International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at :www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET2411444

A Conceptual Machine Learning Approach for Rainfall Pattern Prediction in Umuahia Metropolis

Eunice Chioma Agina[1], Uduak David George[2*], Patience Usoro Usip[3]

[1]Department of Computer Science, Michael Okpara University of Agriculture, Umudike, Nigeria
[2*]Department of Computer Science, University of Uyo, Uyo, Nigeria
[3]Department of Computer Science, University of Uyo, Uyo, Nigeria
[2*] Corresponding author: uduakgeorge@uniuyo.edu.ng

ARTICLEINFO

Article History:

Accepted: 15 Nov 2024
Published: 05 Dec 2024

Publication Issue :
Volume 11, Issue 6
November-December-2024
Page Number :
249-265

ABSTRACT

The erratic nature of rainfall in Umuahia metropolis of Abia State, Nigeria due to the constant variations in atmospheric conditions, results in extreme weather conditions such as drought and flooding which pose dire consequences to human beings and the environment resulting in loss of lives, damage to agricultural produce and vital infrastructure. This study seeks to present a conceptual approach using a machine learning algorithm to support rainfall pattern prediction. A comprehensive reviews of related works was carried out on Artificial Neural Network (ANN), Support Vector Machine (SVM), Multilinear Regression (MLR) and Extreme Gradient Boosting (XGBoost) applications in prediction. The min-max normalization technique was deployed to render the dataset in a common normalized data range. A 4-10-1 architecture of a Multilayer Perceptron (MLP) was designed with four nodes at the input layer, ten nodes at the hidden processing layer, and one node at the output layer for rainfall pattern prediction. Implementation of this study with real data and its comparison with other machine learning algorithms are highly recommended for further study in this domain.

Keywords : Rainfall Pattern, Artificial Neural Network, Multilayer Perceptron, Prediction, Machine Learning.

INTRODUCTION

In recent times there is a global occurrence of rainfall variability affecting several countries especially the developing ones, resulting in extreme weather conditions such as flooding and drought which lead to destruction of lives, properties, infrastructure, and crops. Extreme weather conditions also contribute negatively on human lives, agriculture, water resources management, hydroelectric power supply, and economic growth of certain geographical locations. It adversely affects proper decision making by individual farmers and the government. This study seeks to conceptualize a rainfall pattern prediction system using a machine learning approach, specifically, the Multilayer Perceptron (MLP) in Umuahia metropolis. Umuahia is the capital city of Abia State, a South Eastern state in Nigeria, renowned for hosting a popular agricultural research institute (National Root Crops Research Institute) and Nigeria's most famous commercial city, Aba.

Variations in rainfall patterns are evidence of climate change which could be natural or human-induced through a continuous use and burning of fossil fuel as well as urbanization, thereby contributing largely to an increase in the amount of greenhouse gases released into the atmosphere, leading to global warming [1], [2]. Rainfall is defined as the amount of rain water falling within a specific area at a given time [3]. Its process involves evaporation of water into the atmosphere and condensation of water leading to water droplets. It is the main source of water among the different weather parameters for the hydrological cycle and it is critical in the management and planning of irrigation and agricultural projects. Moreover, the availability of water and sustainable food production can be affected by a change in rainfall distribution [4], [5]. Rainfall distribution in Nigeria has two distinct seasons namely: the dry season (November
to March) and the rainy season (April to October) [6]. Nigeria receives substantial amounts of rainfall during the rainy season with a mean annual total ranging between 1200 mm and 2000 mm [5].

Environmental factors affecting the existence and distribution of rainfall include temperature, wind speed, relative humidity, sunshine, pressure, and evaporation [7]. Climate change also has substantial effect on rainfall patterns [8], as well as on other hydro meteorological variables such as temperature and evaporation [9]. Thus, understanding rainfall patterns is fundamental to evaluating changes in climate condition that often result in extreme weather conditions such as floods and drought [10]. Reports have shown that human related activities such as deforestation, constant use and burning of fossil fuel, industrialization, water pollution and agricultural activities have contributed to the reduction of the amount of carbon absorbed by the atmosphere leading to variations in climate system [11].

Over the years, researchers have adopted different methods to analyze historical rainfall data in order to uncover hidden patterns and trends that can be used to predict future occurrences of climate change. These methods rely on statistical and linear models which have limitations in accuracy. Thus, there is need for advanced machine learning models to accurately predict the nonlinear nature of rainfall pattern change. Accurate prediction of rainfall patterns is critical for increasing agricultural productivity for food security and quality water supply and for managing the implication of rainfall on the society [7]. Since rainfall affects human lives and developments in several ways, estimating the amount of rainfall correctly is essential in managing the effects of rain on agriculture, water resources management, hydroelectric power supply, economic growth and development and management of flood [12], [13].

Due to the spatiotemporal variation in the distribution of rainfall, understanding and modeling rainfall patterns becomes complex and difficult 14]. Also, the complexity of rainfall process and the continuous change in climate conditions make accurate prediction both challenging and demanding [15]. Over the years, researchers have focused on predicting rainfall patterns of Umuahia metropolis based on the frequency of rainfall [13], [16], others attempted to predict rainfall patterns of the area based on rainfall amount which used limited data and different linear and nonlinear approaches [17]. Linear models such as Autoregressive Integrated Moving Average (ARIMA) [12], [13], Seasonal Autoregressive Integrated Moving Average (SARIMA) [18], [19] and Fractional Autoregressive Integrated Moving Average (FARIMA) [20] are based on assumptions. They have limitations in handling the non-linear relationship in rainfall data [21] and are unable to capture extreme values, as well as other constraints in rainfall dynamics, making accurate predictions difficult [22].

To overcome limitations of the linear models, machine learning approach were adopted. Machine learning models are algorithms that learn from data without the need of human intervention [23]. They offer a dynamic and flexible approach in modelling highly nonlinear rainfall data and capture their extreme values. They are also capable of independently solving intellectual tasks traditionally solved by humans using complex mathematical and statistical tools [24]. Machine learning models have played significant roles in rainfall pattern prediction. Machine learning models for modeling rainfall data reported in literatures include: Multiple Linear Regression [17], Support Vector Machine (SVM) [25], [26], Extreme Gradient Boosting [27], [28], Artificial Neural Network [3], [29]. Among several machine learning models, Artificial Neural Networks (ANN) have proven to be reliable in modeling complex and non-linear rainfall data. ANN is capable of modeling the non-linearity of rainfall data, making them a good choice for rainfall pattern forecasting [30]. Neural Networks have been used extensively in pattern-recognition, image processing, data mining, regression/approximation of functions, and information processing [30], [31]. Multilayer perceptron (MLP), a most commonly used ANN, is a Feedforward Neural Network architecture that has the structure of the artificial neuron called the Rosenblatt perceptron [30]. The MLP architecture is structured into the input, hidden and output layers. The input layer receives input signal, the hidden layers process them and the output layer produce output from the network. Each neuron in a layer is connected to neurons in other layers through weighted connections [32]. Because of the structure of the MLP, it is capable of handling the nonlinearity of data to give a more meaningful relationship within the data elements.

The rest of the study are as follows. Section II presents an extensive literature reviews of related prediction models while Section III presents the methodology of the study. In Section IV, conclusion of the study is presented.

LITERATURE REVIEW

In [17] the performance of different forecasting models in forecasting rainfall data of Umudike was compared. The research employed Triple Exponential Smoothing (Holt-Winters Method), Multiple Linear Regression and SARIMA for the comparison, using Mean Error (ME), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Friedman Statistic test to assess the models' performances. Results indicated that the Multiple Linear Regression had the best result. [33] proposed the application of Multilinear Regression for weather forecast. The study was based on daily meteorological data of Uttar Pradesh recorded at different intervals for the period of four years and consisted attributes such as station, latitude, longitude, altitude, time, date, air temperature, wind speed, wind direction, humidity, atmospheric pressure, and rainfall. MLR was applied to the data to predict different weather conditions based on the inputs and when compared with other models, MLR model had the highest accuracy of 88.0% compared to other models. [34] compared the performance of Multi Linear Regression (MLR), Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) for the forecasting of daily rainfall at Ercan Airport, Northern Cyprus. The analysis utilized daily meteorological data obtained from National Aeronautics and Space Administration (NASA) for the period of 10 years. The data consisted daily record of relative humidity, minimum temperature, wind speed, maximum temperature, solar radiation and rainfall. Results demonstrated a superior performance of ANN and ANFIS than MLR models could be good-fit for rainfall prediction. [35] proposed a rainfall prediction method using the MLR. The study utilized publicly available sources of Indian rainfall data. Data preprocessing was applied to the data, and

70% of the data was used for model training while 30% was used for model testing. The Mean Square Error (MSE), RMSE, and correlation were used for model validation. The MLR was compared with other models namely the QPF and LR. Experimental results indicated that the MLR model yielded good results with the least values of MSE and RMSE and the highest values of Correlation. [36] proposed a rainfall prediction using the MLR model in order to predict the rate of precipitation (PRCP). The study aimed at predicting the rate of rainfall for Khartoum state. The data for the study consisted of temperature, wind speed, and dew point obtained from the website of the National Climatic Data Center. Training was carried out using the Pytorch library in python programming language. A comparative analysis of the average value MSE of the training data with the test data was done to estimate the performance of the model. The results obtained showed that when the same amount of data was used for both the training and test phases the average of the Mean Square Error (MSE) improved by 85% during test time. However, when the amount of data at the test phase exceeded the amount of training phase, the MSE value dropped to 59%. [5] studied and analyzed the trend in rainfall variability for South Eastern region of Nigeria, rainfall data from 1922-2008 were analyzed using non-parametric Mann-Kendall test. Experimental result indicated that Umuahia had the lowest rainfall trend rate of -0.1153 mm per year which implies a constant decrease in rainfall amount for the period under study. [37] contributed to the knowledge of rainfall trend in Nigeria by assessing the trends of rainfall in Abuja, Nigeria's capital territory, using daily rainfall record of 31 years and the result showed a downward trend in the rainfall amount received in Abuja over the period under study. [12] analyzed and forecasted rainfall patterns in the Manga-Bawku area of northeastern Ghana using rainfall data of forty years from 1976 to 2016. The researchers aimed to forecast rainfall patterns for the next nineteen years, thus they employed the Simple Seasonal Exponential Smoothing (SSES) for the analysis and ARIMA (0,1,1) model for the forecast. Their study showed a downward trend of rainfall pattern in the next 19 years with a Stationary R2 of 0.698 and 0.669, RMSE of 48.775 and 50.717, and normalized Bayesian Information Criteria (BIC) of 7.800 and 7.904 for both models. The researchers concluded that due to the downward trend in rainfall pattern, adequate measures should be made for good water resources management. [22] compared the performance of three artificial neural networks (ANN), four adaptive neuro-fuzzy inference system (ANFIS) and five support vector machine (SVM) algorithms in predicting monthly and annual rainfalls. The data for the experiment were obtained from the Nigerian Meteorological Agency (NIMET), Lagos for the period of 31 years (1983–2013). Geoclimatic coordinates were selected as inputs to the models. Experimental result showed that the adaptive neuro-fuzzy inference system (ANFIS) models outperformed the rest of the models. [3] utilized the Multilayer Perceptron to forecast week-ahead rainfall using historical rainfall data of Mindanao, Philippines. Their work was motivated by the inability of the Automated Weather Station (AWS) system to provide accurate rainfall forecasting which leads to inefficient management of flooding in the area. The researchers aimed to develop a system that could provide timely and accurate week-ahead rainfall forecast that would serve as a warning system for impending natural disasters. They implemented two MLPNN having 11 input neurons of different weather variables including average temperature, minimum temperature, maximum temperature, average wind speed, maximum wind speed, relative humidity, total rainfall, visibility, the date, month, and year. The first MLPNN model with 50 hidden neurons, trained with Scaled Conjugate Gradient (SCG) yielded a 0.01297 MAE and 0.01512 RMSE. The second MLPNN model with 100 hidden neurons trained with SCG training algorithm and Sigmoid activation function yielded 0.01388 MAE and 0.01557 RMSE. [13] compared the performance of Seasonal Autoregressive Integrated Moving Average (SARIMA) model and Seasonal Artificial Neural

Network (SANN) in forecasting the frequency of rainfall in Umuahia, Abia State. The aim of their study was to contribute to a debate that claims that neural networks were more efficient than statistical models in modelling rainfall time series data. The researchers obtained ten years rainfall data spanning from 2006 to 2016 from the National Root Crop Research Institute, Umudike, Abia State. They used the Forecast Error (FE), Mean Forecast Error (MFE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as metrics to evaluate the performance of both models. Results from their experiment showed no significant variation in the metric values from both models. The researchers however concluded that both models can successfully model rainfall data. [38] evaluated and compared the performances of Back Propagation Neural Network (BPNN), Support Vector Regression (SVR) and Long-Short Term Memory Network (LSTM) in predicting rainfall intensity one month ahead. They aimed to predict the occurrence of landslide in advance. Rainfall data from 1901-2015 obtained from IMD (Indian Meteorological Department) Narendra Nagar, Uttarakhand were used for the analysis. The data was normalized using min-max normalization to improve prediction accuracy. Experimental result showed that the BPNN outperformed the other models used by providing the least MSE, MAE and RMSE values of 0.0148, 56.9481 and 95.5339 respectively for input set 2 and least MSE, MAE and RMSE values of 0.0136, 52.3725 and 90.4859 respectively for input set 3. [39] predicted the annual and non-monsoon rainfall of Odisha using the SVR -MLP models. The study was based on the rainfall and relative humidity data collected from the Department of Forest and Environment, Government of Odisha for the period 1991-2015. The models were implemented for predicting the maximum annual rainfall and the maximum rainfall during non-monsoon season using parameters such as average temperature in month, wind velocity, humidity, and cloud cover. Evaluation metrics such as MSE (mean squared error), correlation coefficient, coefficient of efficiency and MAE (mean absolute error) were used to model assessment. The result of the experiment indicated that both models (SVR, MLP) can effectively model rainfall data of Odisha. [40] proposed a method of rainfall prediction using the Multilayer Perceptron and Auto-Encoders. The study aimed at improving agricultural production in India. The proposed model was compared with Support Vector Machine, Auto-Regressive Integrated Moving Average, Self-Organizing Map. Evaluation metrics such as MSE and RMSE were used to assess the performances of the models. Experimental results demonstrated that the proposed model was a good fit for rainfall prediction of Indian data. [41] proposed an approach for forecasting streamflow by modelling rainfall-stream flow relationship. The dataset used for the study comprised 24 years (1991–2014) rainfall data of five climate stations and stream flow dataset of seven gauging stations obtained from SWHP-WAPDA, Lahore, Pakistan. ANN model was trained using different rainfall patterns to forecast streamflow. The performance of the ANN model was evaluated using Root-Mean-Squared Error (RMSE), Correlation Coefficient (R), the Coefficient of Determination (R^2), and Nash–Sutcliffe efficiency Coefficient. The results indicated that the ANN model developed by presenting rainfall patterns of the previous 4 days can precisely predict the daily streamflow with 0.97 and 0.94 value of R^2 for the validation and test period, respectively. [42] examined the performance of Response Surface Regression (RSR) when compared with ANN and MLR for rainfall prediction. The study was based on the rainfall data of Northern Cyprus from 2011 to 2017. Input to the models included geographical coordinates such as latitude, longitude, and altitude of the location as well as meteorological parameters such as average temperature, maximum temperature, minimum temperature and relative humidity. Coefficient of determination (R^2), root mean squared error, Nash–Sutcliffe Efficiency (NSE), and Willmott's index of agreement were used to assess the models' performances. Comparative analysis showed that the three models demonstrated good performance, but the ANN had the best result. Also, the RSR model outperformed the MLR in representing the relationship between the geographical coordinates, meteorological parameters, and rainfall data. [29] compared the performance of ANN and the traditional SARIMA model in predicting rainfall over Nigeria. The researchers used monthly rainfall data from January 1991 to December 2020. According to the researchers, SARIMA was identified by the ACF and PACF plots as appropriate while ANN trained with Levenberg-Marquardt (LM) was chosen as most appropriate with an Average Absolute Error of 0.000525056 over the ANN trained with Scaled Conjugate Gradient Descent (SCGD) algorithms and Bayesian Regularization (BR) method. According to the researchers, results from their experiment indicated that ANN had a better forecasting result than the SARIMA. [43] developed ANN models using a multilayer perceptron approach to predict rainfall 2 months ahead for six geographically diverse weather stations in the Benin Republic. They used monthly rainfall data from 1959 to 2017 as input variables while rainfall data from 2018–2021 were used for testing. The performance of the MLP model was compared with that of the long short-term memory (LSTM) and climatology forecasts (CFs). The result of the experiment indicated that the MLP model outperformed the LSTM and CFs models with Nash–Sutcliffe efficiency (NSE) coefficient values ranging from 0.373 to 0.885, 0.297 to 0.875, and 0.335 to 0.845 for the MLP, LSTM and CFs respectively. [44] analyzed the relationship between rainfall and runoff using Neuro-Fuzzy (NF) and Support Vector Machines (SVM) models. The dataset was obtained from the Muskegon basin in USA and consisted of 1397 instances of daily rainfall data, temperature and runoff. Statistical measurements such as R^2, MAE and RMSE were used to evaluate the models. Results of experiment showed that SVM outperformed NF with R^2 of 0.92, RMSE of 10.26 m^3 /s and MAE of 4.88 m^3 /s respectively. [45] evaluated and compared the performance efficiency of different machine learning models namely Recurrent Neural Network (RNN), Support Vector Machine (SVM), and ANFIS techniques in modelling rainfall data of Bolangir district, Odisha. The researchers used rainfall data of 48 years for the analysis. Different architectures of RNN were tested, namely: 3-2-1, 3-3-1, 3-4-1, 3-5-1, 3-9-1 and the 3-9-1 performed better than the other architectures. The performances of the models were evaluated using MSE, RMSE, and R^2 respectively. Results indicated that the SVM had better performance based on the values of the Coefficient of Determination. SVM had an R^2 values of 0.9526 and 0.9738 for both testing and training phase while ANFIS had 0.9342 and 0.9488, and RNN gave the least values of 0.9039 and 0.9292 respectively. [46] investigated the effectiveness of the Support Vector Machine to predict rainfall 30 days ahead using map images. Daily rainfall data obtained from National Center for Environmental Prediction (NCEP) taken at 7 a.m. Eastern Standard Time, from January 2012 to October 2019 using radar images with a total of 2,835 images were utilized for the study. Each image contains 16 different rainfall intensity levels, with a size of 400×320 pixels which represents the United States. The study was based on three regions corresponding to three squares from a 5×5 grid covering the map area of the continental US. Rainfall was quantized at 3 levels: light (or no rain), moderate, and heavy rain. Prediction result showed that SVM had good performance for predicting rainfall of the regions. In conclusion, the researchers noted that the SVM showed some evidences that when applied to large-scale precipitation maps, can give useful information for predicting regional rainfall under some conditions, however, care must be taken to avoid pitfalls. [47] developed and compared several advanced machine learning models for predicting daily rainfall in Hoa Binh province, Vietnam. The developed models were Adaptive Network Fuzzy Inference System optimized with Particle Swarm Optimization (PSOANFIS), ANN and SVM. Meteorological parameters such as maximum temperature, minimum temperature, wind speed, relative humidity and solar radiation were used as input to the models to predict daily rainfall as output. Model validation was achieved using correlation coefficient (R) and Mean Absolute Error (MAE), Skill Score (SS), Probability of Detection (POD), Critical Success Index (CSI), and False Alarm Ratio (FAR). Experimental results showed that all the models yielded good predictions of daily rainfall but the SVM had the best performance.

In [48], a combination of Support Vector Machine with Stochastic Gradient Descent (SGD-SVM) to predict rainfall was proposed. The researchers aimed to replace the linear threshold used in traditional rainfall prediction activities with the new model in order to improve the accuracy of rainfall prediction. Data for the study contains daily records of one-year ground-based meteorological data obtained from Maritim Perak Station (ID WMO: 96937). The dataset consisted of weather parameters namely; atmospheric pressure, sea level pressure, wind direction, wind speed, relative humidity and precipitation with a total of 1825 instances. The principal component analysis (PCA) was used for feature selection and dimensionality reduction of the data. The researchers maintained that the proposed model could satisfactorily predict rainfall, however, no values of statistical measures were stated. [49] proposed an ensemble method of Naïve Bayes, decision tree, Support Vector Machine, Artificial Neural Network, and Random Forest for predicting rainfall of Malaysia. The dataset for the study was obtained from the Drainage and Irrigation Department, and the Malaysian Meteorological Department having 1,581 instances. The dataset had some missing values which were populated using a mean average mechanism. The mean average functions are obtained by summing all instances of an attribute that is selected and then dividing the sum by the number of records. The performance of the model was evaluated using average probability, maximum probability, and majority voting. The experimental result indicated that the ensemble model outperformed the individual models, thus, could satisfactorily predict rainfall. [50] utilized the Support Vector Regression Machine to predict heavy rainfall in Philippine. Data for the study was obtained from Philippine Department of Science and Technology - Advance Science and Technology Institute (DOST-ASTI) from 2013 to 2017. The SVRM model utilized Radial Basis Function (RBF) as the kernel function with the parameters of c = 100; g = 1; e = 0.1; p = 0.001 and the lag variable which used 12-hour report with lags up to 672-timesteps (i-672). Result from the experiment showed a Mean Square Error (MSE) of 3.461315 indicating a positive potential in modelling rainfall data. [51] predicted rainfall in Indonesia using the Support Vector Machine. The data used in the study had 758 instances from January 2021 to January 2023. The data was obtained from the Raja Haji Fi Sabilillah meteorological and geophysical station for Tanjungpinang City, Riau Islands, Indonesia. Data attributes considered in the rainfall prediction were temperature (^{o}C), humidity (%), wind speed (m/s), and rainfall (in). Experimental result show 82% precision value for rain and a 0.74 ROC curve score which indicates that the model satisfactorily model rainfall data.

[27] built a multivariate rainfall prediction model using the Extreme Gradient Boosting model. Meteorological data for the study was obtained from the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) for Tanjung Mas, Semarang City, Indonesia for 7 years. During training, the researchers omitted missing values. The data used had 11 attributes, missing values were omitted during training. The experimental result gave a training RMSE of 2.7 mm and the testing MAE of 8.8 mm which demonstrated that the model can satisfactorily produce accurate predictions for daily rainfall estimates. [7] evaluated the performance of three machine learning models (Multivariate Linear Regression, Random Forest, and Extreme Gradient Boost) in predicting daily rainfall amount in Ethiopia. The researchers used rainfall data obtained from a local meteorological office at Bahir Dar City, Ethiopia. Pearson correlation technique was used in selecting the relevant environmental variables

as input to the models. The results of their experiment indicated that the Extreme Gradient Boosting out-performed the other models. [28] predicted rainfall on a monthly scale using XGBoost model. The researchers aimed to contribute to the knowledge of rainfall prediction using the XGBoost model. The data for the analysis consisted of historical rainfall data of 30 years from 1987 to 2017 obtained from India Meteorological Department (IMD), and other Government departments of Vishakapattanam, Andhra Pradesh. The analysis was executed using the R language. Model evaluation and parameter estimation were achieved by means of autocorrelation function (ACF) and partial auto correlation function (PACF) with 95% accuracy for forecast up to 3 to 5 years. This result imply that the XG-Boost Model is a good fit to forecast the rainfall of Vishakapattanam. [52] utilized Logistic Regression, Decision Tree, Neural Network, Random Forest, LightGBM and XG Boosting models to predict rainfall in Australia. Appropriate parameters were selected for each of the models. Meteorological data from the years 1901 to 2015 were used for the study. Results from the models showed accuracy of 85.92% for decision tree, 79.51% for logistic regression, 88.40% for Neural network, 92.90% for Random Forest, 87.43% for LightGBM, and 95.79% for XG Boost. The result clearly shows that the XGBoost model had the best performance and could be effectively utilized for rainfall prediction. [53] proposed a novel method of rainfall prediction called Analogousness Enhanced Rainfall Predictor using XGBoost Backbone. The study focused on improving the accuracy of rainfall prediction using the novel method to deal with data-class imbalance issue. Large datasets totaling 145,459, spanning ten years of observed weather data from Australia was utilized for the study. The Synthetic Minority Oversampling Technique (SMOTE) was employed to deal with data imbalance between the data classes (“Yes”, “No”) while the K-fold Cross-Validation Technique was used to find the best value of the XGBoost parameters. Experimental result demonstrated accuracies of 83.00%, 83.07%, 84.91%, 79.28% and 83.29% for decision tree, random forest, XGBoost, Weighted XGBoost and Analogousness Enhanced Rainfall Predictor utilizingXGBoost Backbone classifiers. [54] predicted daily rainfall of Indian states using the XGBoost and Random Forest approach. The study aimed to contribute to the improvement of agricultural produce for food security. The implementation of the approach was based on daily weather data of Bahir Dar City, Ethiopia. The data contained different variables namely, year, month, date, evaporation, daylight, maximum temperature, minimum temperature, humidity, wind speed, and rainfall. Pearson correlation technique was used to select the relevant atmospheric variables used as input to the machine learning models. The RMSE and MAE were utilized for model evaluation. Results obtained show that the Extreme Gradient Boosting machine learning technique outperformd the Random Forest. [55] proposed the forecast of Australian rainfall using random forest (RF), logistic regression (LogReg), Gaussian Naive Bayes (GNB), k-nearest neighbours (kNN), support vector classifier (SVC), and XGBoost (XGB). The data was pretreated using missing value imputation and feature selection. For evaluation purposes, cross-validation and performance indicators such as accuracy, precision, recall, and F1-score were utilized. Evaluation results indicated that the RF and XGB models performed best, with accuracy ratings of 87% and 85%, respectively. However, GNB and SVC models had the poorest performance with accuracy ratings below 70%. [56] examined the implementation of several machine learning models for rainfall forecasts using Austin weather data. The machine learning models were: Extreme Gradient Boosting, Support Vector Machine, Long Short-Term Memory, and Random Forest models. With a total of 1319 records containing 21 variables, namely: date, temperature, dew point, humidity, sea level pressure, visibility, high miles, wind speed, precipitation sum, and events were obtained. However, only 19 variables were used with precipitate range added for the investigation. Rainfall was classified into different classes: “no rain,” “small rain,” “moderate rain,” and “heavy rain” based on specific value ranges. Extreme Gradient Boosting had the best results of 85.17% accuracy, 83.19% F1 score, 85.17% recall score, and 82.14% precision score respectively.

METHODOLOGY

The study carries out extensive reviews of literature, designs the system architecture, data preprocessing, data normalization, data transformation and designs the MLP architecture.

A. System Architecture

The designed system architecture for the Rainfall Patterns Prediction for Umuahia Metropolis is shown in Fig. 1, consisting of several modules namely: Database, Data Preprocessing, Model Design, Model Implementation and Performance Evaluation.

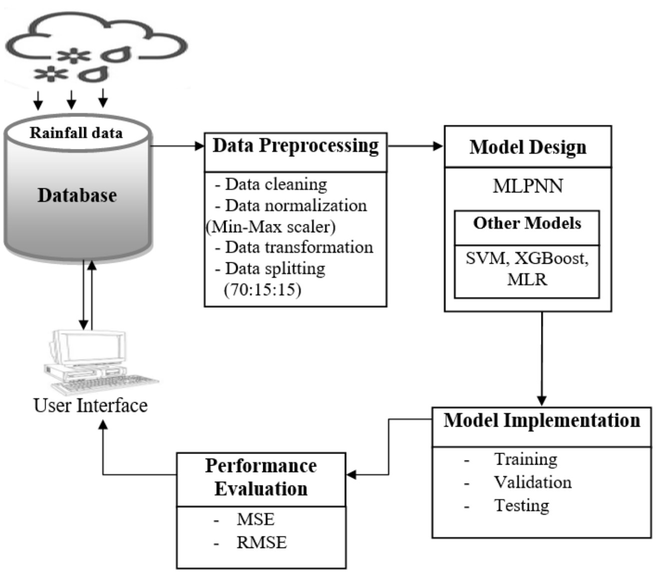

Fig. 1 Architecture of the Rainfall Patterns Prediction System

Database:

Rainfall data is stored in the database from where the necessary dataset is supplied to the system. The system would be able to fetch data from it and also save the results of processing to it. It consists of the following schema in Table 1.

TABLE 1. DATABASE SCHEMA

Attribute	Data Type	Description
Month	Text	Month rainfall occurred
Year	Text	Year rainfall occurred
Amount	Numeric	Amount of rainfall in millimeter (mm)
Location	Text	Location
Altitude	String	Station altitude
Latitude	String	Station latitude
Longitude	String	Station longitude

Data Preprocessing

The goal of data preprocessing is to ensure that correct data format is given as input to the machine learning model. The activities here include elimination of missing data, outliers, and redundant data as well as preventing out of format data and every form of noise in the dataset.

Data Normalization: For machine learning to take place, the data points need to be normalized to a predetermined range, using the min-max normalization given in equation 1 ([57], [58], [59], [60]) as:

$$x' = \left(\frac{x - oldMin}{oldMax - oldMin}\right) * (newMax - newMin) + newMin \qquad (1)$$

where x' is the min-max normalized data sample, x is the original data sample, oldMin is the minimum data among any attribute of the original dataset, oldMax is the maximum data among any attribute of the original dataset, newMin is the minimum of the normalized dataset, and newMax is the maximum of the normalized dataset. The newMin and newMax are the defined new minimum and maximum range in the normalized dataset. Where newMin = 0 and newMax = 1, the min-max normalization model transformed to equation 2 as follows:

$$x' = \left(\frac{x - oldMin}{oldMax - oldMin}\right) \qquad (2)$$

Thus the data is presented within the range [0, 1].

Data Transformation

Since rainfall data is a time series data, the mean monthly rainfall data used in the study is modelled as a time series data which involves sequential recorded observations of rainfall over time. The study is focused on a long-term forecast with one year ahead prediction. The rainfall time series is represented as follows [61]:

$$Y(t+1) = f[x(t), x(t-1), \ldots, x(t-s+1)] \qquad (3)$$

where $Y(t+1)$ is the predicted value, f is a function that describes the prediction model, t is the time series and s is the size of input values. Four input values were selected, the current month and three historical months as input samples. The current value helps to keep an up-to-date record of rainfall pattern while the historical values keep track of the pattern.

The input vectors are represented as:
$[x(t-4); x(t-3); x(t-2); x(t-1)]$
The output vector is represented as: $Y(t+1)$.

The rainfall data attributes with their corresponding codes are presented in Table 2.

TABLE 2
INPUT PARAMETERS FOR RAINFALL PATTERNS PREDICTION

S/N	Parameter	Code
1	Historical Rainfall of Last	RFLTHM
2	Three Months	RFLTM
3	Historical Rainfall of Last	RFLM
4	Two Months	RFCM
	Historical Rainfall of Last	
	one Months	
	Historical Rainfall of	
	Current Month	

MLP Model Design

Building of a rainfall pattern prediction model helps to accurately predict monthly rainfall amount twelve months ahead using past recorded rainfall data as predictors. This study adopts the ANN model using the MLP approach in predicting monthly rainfall patterns of Umuahia metropolis, Abia State, Nigeria. The designed MLP is a three-layer Feedforward Neural Network with four (4) nodes at the input layer (representing the rainfall input data attributes), one (1) node at the hidden layer with varied number of neurons, probably between 10, 20 and 30 (which receive and process the signals from the input layer using the connection weights and hyperbolic transfer function) and one (1) node at the output layer (representing the predicted rainfall amount generated through the connection weights of the hidden layer using sigmoid transfer function). Decision of the best number of neurons in the hidden layer is made based on the values of the Mean Squared Error (MSE) and Regression Correlation (R) of the training, validation and test values. The computed output from the network is compared with the desired output and the difference, which is the error term, is used to adjust the connection weights by means of a back-propagation algorithm. This process is repeated until the error term is within the acceptable threshold. The proposed MLPNN design for rainfall patterns prediction is shown in Figure 2.

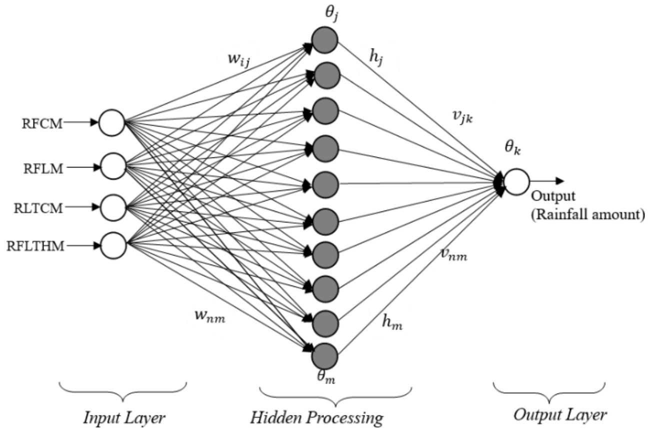

Fig. 2. Structure of the Designed MLP with 10 Hidden Neurons

The MLP neural network has the following layers:

a. Input Layer x_i: $i = 1,2, \dots, n$; where n = 4. From the architecture, $t-1$ is the rainfall input data for current month, -2 , $t-3$, $t-4$ are the rainfall input data for the next three historical months respectively.

b. Hidden Processing layer h_j: $j = 1,2, \dots m$; where m = 10.

c. Output layer o_k: $k = 1$

where n is the number of nodes in the input layer and m is the number of nodes in the hidden layer and k is the number of nodes in the output layer.

The outputs of the input layer I_i, hidden layer (h_j) and output layer (o_k) are obtained as follows (Acharya et al., 2003):

$$I_i = \sum_{i=1}^{n} w_{i,j}\, x_i, \quad i = 1,2, \dots, n \tag{4}$$

$$h_j = f\left(\sum_{i=1}^{n} w_{i,j}\, x_i - \theta_j\right), \quad j = 1,2, \dots, m \tag{5}$$

$$o_k = f\left(\sum_{j=1}^{m} v_{j,k}\, h_j - \theta_k\right), \quad k = 1,2, \dots, p \tag{6}$$

where f is the activation function of each neuron, $w_{i,j}$ is the weight connecting the input layer i and the hidden layer j, x_i is the input vector from the ith input layer nodes, $v_{j,k}$ is the connection weight between the hidden layer j and the output layer k, h_j is the output of the hidden layer while θ_j and θ_k are the bias terms of the hidden and the output layers respectively.

The error terms of the output layer (e_k) and the hidden layer (e_j) are computed as follows [62]:

$$e_k = o_k(1 - o_k)(d_k - o_k) \tag{7}$$

where d_k is the desired output.

$$e_j = h_j(1 - h_j) \sum_{j=1}^{m} w_{k,j}\, e_k \tag{8}$$

The weights at output and hidden layer nodes are updated using Equation 9 while the weight between the hidden and input layer nodes is updated using Equation 10.

$$w_{jk} = w_{j,k} + \eta h_j e_k \tag{9}$$

$$w_{ij} = w_{i,j} + \eta x_i e_j \tag{10}$$

where η is the learnng rate.

The bias at the output layer and hidden layer is updated based on Equations 11 and 12.

$$\theta_k = \theta_k + \eta e_k \tag{11}$$

$$\theta_j = \theta_j + \eta e_j \tag{12}$$

A system of equations formulated for the input and hidden

layers is as follows:

$$\sum_{j=1}^{m} \sum_{i=1}^{n} w_{i,j} x_i = h_j^* \tag{13}$$

where W_{ij} is the matrix of weights on the connection from the jth node in the hidden layer to the ith node in the input layer, x_i is the rainfall patterns input vector and h_j^* is the hidden layer pre-output (output to be acted on by transfer function). The pre-output is subjected to hyperbolic transfer function to obtain the actual output. The hyperbolic transfer function is given in [63] as:

$$h_j = \frac{e^{h_j^*} - e^{-(h_j^*)}}{e^{h_j^*} + e^{-(h_j^*)}} \tag{14}$$

Similarly, the output from the output layer is computed as shown in Equation 3.15.

$$\sum_{k=1}^{p}\sum_{j=1}^{m} v_{j,k} h_j = o_k^* \qquad (15)$$

where $V_{j,k}$ is the weight connecting the k^{th} output layer node to the j^{th} hidden layer node and o_k^* is the pre-output value which is further subjected to a transfer function to obtain the actual output using Equation 3.16 (George et al., 2018).

$$o_k = \frac{1}{1 + e^{-(o_k^*)}} \qquad (16)$$

C. Model Evaluation

The performance of the model will be evaluated by comparing values of the predicted values with the target values using the Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) given in equations (17), (18), and (19) respectively. The MSE measures the average squared difference between the predicted values and the actual observed values. The RMSE evaluates how well predicted values fit the original data as well as represents the standard deviation of errors in the reconstructed values. It is the square root of the average of squared differences between predicted and target data. Its value is expected to be small (near zero) for a good model performance. The Coefficient of Determination (R^2) measures the proportion of error variation in the reconstructed data by the model. Its value is between 0 and 1, a number close to 1 shows a good model performance while a number close to zero (0) indicates a poor model performance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \qquad (17)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \qquad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})^2} \qquad (19)$$

where n is the number of observations, x_i is the original rainfall data and $\hat{x}_i$ is the reconstructed rainfall data.

D. Rainfall Prediction Workflow

The workflow diagram in Figure 3 outlines the procedure for executing the rainfall patt ern prediction system. The raw dataset would be preprocessed to render it in a form suitable and easier for machine learning activities to take place. Preprocessing include activities such as data cleaning, data normalization, data transformation and data splitting. The dataset would be split into training and testing data, where the training data takes a greater proportion of the data to build a more reliable prediction model from it. In most cases a portion of the data is also set aside for validation in order to guard against over-fitting of the model during the training process. The recommended ratio is 70: 15: 15 or 80: 10: 10 for training, testing and validation respectively.

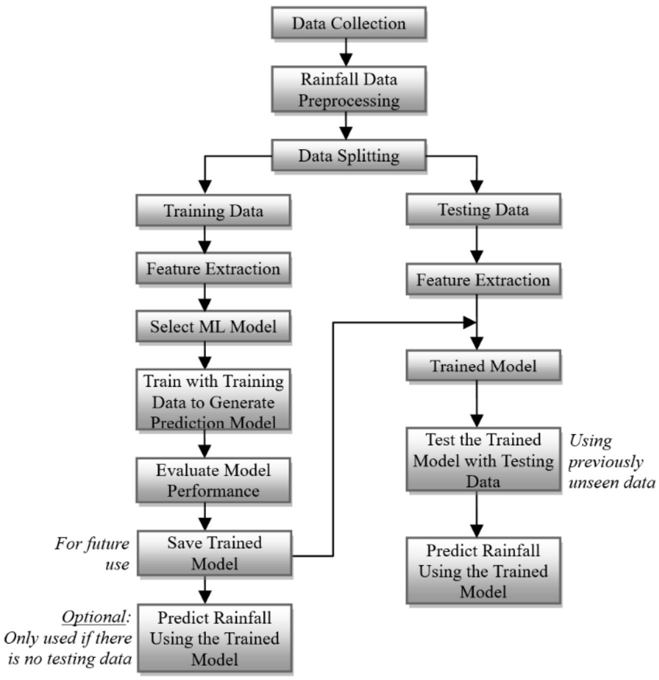

Fig. 3 System Work Flow Diagram

Subsequent upon the generation of the prediction model and its performance evaluation to be of acceptable level, the model is saved for future use and can be used to implement the prediction if the dataset is so small such that there is no testing data. The saved model is utilized in the testing phase on the dataset the model has not previously seen. This gives a more reliable prediction.

CONCLUSION

Consequent upon the erratic rainfall pattern in Umuahia metropolis of Abia State, Nigeria, and its resulting negative effects on humans and the environment, the research seeks to conceptualize a machine learning approach to assist in the prediction of rainfall in the area. The Multilayer Perceptron (MLP) served as the ML algorithm and a 4-10-1 MLP architecture was designed to prepare the ground for possible implementation with real data. Other ML algorithms are recommended to be implemented for performance comparison purpose. The use of real data on this conceptual design is recommended for future work on this domain.

Conflict of Interest: The authors declare that there is no conflict of interest.

Source of Funding: There is no external funding for this research.

REFERENCES

[1] Abe, A. O., Adeniji, Q. A., Rabiu, J. A., Adegboyega, O., Raheem, I. O., Rasaki, M.G., Sada, S. M., and Fidelis, L. (2022). Statistical Analysis and Forecasting of Rainfall Patterns and Trends in Gombe North-Eastern Nigeria. Iraqi Journal of Physics, 20 (2), 64-77.

[2] UNFCCC (2007). Climate Change: Impacts, Vulnerabilities and Adaptation in Developing Countries.

[3] Velasco, L. C. P., Serquiña, R. P., Zamad, M. S. A. A., Juanico, B. F., and Lomocso, J. C. (2019). Week-ahead Rainfall Forecasting Using Multilayer Perceptron Neural Network. Procedia Computer Science 161 (2019) 386–397. https://www.doi:10.1016/j.atmosres.2020.10484

[4] Khaing W. M. and Thinn T. N. (2008). Optimum Neural Network Architecture for Precipitation Prediction of Myanmar, World Academy of Science, Engineering and Technology, 48, 130 – 134.

[5] Ekwueme, B. N. and Agunwamba, J. C. (2021). Trend Analysis and Variability of Air Temperature and Rainfall in Regional River Basins. Civil Engineering Journal, 7(5), 816–826.

[6] Ibebuchi, C. C. and Abu, I. (2023). Rainfall Variability Patterns in Nigeria During the Rainy Season. Scientific Reports, 13:7888

[7] Liyew, C. M. and Melese, H. A. (2021). Machine Learning Techniques to Predict Daily Rainfall Amount. Journal of Big Data 8, 153, 1-12.

[8] Mislan, M., Haviluddin, H., Hardwinarto, S., Soeparto, S. M. and Aipassa. M. I. (2015). Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan – Indonesia. International Conference on Computer Science and Computational Intelligence (ICCSCI 2015). Procedia Computer Science 59 (2015), 142 – 151.

[9] Falaiye O. A., Olaitan A. G., and Nwabachili S. C. (2021). Parametric Analysis of Rainfall Variability Over Some Selected Locations in Nigeria. International Journal of Climate Research, 5(1), 35-48.

[10] Idowu, A. A., Ayoola, S.O., Opele, A.I., Ikenweiwe, N.B (2011) Impact of climate change in Nigeria. Iranica Journal of energy and environment 2(2): 145-152.

[11] Houghton J. T., Meiro Filho L. G., Callander B. A., Harris N., Kattenburg A., and Maskell K. (1996). Climate change 1995: the science of climate change. Climatic Change 584.

[12] Dankwa, P., Cudjoe, E., Amuah, E. E. Y., Kazapoe, R. W., and Agyemang, E. P. (2021). Analyzing and forecasting rainfall patterns in the Manga-Bawku area, northeastern Ghana: Possible implication of climate change. Environmental Challenges 5 (2021) 100354.

[13]Nwokike, C. C., Offorha, B. C., Obubu, M., Ugoala, C. B., and Ukomah, H. I. (2020). Comparing SANN and SARIMA for forecasting frequency of monthly rainfall in Umuahia. Scientific African, 10, e00621. https://doi:10.1016/j.sciaf.2020.e00621

[14] French, M. N., Krajewski, W. F., and Cuykendall, R. R. (1992). Rainfall Forecasting in Space and Time Using Neural Network, J. Hydrol., 137, 1–31.

[15] Abdulkadir, T. S., Salami, A. W., and Kareem, A. G (2012). Artificial Neural Network Modeling of Rainfall in Ilorin, Kwara State, Nigeria. Journal of Research Information in Civil Engineering, 9(1), 108-120.

[16] Offorha, B. C., Uche-Ikonne, O. O., Akpanta, A. C., Chikezie, D. C., and Okoye N. N. (2018). Comparison of Forecasting Methods for Frequency of Rainfall in Umuahia, Abia State, Nigeria. European Journal of Statistics and Probability, 6(1), 1-15.

[17] Ekpenyong, E. J., Okereke, E. W., and Udeagbara, C. (2019). A Comparison of the Forecasting Models of Rainfall Data of Umudike, Abia State Nigeria. International Journal of Basic Science and Technology, 5(1), 49 – 57.

[18] Adams, S. O. and Bamanga, M. A. (2020). Modelling and Forecasting Seasonal Behavior of Rainfall in Abuja, Nigeria; A SARIMA Approach. American Journal of Mathematics and Statistics, 10(1): 10-19.

[19] Kokilavani, S., Pangayarselvi, R., Ramanathan, S. P., Dheebakaran, Ga, Sathyamoorthy, N. K., Maragatham, N. and Gowtham, R. (2020). SARIMA Modelling and Forecasting of Monthly Rainfall Patterns for Coimbatore, Tamil Nadu, India. Current Journal of Applied Science and Technology, 39(8): 69-76.

[20] Montanari, A., Rosso, R., and Taqqu, M. S. (1996). Some long-run properties of rainfall records in Italy, J. Geophys. Res. 101 (D23) (1996) 29431–29438.

[21] Dada, E.G., Yakubu, H.J. and Oyewola, D.O. (2021). Artificial Neural Network Models for Rainfall Prediction. European Journal of Electrical Engineering and Computer Science. 5, 2 (Apr. 2021), 30–35. https://doi.org/10.24018/ejece.2021.5.2.313.

[22] Ojo, O. S. and Ogunjo, S. T. (2022). Machine learning models for prediction of rainfall over Nigeria. Scientific African 16 (2022) e01246.

[23] Usman, M. N. (2021). Rainfall Prediction for Minna Metropolis Using Artificial Neural Network. MSC Thesis, Federal University of Technology, Minna, Nigeria

[24] Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M., and Tornatore, M. (2018). An Overview on Application of Machine Learning Techniques in Optical Networks. IEEE Communications Surveys & Tutorials, 1–1. https://www.doi:10.1109/comst.2018.2880039

[25] Mohanty, J. R., and Mohapatra, M. R. (2018). Rainfall Prediction Using Support Vector Machine (SVM). IOSR Journal of Computer Engineering (IOSR-JCE) 20(3), 06-13

[26] Rachel, N. and Lakshmi, M. (2016). Landslide Prediction with Rainfall Analysis using Support

Vector Machine. Indian Journal of Science and Technology, 9(21), 1-6.

[27] Anwar, M. T., Winarno, E., Hadikurniawati, W., and Novita, M. (2020). Rainfall prediction using Extreme Gradient Boosting. Journal of Physics: Conference Series, Volume 1869, 2nd Annual Conference of Science and Technology (ANCOSET 2020), 28 November 2020, Malang, Indonesia.

[28] Poola, K. and Sekhar, P. H. (2021). Prediction of rainfall by using extreme gradient boost (XG boost) in Vishakapattanam area, Andhra Pradesh. International Journal of Statistics and Applied Mathematics, 6(3): 83-86.

[29] Ikpang, I. N., Okon, E. J. and George, E. U. (2022). Modeling Average Rainfall in Nigeria with Artificial Neural Network (ANN) Models and Seasonal Autoregressive Integrated Moving Average (SARIMA) Models. International Journal of Statistics and Probability, 11(4), 53-62.

[30] Costa, L., Guerreiro, M., Puchta, E., Tadano, Y. S., Alves, T. A. A., Kaster, M., and Siqueira, H. V. in Tayyar, M. H. (2023), k-Means. In: LL Minku, G Cabral, M Martins & M wagner (eds) (2023). Introduction to Computational Intelligence. An IEEE Computational Intelligence Society, Chapter 9, 105 – 116. https://www.doi: 10.5281/zenodo.7537827.

[31] Leonardo Noriega (2005). Multilayer Perceptron Tutorial. Beaconside Staffordshire ST18 0DG.

[32] Anh, D. T., Dang, T. D. and Van, S. P. (2019). Improved Rainfall Prediction Using Combined Pre-Processing Methods and Feed-Forward Neural Networks. Multidisciplinary Scientific Journal, 2, 65–83; https://doi:10.3390/j2010006.

[33]Anusha, N., Chaithanya, M. S., Reddy, G. J. (2019). Weather Prediction Using Multi Linear Regression Algorithm. IOP Conf. Series: Materials Science and Engineering 590 (2019) 012034, 1-8, https://www.doi:10.1088/1757-899X/590/1/012034

[34] Abdulkadir, R. A., Ali, S. I. A., Abba, S. I., and Esmail, P. (2020). Forecasting of daily rainfall at Ercan Airport Northern Cyprus: A Comparison of Linear and Non-Linear Models. Desalination and Water Treatment, 177 (2020) 297–305.

[35] Grace, R. K., and Suganya, B. (2020). Machine Learning based Rainfall Prediction. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). https://www.doi:10.1109/icaccs48705.2020.907

[36] Ahmed, H. A. Y., & Mohamed, S. W. A. (2021). Rainfall Prediction using Multiple Linear Regressions Model. 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). https://www.doi:10.1109/iccceee49695.2021.942.

[37] Itiowe, T., Hassan, S. M. and Agidi, V. A. (2019). Analysis of rainfall trends and patterns in Abuja, Nigeria. Current Journal of Applied Science and Technology, 34(4): 1-7.

[38] Srivastava, S., Anand, N., Sharma, S., Dhar, S., and Sinha, L. K. (2020). Monthly Rainfall Prediction Using Various Machine Learning Algorithms for Early Warning of Landslide Occurrence. 2020 International Conference for Emerging Technology (INCET). https://www.doi:10.1109/incet49848.2020.9154184.

[39] Zhang, X., Mohanty, S. N., Parida, A. K., Pani, S. K., Dong, B., and Cheng, X. (2020). Annual and Non-Monsoon Rainfall Prediction Modelling Using SVR-MLP: An Empirical Study from Odisha. IEEE Access, 8, 30223–30233. https://doi:10.1109/access.2020.2972435

[40] Basha, C. Z., Bhavana, N., Bhavya, P., and V, S. (2020). Rainfall Prediction using Machine

Learning & Deep Learning Techniques. Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020), 92- 97. https://www.doi:10.1109/icesc48915.2020.91558

[41] Ali, S., and Shahbaz, M. (2020). Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. Modeling Earth Systems and Environment. https://www.doi:10.1007/s40808-020-00780-3

[42] Kassem, Y., Gökçekuş, H., Çamura, H., and Esenel, E. (2021). Application of Artificial Neural Network, Multiple Linear Regression, and Response Surface Regression models in the Estimation of Monthly Rainfall in Northern Cyprus. Desalination and Water Treatment, 215 (2021) 328–346. https//www.doi: 10.5004/dwt.2021.26525

[43] Aïzansi, A. N., Ogunjobi, K. O., and Ogou, F. K. (2024). Monthly rainfall prediction using artificial neural network (case study: Republic of Benin). Environmental Data Science 3: e11, 1–25. https://doi:10.1017/eds.2024.10

[44] Taşar, B., Unes, F., and Varcin, H. (2019). Prediction of the Rainfall – Runoff Relationship Using Neuro-Fuzzy and Support Vector Machines. Air and Water – Components of the Environment Conference Proceedings, Cluj-Napoca, Romania, 237-246, https://www.doi: 10.24193/AWC2019_24.

[45]Samantaray, S., Tripathy, O., Sahoo, A., and Ghose, D. K. (2020). Rainfall Forecasting Through ANN and SVM in Bolangir Watershed, India. Smart Intelligent Computing and Applications, Smart Innovation, Systems and Technologies 159, 767-774

[46] Hussein, E., Ghaziasgar, M. and Thron, C. (2020). Regional Rainfall Prediction Using Support Vector Machine Classification of Large-Scale Precipitation Maps. arXiv:2007.15404v1 [cs.LG] 30 Jul 2020

[47] Pham, B. T., Le, L. M., Le, T.-T., Bui, K.-T. T., Le, V. M., Ly, H.-B., and Prakash, I. (2020). Development of advanced artificial intelligence models for daily rainfall prediction. AtmosphericResearch,104845.

[48] Hapsari, D. P., Utoyo, M. I. and Purnami, S. W. (2020). A Prediction of Rainfall Data Based on Support Vector Machine with Stochastic Gradient Descent. IICST2020: 5th International Workshop on Innovations in Information and Communication Science and Technology, Malang, Indonesia

[49] Sani, N. S. Rahman, A. H. A., Adam, A., Shlash, I., and Aliff, M. (2020). Ensemble Learning for Rainfall Prediction. International Journal of Advanced Computer Science and Applications, 11(11), 153-162.

[50] Velasco, L. C., Aca-ac, J. M., Cajes, J. J., Lactuan, J. L. and Chit, S. C. (2022). Rainfall Forecasting using Support Vector Regression Machines. International Journal of Advanced Computer Science and Applications, 13(3), 231-237.

[51] Hayaty, N., Kurniawan, H., Rathomi, M. R., Chahyadi, F. and Bettiza, M. (2023). Rainfall Prediction with Support Vector Machines: A Case Study in Tanjungpinang City, Indonesia. BIO Web of Conferences 70, 01003 https://doi.org/10.1051/bioconf/20237001003

[52] Mahesh, V. S., Ramya, R., Vardhan, P. A., Kumar, T. V. A., and Krishna, M. B. (2023). Rainfall Prediction Using Machine Learning Algorithms. International Journal of Research Publication and Reviews, 4(10), 940-945.

[53] Meti, G. and Ravi Kumar. G. K. (2023).Analogousness Enhanced Rainfall Predictor using XGBoost Backbone. International

Journal of Intelligent Systems and Applications in Engineering, 11(2), 329-335.

[54] Varun, K. P., Reddy, M. M., Sushanth, T., Kumar, R. S. and Kumar, T. S. (2023). Real Time Rainfall Prediction for Indian States Using XGBoost and Random Forest Approach. International Journal of Creative Research Thoughts (IJCRT), 11(4), 672-679

[55] Vinta, S. R. and Peeriga, R. (2024). Rainfall Prediction using XGB Model with the Australian Dataset. EAI Endorsed Transactions on Energy Web Research Article, 11, 1-4.

[56] Tin, T. T., Sheng, E. H. C., Xian, L. S., Yee, L. P., and Kit, Y. S. (2024). Machine learning classification of rainfall forecasts using Austin weather data. International Journal of Innovative Research and Scientific Studies, 7(2), 727-741.

[57] Ali, P. J. M. (2022). Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements. X(1), 85–91. http://dx.doi.org/10.14500/aro.10955

[58] Henderi, Wahyuningsih, T., & Rahwanto, E. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. 4(2), 13–20.

[59] Kamani, G. J., Parmar, R. S., & Ghodasara, Y. R. (2019). DATA NORMALIZATION IN DATA MINING USING GRAPHICAL USER INTERFACE: A PRE-PROCESSING STAGE. 30(2), 106–109.

[60] Patro, S. G. K., & Sahu, K. K. (2015). Normalization: A Preprocessing Stage. 2(3), 20–22. https://doi.org/10.17148/IARJSET.2015.2305

[61] Eyo, I., Eyoh, J., and Umoh, U. (2021). On the Prediction of COVID-19 Time Series: An Intuitionistic Fuzzy Logic Approach. Journal of fuzzy extension and application, 2(2), 171-190.

[62] Akpan, M. G., George, U. D. and Onwe, D.N. (2023). Machine Learning Approach for Classification of Dalium Guineense Fruits. World Journal of Applied Science and Technology, 15(1), 38-42

[63] George, U. D., Inyang, U. G. and Akinyokun, O. (2018). Conceptualization of Intelligent Clustering Methodology for Terrorist Acts Classification in Nigeria, International Journal of Current Advanced Research, 07(7), 14200-14208. DOI: http://dx.doi.org/10.24327/ijcar.2018.14208.2567.