta_i x_i) \quad (5)$$

Where $h_\theta(x)$ is the predicted value. The cost function ($j(\theta)$) is presented as;

$$j(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=0}^{m} h_\theta(x^i) - y^i)^2 \quad (6)$$

Where y^i is the actual value, while j has to be reduced to ensure better accuracy. The optimization algorithm which minimizes the cost function (it calculates how wrong the model output is) is the gradient descent which is presented as in equation 7;

$$flood(\theta_j) = \theta_j - \alpha \frac{\delta}{\delta \theta i} j(\theta_0, \theta_1)) \quad (7)$$

Where m is the total samples in the flood datasets, α presents the rate of learning by the LR, which varies between 0 and 0.5. This process changes for every value of equation 6 until convergence is achieved. The figure 2 presents the flow chart of the LR algorithm. The weights of the algorithm are initialized and then values are randomly assigned for the learning process using the variant of the learning rates (0-0.5), while the error rate and cost function are monitored and minimize using the gradient descent optimization algorithm in equation 7, which continuously minimize the cost function (j) in equation 6 until convergence is achieved and the detection outcome in equation 5 generated.

3.6 Decision Tree (DT)

DT uses flowchart like tree structure to solve regression problem from series of splitted features. The DT begin with the root where the population are divided according to features of the datasets and then decisions are made to generate new leafs, while the other leaf which makes the wrong decisions based on the output of the entropy function are pruned to stop over-fitting problem.

$$E(S) = -p_{(+)} \log_{(+)} - p_{(-)} \log_{(-)} \quad (8)$$

Where $p_{(+)}$ presents the probability of positive class is, $p_{(-)}$ is the probability of negative class and S is the training data subset.

3.7 Random Forest (RF)

The RF is an ensemble machine learning algorithm which is made of many decision trees and can be employed to solve both regression and classification problem. In this study the DF was adopted and used for the training of the flood data collected. The operation of the RF identified the training classes and the number of variables in the classifier, which is used as a determinant for the tree node creation. For each tree the Gini index model in equation (9) is used to vote for the detection performance and determine the tree with poor probability of correct detection and voted for the next tree generation. At the end of the votes the average of all the voted tree output is used to determine the detection model. The Gini model is presented as [29];

$$Gini(split) = \sum_{i=1}^{n} P_i * (1 - P_i) \quad 9$$

Where i presented the class of the dataset, n is the number of class labels, P is the proportion of class label.

3.8 Result of the flood detection models

This section evaluated the performance of the flood detection model using Recall, accuracy and Receiver Operator Characteristics (ROC) curve. The models used for the formulation of the parameters are;

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FN + TN} \quad (11)$$

Where TP is true positive rate, FP is false positive rate, FN is false negative rate, TN is true negative rate. The results were also validated using tenfold cross validation approach, which tested the detection model, ten times and calculate the average. While the ROC depicts the sensitivity of the classifier model, which compares the changes between the TPR and FPR.

3.9 Training of the machine learning models

Having presented the ML models in the previous sections, they were respectively train with the flood data collected for the generation of the detection model. Prior to the training phase, the loaded data were processed using the imputation data replacement approach which used regression approach to predict the values of missing data based on the currently available data and the replace. Similarly the max-min feature extraction approach was used to extract the data into feature compact statistical vectors and then feed for training. The training datasets feed into LR was trained with gradient descent model in equation 6 to minimize the cost function in equation 5 and generate the detection model. For the DT, entropy model in equation 8 was used to determine the probability of false classification and when the cost function is minimized, the flood detection model is generated. Likewise with the RF which used Gini index probability function in equation 9 to solve the optimization problem and generate the detection model for flood.

3.10 The Early warning system

This section of the study presents the early warning system which is used for the notification of habitats in flood prone environment, informing them ahead of time against flood incoming. To achieve this many methods such as internet of things, SMS alert system, alarm notification, danger indicator etc can be used. These early warning systems are simply incorporated with the flood detection system as input and then notify the environment of the problem. The algorithm of the early warning system is presented as (Algorithm 1);

1. Start
2. Identify output of flood detection model
3. Initialize parameters (time (t) and transmission control protocol (TCP)
4. Set t=0; set subscriber phone number
5. If flood input=true
6. Then
7. Process data
8. Else
9. Return to load new input
10. End if
11. If process data = true
12. Then
13. Activate TCP
14. Identify subscriber phone number
15. Notify management agency
16. Text ”flood incoming”
17. Else
18. Count down t-1
19. For t=0
20. Sound alarm
21. End for
22. Return to input model
23. End

3.11 Flood Detection and Early Warning System

The section presents the complete flood detection and early notification system developed with the hydrological flood model developed with pressure and water volume as key elements, machine learning model generated for flood detection and the early warning system. The test flood data loaded was processed to address data imbalance problem as a result of missing data and extracts the feature into the machine learning based detection models developed or regression. The ML algorithm used the reference flood detection model to detect time series flood by comparing the trained features of the hydrological flood model with the features of the test flood dataset. When flood is detected the early warning system initialized time and Transmission Control Protocol (TCP) function to notify the relevant agencies of the problem and also sound alarm as reported in the algorithm 1. The pseudo-code of the flood detection and early warning system was presented as;

1. Start
2. Load flood model
3. Processing data
4. Feed to flood detection model
5. If flood =true
6. Then
7. Activate algorithm 1
8. Else
9. Return to step 2
10. End

IV. RESULTS OF THE MACHINE LEARNING MODELS

The result will evaluate the training performance of the three machine learning algorithms trained with the hydrological flood model considering the accuracy of detection, recall and receiver operator characteristics as defined by equation 10 and equation 11 respectively.

4.1 Result of the RF Detection Model

This section presented the result of the RF detection model in figure 3.4. The result showed the performance of the model when loaded with the data of the flood hydrological model in equation 3 for training. During the training process, the RF used the Gini index function in equation 9 to determine the probability of correct classification for each tree and the vote out the detection model. The detection model was evaluated with recall, accuracy and receiver operator characteristics (ROC), and the result recorded is presented in table 1;

Table 1: Performance of flood detection model with RF

Parameters index	Results (%)
Recall	100
ROC	83.33
Accuracy	91.67

The table 1 showed the performance of the RF flood detection model. The implication of the result showed that the probability of correct detection for flood with RF is 83.33 which are good as the recall which is the probability of correct flood detection recorded 100%. In addition, the accuracy of the detection outcome is 91.67% which is also very good and implied good training process and correct detection model. What this means is that the RF model as an algorithm for flood detection is very good and will produce correct detection outcome with high efficiency. To validate the result, tenfold cross validation approach was adopted and used to iteratively test the model and the average computed (Refer to table 4 for all validation results).

4.2 Result of the LR Detection Model

The result of the LR detection model was discussed in this section. The model in figure 3.2 was loaded with the training flood hydrological model in equation 3 and trained using gradient descent algorithm in equation 7 to minimize the cost function model in equation 6, until the desired detection model was achieved. The result generated from the model during the training process for its evaluation is reported in table 2;

Table 2: Performance of flood detection model with LR

Parameters index	Results
Recall	88.88
ROC	84.44
Accuracy	0.9167

The table 2 presented the performance of the LR detection model for flood. The model outcome after evaluation showed that the probability of true classification of flood as shown in the recall is 88.88% which is good. In addition the probability of correctly flood detection as in the ROC is 84.44% which is also good, while the overall accuracy of the model is 91.67% which implied good success rate for detection of flood. The implication of the result showed that LR achieved good detection performance for the flood early detection, which is very good. To validate the result, tenfold cross validation approach was adopted and used to iteratively test the LR model and the average computed (Refer to table 4 for all validation results).

4.3 Result of the DT Detection Model

The performance of the DT detection model for early flood detection and also examined in this section considering similar performance evaluation matrices used for other ML counterparts. The result recorded was presented in the table 3;

Table 3: Performance of flood detection model with DT

Parameters index	Results
Recall	77.77
ROC	72.22
Accuracy	75.00

The table 5 presented the performance of the DT, showing how the detection model which used entropy model in equation 3.8 for the minimization of error probability during the training process to generate the flood detection model. The result reported recall of 77.77%, ROC of 71.22% and overall flood detection accuracy of 75%. This implied that the DT was also able to correctly learn and predict the flood problem. To validate the performance of the DT flood detection model, the table 4 was used.

Table 4: Validation Performance of flood detection model with LR, RF and DT

Fold	LR			RF			DT		
	Recall	Accuracy	ROC	Recall	Accuracy	ROC	Recall	Accuracy	ROC
1	88.88	91.67	84.44	100.00	91.67	83.33	77.77	75.00	72.22
2	88.77	90.44	83.56	98.18	90.18	83.33	88.77	74.44	71.56
3	88.13	92.78	83.34	99.12	90.78	83.33	88.13	75.78	71.34
4	88.04	91.23	85.18	98.04	91.84	85.17	88.04	75.23	72.18
5	86.94	94.66	84.23	100.00	93.18	84.12	86.94	76.66	72.23
6	87.45	92.77	83.65	100.00	92.77	83.33	87.45	75.77	73.65
7	88.77	91.67	84.43	98.19	91.67	84.10	88.77	76.67	72.43
8	88.98	90.65	80.65	98.98	91.67	80.98	88.98	74.65	72.65
9	89.17	94.12	83.77	99.17	94.12	83.33	89.17	74.12	72.77
10	89.10	94.10	81.13	99.78	90.89	81.28	89.10	74.10	72.13
Avg.	88.423	92.409	83.438	99.146	91.877	83.23	77.77	75.00	72.22

From the result in table 4, the average recall obtained for LR is 99.146%, average accuracy of flood detection is 91.87% and average ROC result which is probability of correct flood detection is 83.23%. Similarly the validation result of the LR detection model for flood early detection reported average recall of 88.42%, average accuracy of flood detection is 92.409% and average ROC result which is probability of correct flood detection is 83.44%. In the same vein, the DT reported average recall of 77.77%, average accuracy of flood detection is 75% and average ROC result which is probability of correct flood detection is 72.22%. The table 5 was used to compare the models to recommend the best which was used for the modeling of the new system.

Table 5: Comparative Analysis

Parameters index	Recall (%)	Accuracy (%)	ROC (%)
DT	77.77	75.00	72.22
RF	99.146	91.877	83.23
LR	88.423	92.409	83.438

The table 5 presented a comparative analysis of flood detection model considering DT, RF and LR. The result when compared considering accuracy, recall and ROC performance. From the result, it was observed that the RF achieved better flood detection performance when compared with the counterparts. This implied that the flood detection and early warning system was developed with the RF as the machine learning model which detect the flood problem and then signal the early warning algorithm in algorithm (1) to notify the relevant agencies for control measures. In addition, further validation of the models was performed through comparative analysis with state of the art flood detection algorithms as reported in table 8.

Table 8: Comparative state of the art flood detection algorithms

Author	Algorithms	Accuracy (%)
[31]	SVM	25.07
	DT	72.25
	ANN	77.10
[20]	CDA	87.00
[21]	CNN	91.00
[22]	SVM	79.00
	Logistic regression	79.00
[18]	CNN	85.53
[30]	RF	98.70
New system	New RF	91.87

The table 8 presented come of the reviewed flood detection model developed with recent techniques such as machine learning and deep learning. From the result, it was observed that only the RF in [30] achieved better accuracy when compared with the new LR. However the new RF is more reliable as pressure which is key element o flood was considered in the hydrological model of the problem formulation in the equation 3 as against the case of the RF in [30].

REFERENCE

[1]. Renaud B., Hayley J., Stephen B. (2019)" A synthesis of hourly and daily precipitation extremes in different climate regions" Elsevier; weather and climate extremes; 26; 100219

[2]. Korbinian B., Gunter B. (2021)" Understanding the relationship between rainfall an flood prabilities though comined intensity duration frequency analysis"Journal of hydrology; vol 602; Science direct; 126759

[3]. Ali H. and Mishra V. (2018)" contribution of dynamic and thermodynamic scalling in sub-daily precipitation extremes in india "Geophys. Res. Lett., 45.

[4]. United Nations Environmental Programme (UNEP), (2020) "How climate change is making recorded breaking floods the new normal" Unep.org.

[5]. Lanre O. (2022)" Facts check: Is the devastating flood in Nigeria influenced by climate change?"; development cable; Article; available at www.thecable.ng. accessed 4/14/2022.

[6]. Eddy S., Suryadi F., Tarigan K. (2018) "Strategy of drainage and flood control in Palembang city"; Drainage and flood control section in water resource management; Palembang; pp. 1-14.

[7]. Ian R., Bruce A. (2006)" Forest and floods: Moving to an evidence based approach to watershed and integrated flood management" International Water Resources Association 31(1). Pp. 1-8.

[8]. Maryann O., Avemaria M. (2022) "A basic design aproach of drainage facilitie for effective prevention and control of flooding" Nigerian journal of engineering science research; 2(1): pp. 01-12.

[9]. Duminda P., Ousmane S., Jetal A., Hamid M. (2020) "Challenge and technical advances in flood early warning system" Inbook: flood impact mitigation and resilience enhancement; intchOPen; pp. 1-18.

[10]. Mousa M., Zhang X., & Claudel C., (2016). "Flash Flood Detection in Urban Cities Using Ultrasonic and Infrared Sensors", IEEE Sensors Journal. DOI 10.1109/JSEN.2016.2592359 http://hdl.handle.net/10754/618396

[11]. Syifa M., Park S., Achmad A., Lee C., & Eom J., (2019) "Flood Mapping Using Remote Sensing Imagery and Artificial Intelligence Techniques: A Case Study in Brumadinho, Brazil", Advances in Remote Sensing and Geoscience Information Systems of Coastal Environments. Journal of Coastal Research SI 90. Pp 197-204 Coconut Creek, Florida ISSN 0749-0208.

[12]. Dejan N., (2015) "Ultrasonic Sensor HC-SR04 and Arduino – Complete Guide", https://howtomechatronics.com/tutorials/arduino/ultrasonic-sensor-hc-sr04/ Accessed December 2022

[13]. Ahmad M., Al-Mehedi M., Yazdan M., & Kumar R., (2022) "Development of Machine Learning Flood Model Using Artificial Neural Network (ANN) at Var River". Liquids 2022, 2, Pp 147–160. https://doi.org/ 10.3390/liquids2030010

[14]. Jahangir M., Reineh S., & Abolghasemi M., (2019). "Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm", Weather and Climate Extremes journal homepage: www.elsevier.com/locate/wace

[15]. Samikwa E., Voigt T., & Eriksson J., (2020) "Flood Prediction Using IoT and Artificial Neural Networks with Edge Computing", IEEE iThings-2020: Permanent link to this version: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-427280

[16]. Elsafi S., (2014). "Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan", Alexandria University Alexandria Engineering Journal www.elsevier.com/locate/aej www.sciencedirect.com 3, Pp 655–662

[17]. Shahabi H., Shirzadi A., Ronoud S., Asadi S., Pham B., Mansouripour F., Geertsema M., Clague J., Bui D., (2021). "Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic

algorithm", Geoscience Frontiers journal homepage: www.elsevier.com/locate/gsfhttps://doi.org/10.1016/j.gsf.2020.10.007

[18]. Smys S., Basar A., & Wang H., (2020) "CNN based Flood Management System with IoT Sensors and Cloud Data", Journal of Artificial Intelligence and Capsule Networks (2020) Vol.02/ No.04 Pages: 194-200 http://irojournals.com/aicn/ DOI: https://doi.org/10.36548/jaicn.2020.4.001

[19]. Han J., (2002). "Application of Artificial Neural Networks for flood warning systems", North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy Civil and Environmental Engineering Raleigh, North Carolina.

[20]. Tanim A., McRae C., Tavakol-Davani H., Goharian E., (2022) "Flood Detection in Urban Areas Using Satellite Imagery and Machine Learning", Water 2022, 14, 1140. https://doi.org/10.3390/w14071140

[21]. Munawar H.S., Ullah F., Qayyum S., Khan S., Mojtahedi M., (2021) "UAVs in Disaster Management: Application of Integrated Aerial Imagery and Convolutional Neural Network for Flood Detection", Sustainability 2021, 13, 7547. https://doi.org/10.3390/su13147547

[22]. Mane P., Katti M., Nidgunde P., & Surve A., (2020). "Early Flood Detection and Alarming System Using Machine Learning Techniques", International Journal of Research in Engineering, Science and Management Volume-3, Issue-10, October-2020 journals.resaim.com/ijresm | ISSN (Online): 2581-5792 | RESAIM Publishing Pp 29-32

[23]. Priscillia S., Schillacci C., & Lipani A., (2021). "Flood susceptibility assessment using artificial neural networks in Indonesia", Artificial Intelligence in Geosciences journal homepage: www.keaipublishing.com/en/journals/artificial-intelligence-in-geoscience

[24]. Zuo-tao XIE, Fang-li YANG , Xiao-li FU (2012)"Mathematical model for flood routing in Jingjiang River and Dongting Lake network" Water Science and Engineering, 2012, 5(3): 259-268 doi:10.3882/j.issn.1674-2370.2012.03.002

[25]. AvijeetB.(2022)"Whatisexploratorydataanalysisstepsandmarketanalysis"Articleofdata science and Business anaystics; available at http:// www.simplilearn.com; accessed 11/17/2022.

[26]. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013 May;64(5):402-6. doi: 10.4097/kjae.2013.64.5.402. Epub 2013 May 24. PMID: 23741561; PMCID: PMC3668100.

[27]. Pudil P., Novovivova J., Choakjarernwanit N., Kittler J. (1993)" An analysis of the max-min approach to feature selection and ordering" Pattern recognition letters; vol14, issue 11;pp. 841-847.

[28]. Banu R, Sidram MH (2022) Window Based Min-Max Feature Extraction for Visual Object Tracking. Indian Journal of Science and Technology 15(40): 2047-2055. https://doi.org/ 10.17485/IJST/v15i40.1395

[29]. Nocholas D. (2022)" Random forest: A complete gude for Machine learning" Data science; builtin articles; accessed 17/3/2023.

[30]. Hashi A., Abdirahman A., Elmi M., Hashi S., & Rogriguez O., (2021). "A Real-Time Flood Detection System Based on Machine Learning Algorithms with Emphasis on Deep Learning", International Journal of Engineering Trends and Technology Volume 69 Issue 5, 249-256, May 2021 ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I5P232 Pp 249-258

[31]. Fadratul Hafinaz Hassan, Nur Amiera Azelan (2018) "Comparing Performance of Machine Learning Algorithms in a Flood Prediction Model with Real Data Sets" International Journal of Advanced Trends in Computer Science and Engineering" vol 8; no 1, 4; pp 152-158.