CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2024.297

Evaluating the performance of Random Forest, Decision Tree, Support Vector Regression and Gradient Boosting for streamflow prediction

Osahon Idemudia, Jacob Odeh Ehiorobo, Osadolor Christopher Izinyon, and Idowu Rudolph Ilaboya*

Department of Civil Engineering, Faculty of Engineering, University of Benin, Nigeria

**Corresponding author (rudolph.ilaboya@uniben.edu)*

Article info.

Received 29 Nov 2023
Revised 29 Feb 2024
Accepted 28 Mar 2024

Keywords

Machine Learning, Random Forest (RF), Decision Tress (DT), Support Vector Regression (SVR), and Gradient Boosting (GB)

ABSTRACT

This study utilized a range of machine learning algorithms to predict the hourly streamflow in the Ikpoba River. Data gathering relied on a Hydromet System installed along the river, collecting hourly measurements of gage height, ambient temperature, and atmospheric pressure. To convert the gage height to streamflow data, historical gage and streamflow data covering the period from 2015 to 2020 were extracted from the Ikpoba River rating curve and were analyzed using curve fitting techniques to establish the precise relationship between streamflow and gage height. Various goodness-of-fit measures, such as adjusted R-squared value, standard error of estimate, and coefficient of determination, were utilized to identify the best-fit relationship. The estimated streamflow data were subsequently validated using the Soil and Water Assessment Tool, incorporating the digital elevation model of the study area, along with other input parameters like soil, slope, daily maximum precipitation, and daily maximum temperature. Validation results were illustrated using regression plots generated in Microsoft Excel. From the machine learning results, random forest algorithm outperformed other methods in predicting streamflow, with a mean square error of 0.02 and a coefficient of determination of 0.98. Conversely, decision trees showed superior accuracy in predicting individual data points, with the lowest root-mean-square error of 0.02.

1. INTRODUCTION

As a result of its recurring patterns, the hydrological cycle offers opportunities for water flow optimization and estimation (Sohn et al., 2020). Streamflow is extensively utilized in the planning and managing water resources, including hydropower operations and planning, urban, agricultural, and environmental water supply operations, drought management, and flood mitigation. River flow calculations must be precise in order to support the efficient design and maintenance of vital water infrastructure, such as culverts and dams (Hyndman & Athanasopoulos, 2018). Predicting streamflow in an accurate, timely, and continuous manner during anomalous events acts as a preventative strategy against impending disaster and enables the taking of preventative action in order to save lives and valuable property. Additionally, it gives decision-makers and stakeholders the information they need to manage complex water resources systems effectively (Ghobadi & Kang, 2022). Still, predicting erratic river flows is made more difficult by the intricate relationship between water and environmental factors, which are regularly influenced by pollution and human activity. Streamflow is influenced by a wide range of factors, including temperature, evapotranspiration, and land use, topography, and soil properties. Streamflow and watershed features are related in a non-linear way (Adnan et al., 2019; Shah et al., 2021). Streamflow is frequently monitored at stream gauge stations, computed using hydrological models with a physical foundation, or statistically predicted using empirical models based on data. The spatiotemporal availability of stream gauge data for desired sites might be restricted, despite the fact that in situ measurements are crucial for getting precise streamflow records. Furthermore, due to human or equipment error, these stations may mis-measure streamflow necessitating reevaluation and flow prediction for the periods when data is unavailable. Hydrological models involve physical formulas to represent the intricate meteorological and hydrological processes; they are based on explicit correlations between inputs and outputs and demand a high degree of system knowledge (Feng et al., 2022; Cacal et al., 2023). For a very long time, flow prediction has made extensive use of numerous linear statistical prediction models, such as the AR and ARMA models, whose accuracy is frequently subpar. With the development of machine learning algorithms in recent decades, non-linear prediction models like support vector machines (SVM) and artificial neural networks (ANN) have been used in the hydrological field (Lin et al., 2021; Chui & Han, 2021; Wu & Wang, 2022), improving time series prediction accuracy to some degree. Over the past 20 years, machine learning models have become more and more popular for estimating the past or forecasting future streamflows. Regression-based machine learning models, also known as training models, identify statistical correlations between target and input data for a historical period and forecast future periods. A number of machine learning algorithms have been applied to predict future streamflow. Khullar and Singh (2021) and Sit et al., (2020) reviewed early and new machine learning applications in water resources and hydrology. Artificial neural networks (ANNs), support vector machines (SVMs), random forests (RF), extreme gradient boosting (XGBoost), and deep learning (DL) are examples of frequently used algorithms. Noori and Kalin (2016) used the ANN model in conjunction with the physically based hydrological model Soil and Water Assessment Tool (SWAT) to forecast daily streamflow for ungauged watersheds. RF was used by Petty and Dhingra (2018) to forecast streamflow for floods. Adnan et al., (2019, 2020) used an optimally pruned extreme learning machine (ELM) to predict daily and monthly streamflow. Using ANN, wavelet neural networks (WNNs), and an adaptive neuro-fuzzy inference system, Dalkiliç and Hashimi (2020) estimated daily streamflow and found that WNNs provide more precise estimates. A DL model was created by Ghobadi and Kang (2022) to forecast long-term streamflow on a monthly timescale. Xu et al., (2022) forecasted monthly streamflow, taking into account variables from general circulation models, using the machine learning (ML) method. After using two ML models and two hydrological models to simulate the rainfall-runoff process, Sayed et al. (2023) concluded that ML models are useful forecasting tools. Ikpoba River is an ungauged river and getting a reasonable volume of streamflow data for hydrological studies is a huge task. This study attempts to investigate the performance of selected ML algorithm for streamflow data prediction. By employing the Soil and Water Assessment Tool (SWAT) to validate streamflow predictions derived from both empirical rating curve analysis and machine learning techniques, we offer a robust framework for estimating streamflow even in data-scarce areas. For instance, in regions with limited monitoring infrastructure, such as remote or developing areas, our approach provides a valuable tool for assessing streamflow and informing resource management decisions. Moreover, by explicitly discussing the implications of our methodology for ungauged rivers in the paper, we contribute to advancing knowledge in hydrological modeling by addressing a critical gap in the existing literature and offering practical solutions for estimating streamflow in challenging environments.

2. SWAT APPLICATION IN HYDROLOGICAL MODELING

The relevance of utilizing the Soil and Water Assessment Tool (SWAT) for streamflow prediction extends beyond gaged stations to ungauged stations as well. While gaged stations provide direct streamflow measurements, ungauged stations lack this data, presenting a challenge for water resource management and decision-making in many regions. SWAT's ability to utilize readily available spatial data, such as topography, land use, and soil properties, allows for the estimation of streamflow in areas lacking direct measurements. Therefore, the methodology developed in this study, validated through SWAT, holds relevance for ungauged stations by providing a framework for estimating streamflow and informing water resource management practices in data-scarce regions.

In our study, the Soil and Water Assessment Tool (SWAT) played a crucial role as an evaluation tool for validating streamflow data obtained from both empirical rating curve analysis and machine learning predictions. We employed SWAT to assess the accuracy of our streamflow predictions in the Ikpoba River, a task that previous studies often overlooked or conducted with less comprehensive methodologies. By utilizing SWAT, we were able to enhance the credibility and reliability of our findings by incorporating various environmental and hydrological parameters, such as digital elevation models, soil characteristics, and precipitation data, into our validation process. This approach not only addressed the limitations of traditional rating curve analysis but also provided valuable insights into the performance of our modeling approach, thus contributing significantly to the advancement of knowledge in hydrological modeling and streamflow prediction methodologies.

3. MATERIALS AND METHODS

3.1. Study area

Ikpoba River is the study area. The river is located in southern Nigeria's Edo State, which is surrounded by rainforests. The river originates in the northern portion of the Ishan Plateau and flows south-west through a valley that is sharply carved and through sandy areas before entering Benin City and merging with the Ossiomo River. Edo State lies roughly between longitude 06° 04'E and 06° 43'E and latitude 05°44' N and 07°34' N. The wet and dry seasons are the two distinct seasons that define Edo State's tropical climate. The wet season runs from April to October with a break in August. The average amount of rainfall during this time ranges from 150mm in the far north of the state to 250mm in the south. November through April is considered the dry season, with a cold harmattan period in December and January. During the rainy season, the average temperature is roughly 25°C (77 °F), while during the dry season; it is roughly 28°C (82 °F). In the south, the climate is humid tropical, while in the north, it is sub-humid. Because of the dense population and reliance on the stream, the Ikpoba River is severely disturbed as it passes through Benin City. According to Victor and Dickson (1985), the stream passes through a dense rainforest in its upper reaches, where organic input is influenced by surface runoff and organic matter from the surrounding vegetation. Because riparian settlements are sparsely populated on the outskirts of the city, disturbance from human activity is minimal and restricted.

3.2. Data collection

For accurate data collection, a multi-parameter instrument (Hydromet System) for the acquisition of hourly gage height, ambient temperature, and atmospheric pressure was installed along the river's flow section. For this study, data gathered from September to December 2022 was used.

3.2.1. Estimation of streamflow from gage data

In order to convert the gage data into streamflow, historical gage and discharge data from (2015-2020) were extrapolated from the Ikpoba River rating curve. Thereafter, the data were subjected to curve fitting analysis to ascertain the precise mathematical relationship between the gage height (m) and the streamflow (m^3/s). Selected goodness-of-fit statistics, including coefficient of determination (R^2), coefficient of correlation (r), adjusted-R^2 value, and standard error of estimate (SEE), were used to pinpoint the precise mathematical relationship between the gage data and streamflow. The measured gage data were then converted to streamflow using the best-fit relationship.

3.2.2. Validation of streamflow data

To validate the estimated streamflow data, the Soil and Water Assessment Tool (SWAT) was utilized. SWAT employed various datasets, including the digital elevation model (DEM), slope map, soil map, and land use land cover (LULC) map, alongside daily maximum temperature and daily maximum rainfall data, to simulate river flow (Shijun et al., 2020, Anna et al., 2021). Subsequently, a regression plot comparing the estimated streamflow (observed) with a SWAT-simulated streamflow was generated using Microsoft Excel Spreadsheet.

The SWAT model is a physically based, computationally efficient model and capable of simulating a high level of spatial details by allowing the watershed to be divided into a number of sub-watersheds. Major model components include weather, hydrology, soil temperature, and plant and land management.

Following validation, various machine learning techniques, including RF, DT, SVR, and GB, were employed to forecast future streamflow. The slope map and land use/cover map for the study area were created according to the procedures described in Figures 2 and 3, with Figure 4 illustrating the flowchart for streamflow prediction.

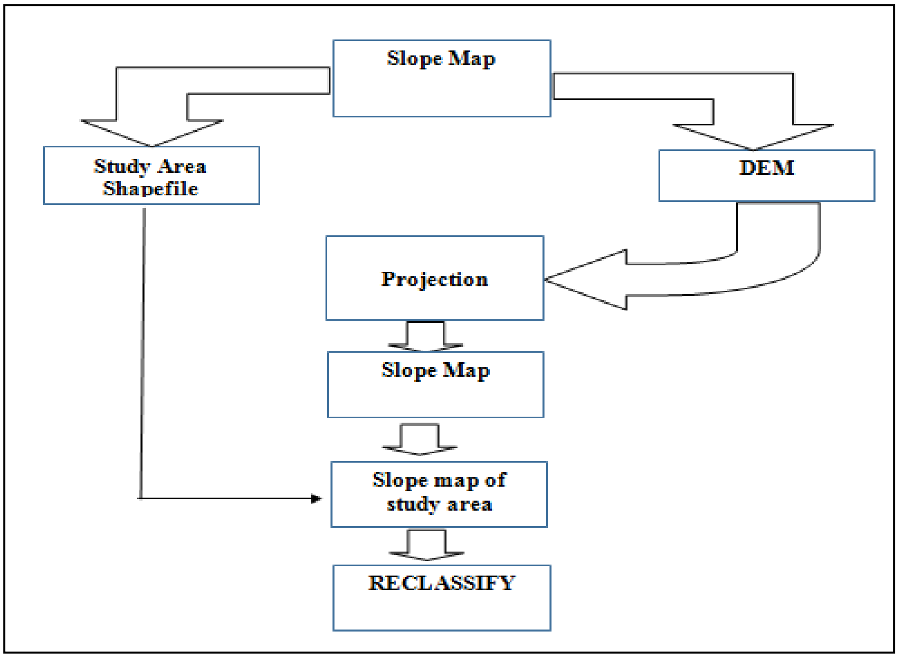

Figure 2. Schematic for generating slope map

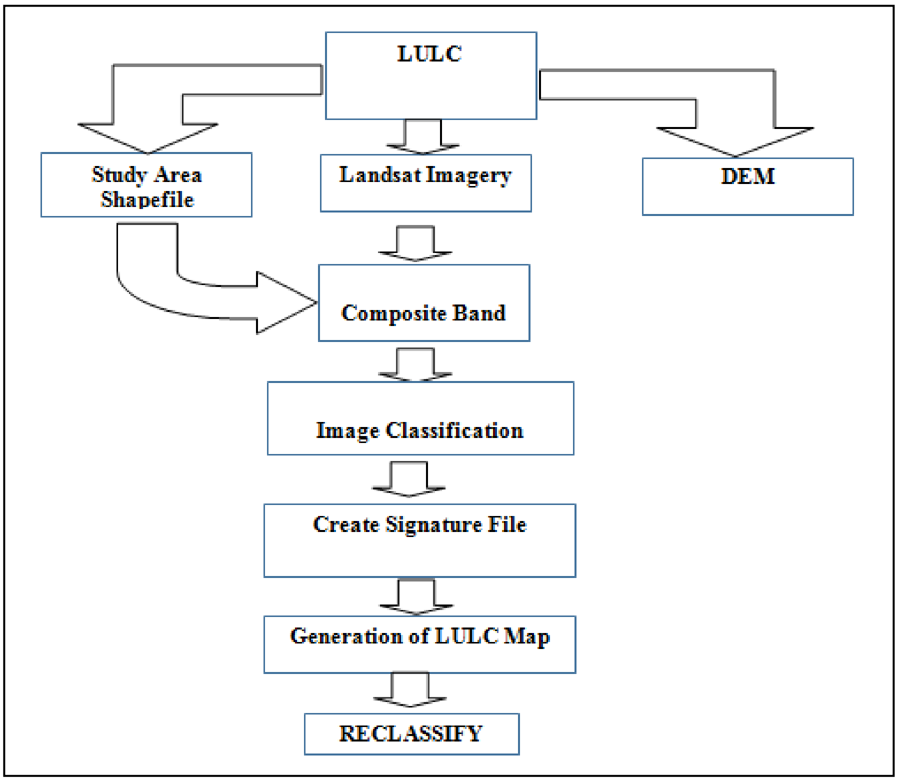

Figure 3. Schematic for generating lulc map

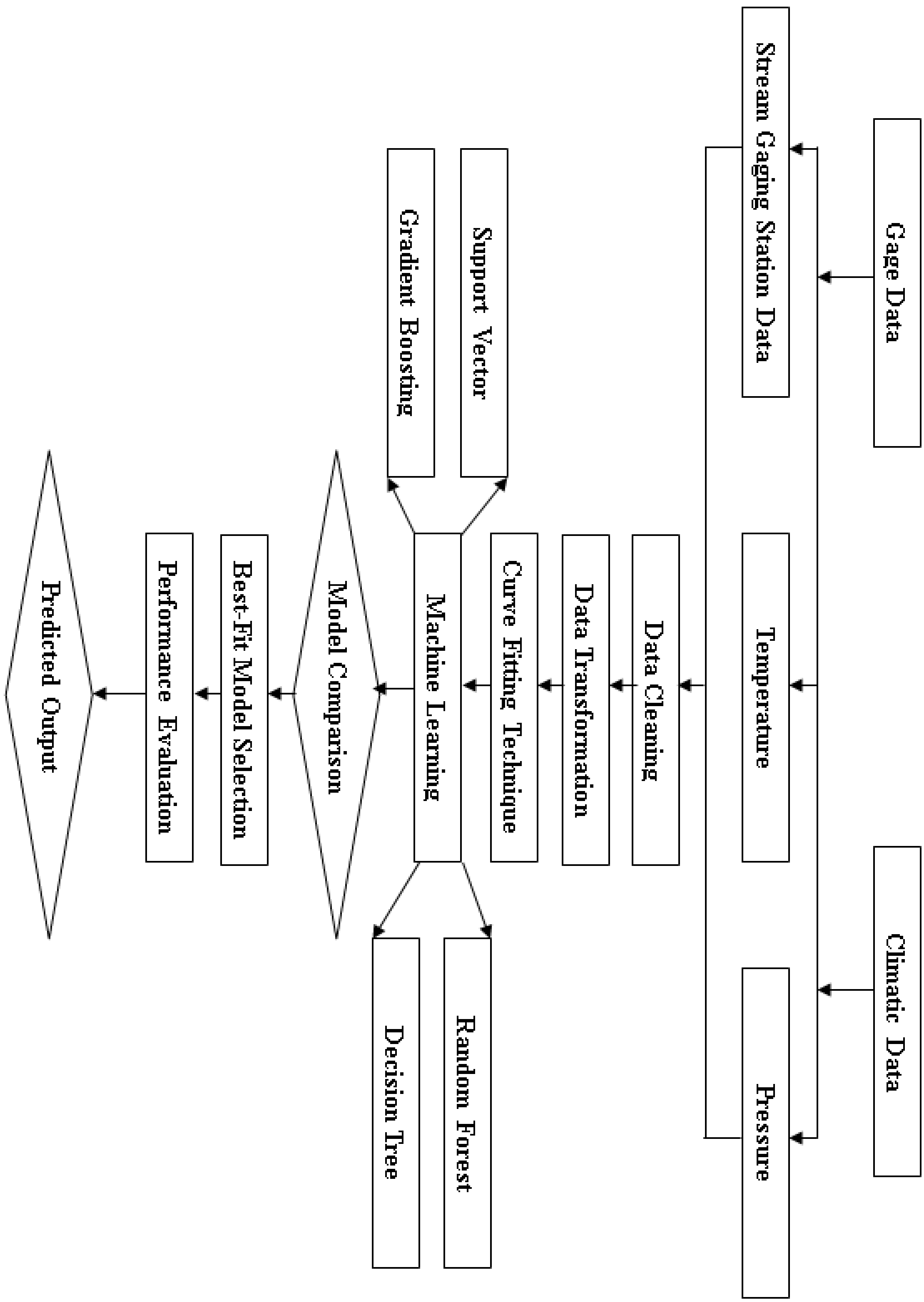

Figure 4: Streamflow Prediction Process Flowchart

3.3. Data preparation for machine learning

The initial stage in preparing the data for machine learning involved normalization. Normalization, a process of scaling numerical features to a standardized range, typically between 0 and 1, ensures that all features contribute equally to the model training process. For this study, the Min-Max scaling approach was utilized, and implemented as follows:

```
from sklearn.preprocessing import MinMaxScaler
>>scaler = MinMaxScaler ()
>>X_train_normalized = scaler.fit_transform(X_train)
```

Following normalization, the next step in data preparation was feature scaling. Feature scaling is imperative for algorithms sensitive to feature magnitudes, such as support vector machines (SVM) or neural networks. The chosen technique for feature scaling was standardization, which scales features to have a mean of 0 and a standard deviation of 1: implemented as follows;

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler ()
X_train_scaled = scaler.fit_transform(X_train)
```

By executing these preprocessing procedures, the data was appropriately formatted and encoded to ensure compatibility with various machine learning algorithms, thereby enhancing model performance and interpretability. Additionally, descriptive statistical analysis, employing metrics such as the arithmetic mean, mode, median, quartiles, and box diagrams, was utilized to unveil the underlying relationships between variables in the dataset (Qin & Huang, 2021). These analyses facilitated the comprehension of patterns and trends within the data, thereby aiding in the interpretation of the study's findings.

3.4. Data analysis using machine learning

Using four different machine learning algorithms was the first step. These algorithms made it easier to create models and predictive techniques based on historical data, which was helpful for projecting river flow rates in the future (Shrestha et al., 2021). The selection of the gage data, temperature, and pressure data for each hour was based on their proven relationship with river behavior. The hourly gage height captures the essence of water movement, whereas pressure and temperature are surrogates for meteorological factors that have a major impact on flow dynamics. Temperature is a key factor in determining flow patterns because it is a powerful predictor of water phase transitions and evaporation rates. In contrast, the inflow rates into the river system are directly impacted by pressure. Through the integration of these features, we leverage the complex interactions between meteorological factors and the dynamics of water resources, thus, improving our models (Fayaz & Goswami, 2019).

We used cross-validation techniques to navigate the complex world of model selection and hyperparameter calibration. This tactic involved dividing the dataset into subsets for training and validation in order to enable iterative validation across various segments. A variety of metrics were used to evaluate the performance of the model. The dispersion between predicted and actual values was explained by mean squared error, the fundamental indicator of prediction accuracy (Kumar et al., 2021). To further measure predictive efficacy, the coefficient of determination was employed as a benchmark (Liu & Hsieh, 2021).

4. RESULTS AND DISCUSSION

4.1. Streamflow generation

The outcome of the curve fitting analysis is presented in Table 2.

Table 2. Curve fitting analysis

Model	Correlation Coefficient (r)	Coefficient of Determination (R^2)	Adj. R^2	Standard Error of Estimate (SEE)	P-value
Linear Function	0.858	0.736	0.721	120.538	0.000
Quadratic Function	0.986	0.972	0.968	40.812	0.000
Cubic Polynomial	0.996	0.992	0.990	0.2777	0.000
Power Function	0.985	0.971	0.969	0.453	0.000

Table 2 analysis revealed the cubic polynomial model as the optimal relationship between streamflow and water gage. Consequently, this model was utilized for converting gage data into streamflow. The univariate statistics of the resulting data post-conversion are detailed in Table 3.

Table 3. Univariate analysis of the dataset

Index	Pressure (mm)	Temperature (deg.C)	Gage Height (m)	Streamflow (m^3/s)
Count	14521.00	14521.00	14521.00	14521.00
Mean	12.936	25.333	17.304	5481.070
Std	0.186	0.914	0.438	249.397
Min	12.669	21.780	2.706	14.658
25%	12.798	25.023	17.219	5379.175
50%	12.872	25.477	17.374	5547.794
75%	13.041	25.920	17.445	5626.447
Max	13.654	27.123	17.544	5736.980

The univariate analysis revealed the distribution characteristics of each column's data. This analysis included the mean and standard deviation, providing insight into the central tendency and variability. Additionally, it detailed the range of the data, spanning from the minimum to the maximum values. The analysis also highlighted key percentiles: 25th, 50th (median) and 75th percentiles, indicating the most common values within these quartiles. As observed in Table 3, each column distribution had a mean value approximately similar to the most common value in the 50% data distribution quartile. For example, the data distribution under the label "pressure" had a mean value of 12.936317, and the most common value in the 50% data distribution was 12.872000, which are approximately 12.9 for both. Similarly, column labeled "temperature" had approximately similar value of 25.00, "gage height" had 17.30, and "discharge" had 5500.00. This indicates the centralized positioning of the mean value on the data distribution curve. Also, the minimum value of 12.669000 on the label "pressure" in terms of range was close to 12.798000, the most common data in its 25% distribution quartile. This information represents a closely packed data distribution and a similar pattern can be observed in the data column labeled "temperature". However, the data column labeled "gage height" and "discharge" had minimum values farther away in terms of range from the most common data in their 25% distribution quartile. This indicates the presence of outliers in such data columns. The combined box-and-whisker plot of the dataset is presented in Figure 5.

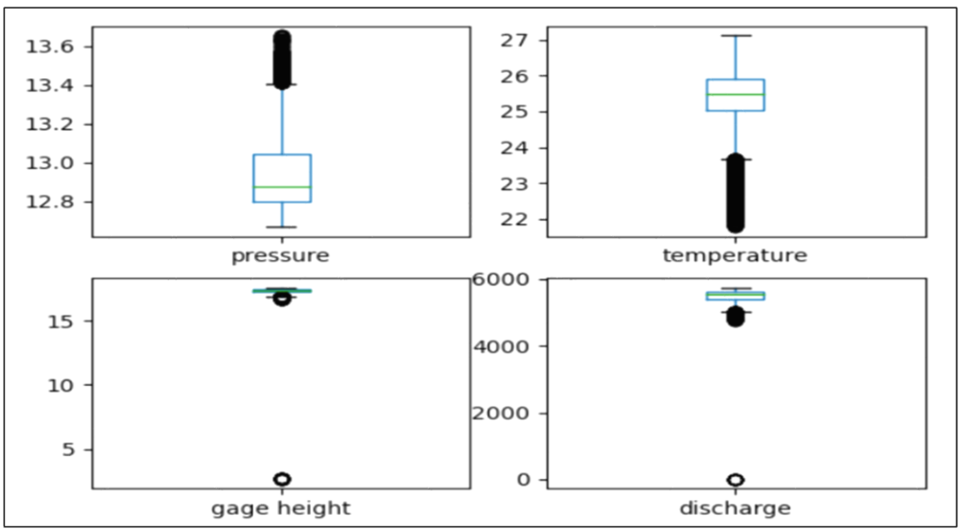

Figure 5. Box and Whisker plot of the dataset

The box-and-whisker plot confirmed the closely packed data distribution of columns labeled “pressure” and “temperature” at the upper and lower limits respectively. However, data columns labeled “gage height” and “discharge” have some outlying data farther away from the lower data limits. In terms of range, outliers present in the data column labeled “gage height” seem farther from its lower data limit compared to the other column data distributions. To identify these outliers, present in the data column labeled “gage height”, Z standard score method was employed. The Z standard score is the number of standard deviations by which the value of a raw score is above or below the mean value of the observed data distribution. Consequently, the lower and upper limit of 3; standard deviation away from the mean, to capture at least a 99.73% data distribution of the “gage height” was set, and a total of 11 outliers were identified. Unfortunately, there is no straight forward best solution for dealing with outliers because it depends on the severity of the outliers and the goals of the analysis. One way of handling outlier is to replace the outlier using imputation as if they were missing values using the interpolated mean values which were employed in this study.

4.2. Streamflow data validation

The result of the regression plot comparing the estimated streamflow (observed) with SWAT-simulated streamflow using Microsoft Excel Spreadsheet is presented in Figure 6

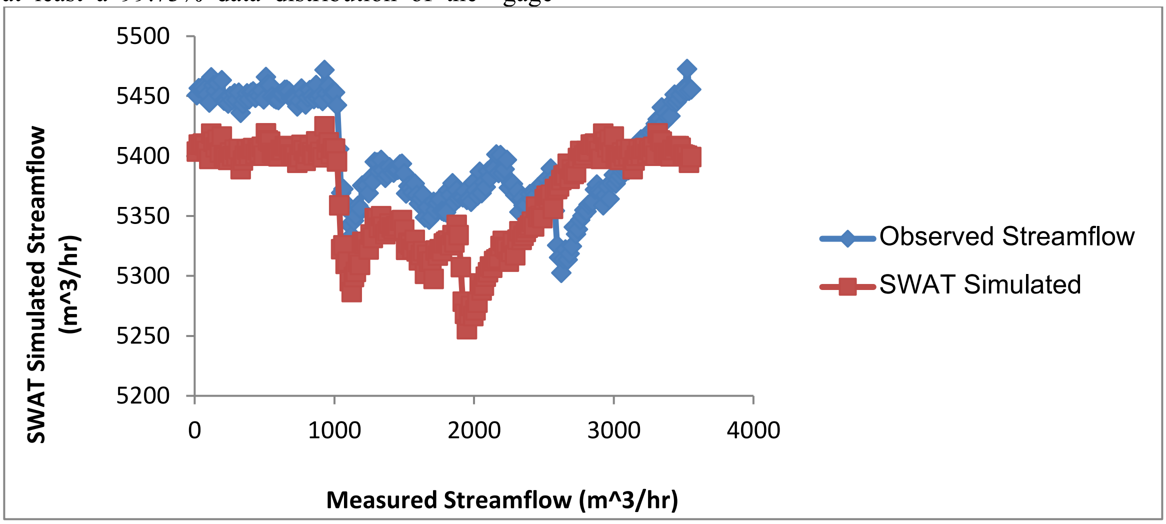

Figure 6. Observed versus SWAT simulated streamflow

The trend of the streamflow data for estimated versus SWATsimulated as observed in Figure 6 was employed to conclude that the estimated streamflow data were accurate and reliable. Hence, the estimated streamflow data were employed as training and validation datasets for future streamflow prediction using selected machine learning algorithms.

4.3. Streamflow prediction using machine learning technique

The first step in the development of the model using a machine learning approach is to split the datasets into dependent (y) and independent (x) variables. For streamflow prediction modelling, we are trying to predict the streamflow based on the water pressure, gage height and temperature. Hence, the streamflow data was used as the dependent variable while water pressure, gage height and temperature were used as the independent variables. Using the seabon library the pairplot which shows the exact relationship between streamflow and other independent variables was obtained.The relationship defined revealed a perfect linear relationship between streamflow and gage height. It was also observed that temperature range of between 24^0C and 27^0C defined the boundary for streamflow data collection and underscored the prevailing condition of the study area. To understand the variables that are positively or negatively correlated with streamflow, the correlation matrices between dependent and independent variables were obtained and presented in Table 4.

Table 4. Correlation matrix of regression

	Pressure	Temperature	Gage	Discharge
Pressure	1.000	-0.304	-0.982	-0.986
Temperature	-0.304	1.000	0.271	0.274
Gage	-0.982	0.271	1.000	0.992
Streamflow	-0.986	0.274	0.992	1.000

It was observed from the result of Table 4 that gage height is strongly positively correlated with streamflow with a correlation coefficient of 99.2%. Although temperature is also positively correlated with streamflow, it was observed that the correlation coefficient is very low (27.4%). For pressure, a strong negative correlation of about 99% was observed. Knowing the correlation between the dependent and independent variables, the datasets were then split into training, validation and testing data. RF, DT, SVR and GB like every other machine learning algorithm adopt the 60% training, 25% validation and 15% testing method. One unique quality of random forest is the tendency to determine the optimum number of n_estimator (optimum number of trees required to obtain the most desirable result. To make a prediction with a Random Forest model, each tree in the forest independently predicts the output variable, and the final prediction is based on the majority vote or average of these individual predictions. The graphical relationship between the performance accuracy of random forest and the number of n_estimators is presented in Figure 7.

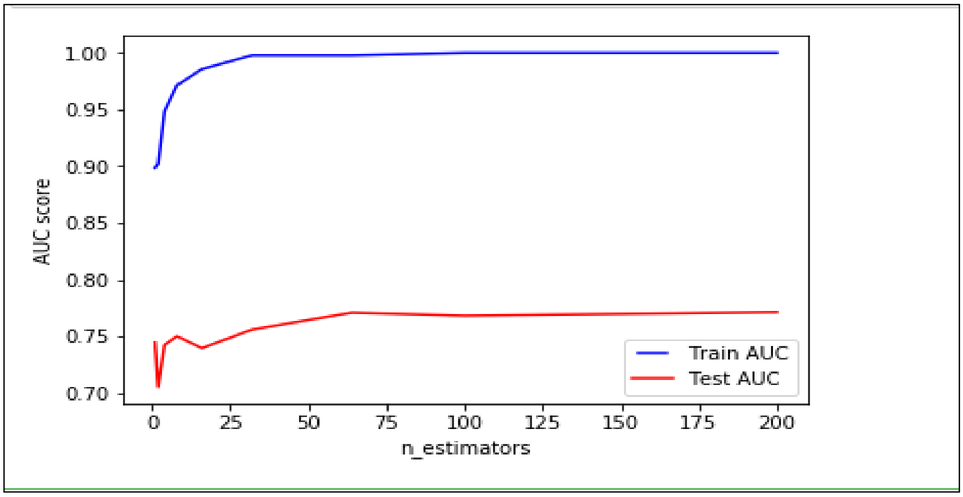

Figure 7. Model accuracy versus n_estimators

From the plot of Figure 7, the number of n_estimators for RFR was taken as 30. Using the sklearn library, the dataset was then fitted to RFR, DT, SVR and GB.

4.3.1. Model performance evaluation

To find out how well each model can forecast future streamflow, performance evaluation was carried out on both the training and testing datasets. Evaluation metrics and graphical visualization were used for performance assessment. Generating the learning curve that was used to train the model was the first step in assessing its performance. If the model is underfitting (not capturing all the patterns in the data), or overfitting (fitting too closely to the training data and not generalizing well to new data), the learning curve can shed light. By looking at the learning curve, you can determine whether your model performs better with more data, or with changes made to its architecture, or with adjusted hyperparameters. The learning curves developed for this study are shown in Figures 8, 9, 10 and 11 representing the curve for RF, DT, SVR and GB.

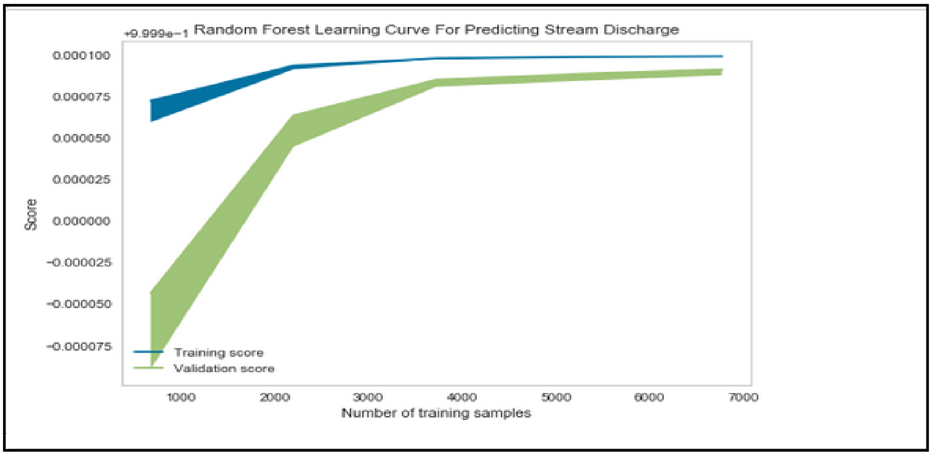

Figure 8. Random Forest Learning Curve

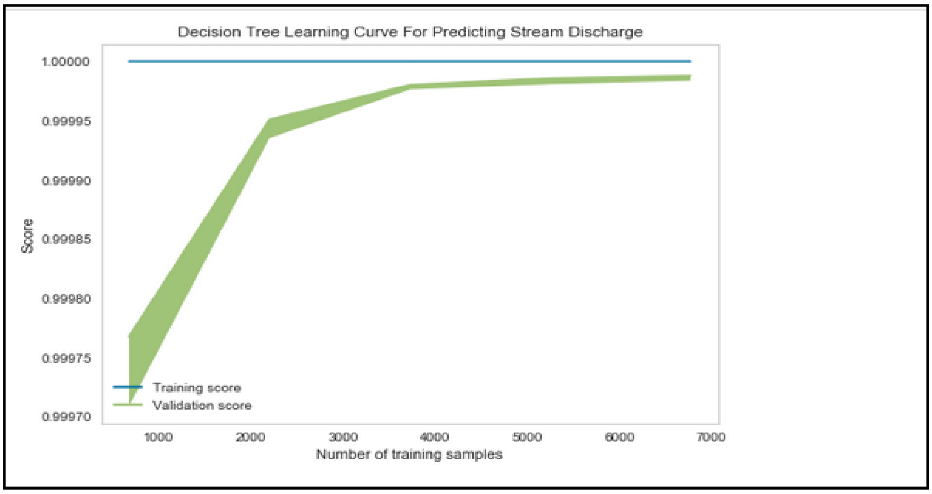

Figure 9. Decision Tree Learning Curve

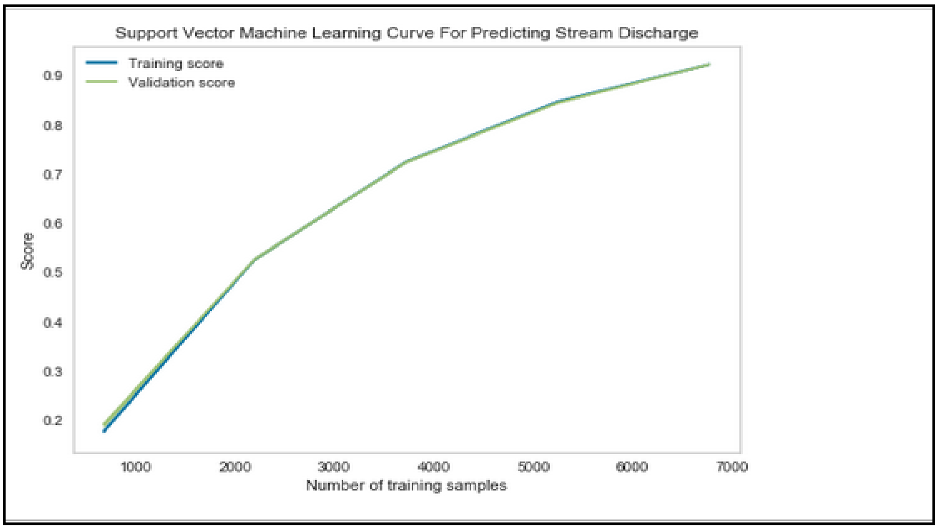

Figure 10. Support Vector Learning Curve

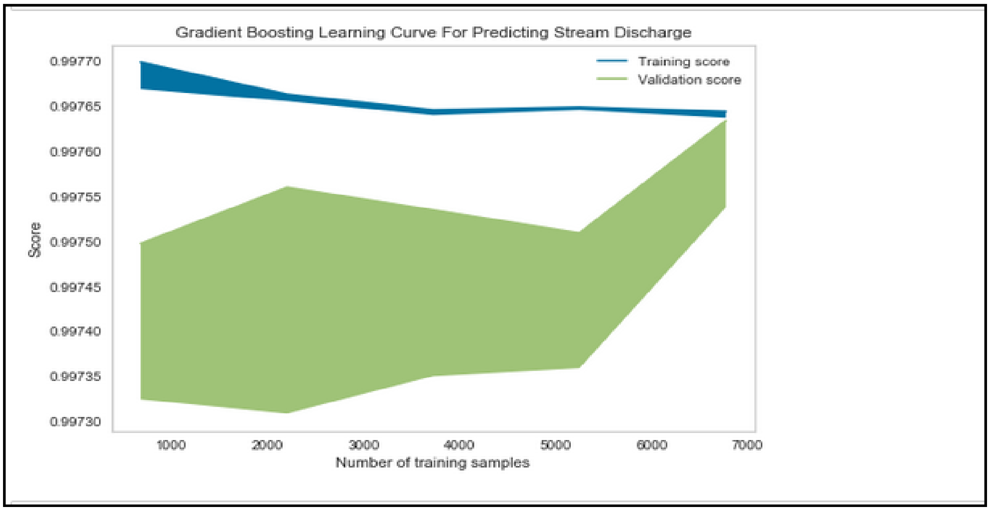

Figure 11. Gradient Boosting Learning Curve

It was noted from the presented learning curves that the chosen algorithm exhibits good data fit. Nonetheless, the SVR model's learning curve demonstrates that the validation and training curves overlapped. Curves that overlap could indicate that the model is effectively generalizing to new data and is not overfitting the training set. Though the curves may overlap, a model may still be overfitting the data; this is particularly likely to happen if the model has many parameters or if the data contains a lot of noise or variability. A machine learning model is generally trained with the aim of minimizing the difference between the training and validation curves and obtaining high accuracy on both sets. The model may be performing fairly well on the training and validation sets if the curves overlap, but it's crucial to keep an eye on the curves and tweak the model's architecture or hyperparameters as necessary to make sure the model is adapting to new data. In a study published in 2021, Guo et al. provided a clear explanation of the overlapping problem and how it affects the machine learning algorithm's overall performance. They confirmed that overlapping data can significantly affect how well deep neural networks generalize information from the training and validation sets. In particular, the authors discovered that a greater quantity of overlapping data can raise the model's generalization error. This finding implies that it's critical to choose the training and validation sets carefully to prevent the model from becoming overly fit. The study also emphasizes how helpful it is to keep an eye on the training and validation curves in order to evaluate the model's generalization performance, and it proposes that overlapping curves may be a helpful sign of the model's generalization error. Future research, according to the authors, should concentrate on creating techniques for choosing training and validation sets that reduce the quantity of overlapping data while still guaranteeing that the model can effectively generalize to new data. It was determined, based on the authors' submission, that SVR did not fit the data as well as RF, DT, and GB did. When RF, DT, and GB learning curves are compared, it is also observed that the Decision Tree Regression model's learning curve indicates that the model did not fit the data as well as RF and GB did. All things considered, RF was declared the best model and yields the best learning curve. Shrestha et al., (2021) are among the other authors who have assessed the effectiveness of the machine learning model by employing the learning curve approach. In their investigation, the effectiveness of clinical natural language processing (NLP) systems was assessed using a learning curve technique. Using a sizable clinical corpus, they trained and assessed their NLP system using supervised machine learning techniques. They also produced learning curves to determine the ideal sample size required to get the system to operate at the required level. They discovered that their learning curve approach worked well for maximizing the NLP system's performance on various clinical tasks and showed how useful it was for assessing and enhancing NLP system performances in clinical settings. A learning curve-based technique for evaluating the generalization effectiveness of machine learning models was presented by Liu and Hsieh in 2021. They created learning curves to calculate the models' generalization error and trained and assessed their models using a variety of benchmark datasets using supervised machine learning techniques. They showed the value of their method for evaluating the generalization capabilities of various machine learning models and discovered that their learning curve approach was successful in pointing out overfitting and underfitting issues in the models. A learning curve method was presented by Qin and Huang (2021) to assess deep neural network generalization performance. They created learning curves to calculate the models' generalization error and trained and assessed their models using a variety of deep neural network architectures on a range of benchmark datasets. They showed the value of their learning curve approach for optimizing the performance of deep neural networks in various applications and discovered that it was useful for comparing the performance of various neural network architectures and hyperparameters. The validation of model performance using the learning curve approach was judged to be satisfactory in accordance with the findings of Shrestha et al., (2021), Liu and Hsieh (2021), and Qin and Huang (2021).

A variety of metrics were used to evaluate the performance of the model in addition to the graphical method. The dispersion between predicted and actual values was explained by the mean squared error (MSE). To further measure predictive efficacy, coefficient of determination (R^2) and other selected metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Explained Variance Score (EVS), were employed as benchmarks. RMSE (Root Mean Square Error) is a measure of the differences between predicted and observed values, and it represents the square root of the average squared differences between the predicted and observed values. Lower RMSE values indicate better model performance. MAE (Mean Absolute Error) is a measure of the differences between predicted and observed values, and it represents the average absolute differences between the predicted and observed values. Lower MAE values indicate better model performance. R^2 (Coefficient of Determination) is a measure of how well the model fits the data, and it represents the proportion of the variance in the observed data that can be explained by the model. R^2 values range from 0 to 1, with higher values indicating better model performance. EVS (Explained Variance Score) is another measure of how well the model fits the data, and it represents the proportion of the variance in the observed data that is explained by the model. EVS values range from 0 to 1, with higher values indicating better model performance. Results of the estimated metrics are presented in Table 5.

Using the selected algorithms, the observed and predicted streamflow were obtained and presented in Figures 12, 13, 14 and 15 respectively.

Table 5. Training goodness of fit statistics based on the selected models

GoF Statistics	R^2	MAE	MSE	RMSE	EVS
RFR	0.98	0.01	0.02	0.05	0.97
DTR	0.96	0.02	0.03	0.02	0.94
SVR	0.68	18.71	906.23	30.1	0.65
GB	0.77	6.74	82.36	9.08	0.69

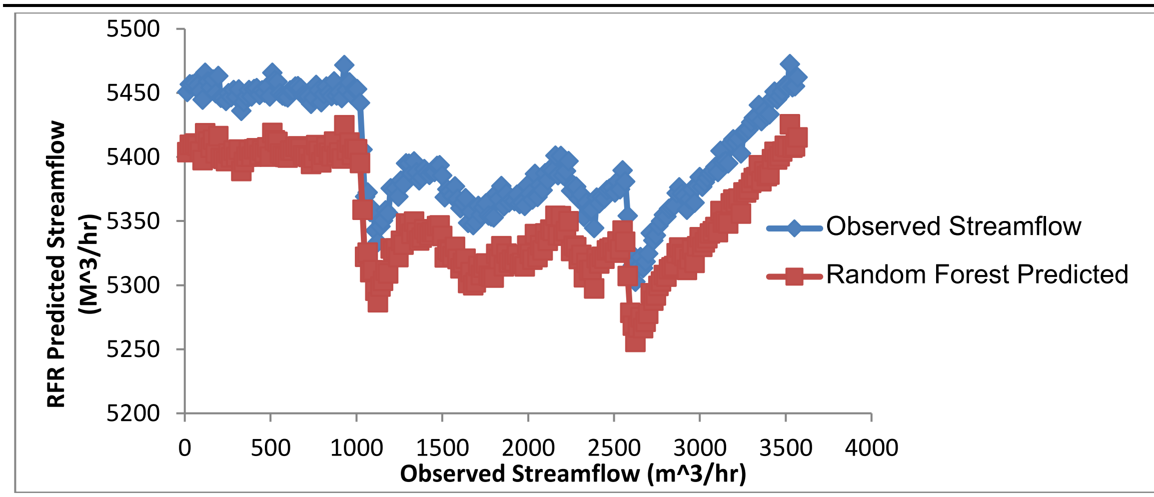

Figure 12. Observed versus RF predicted streamflow

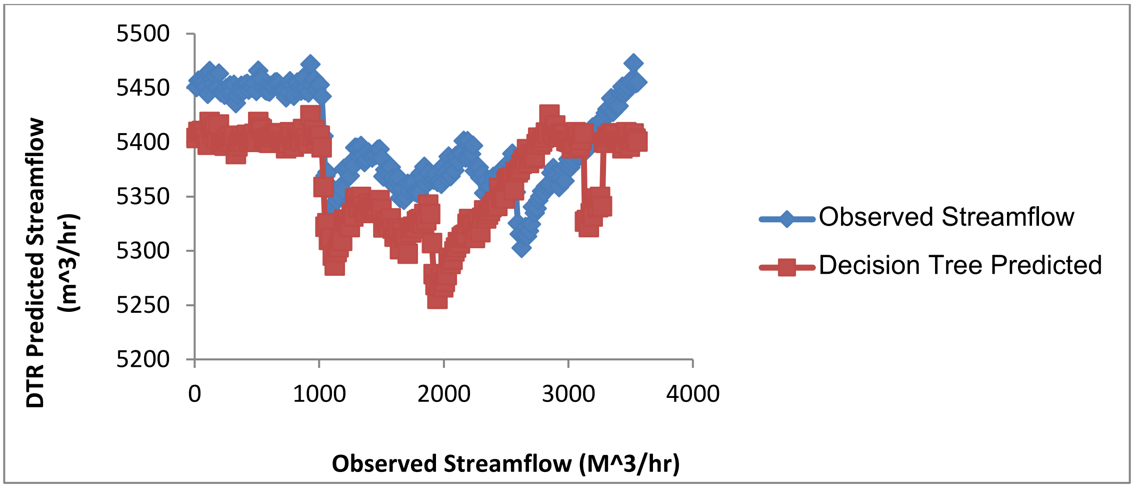

Figure 13. Observed versus DT predicted streamflow

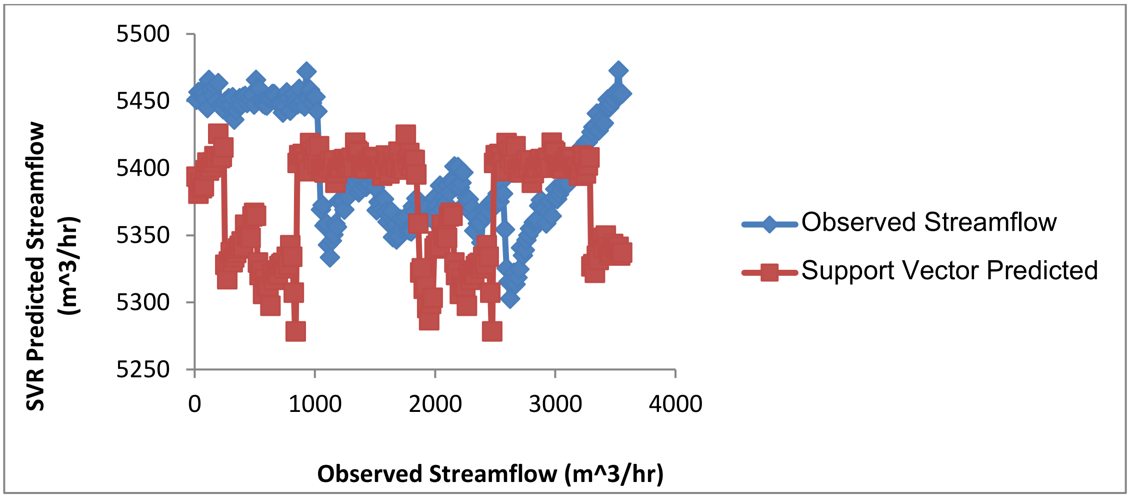

Figure 14. Observed versus SVR predicted streamflow

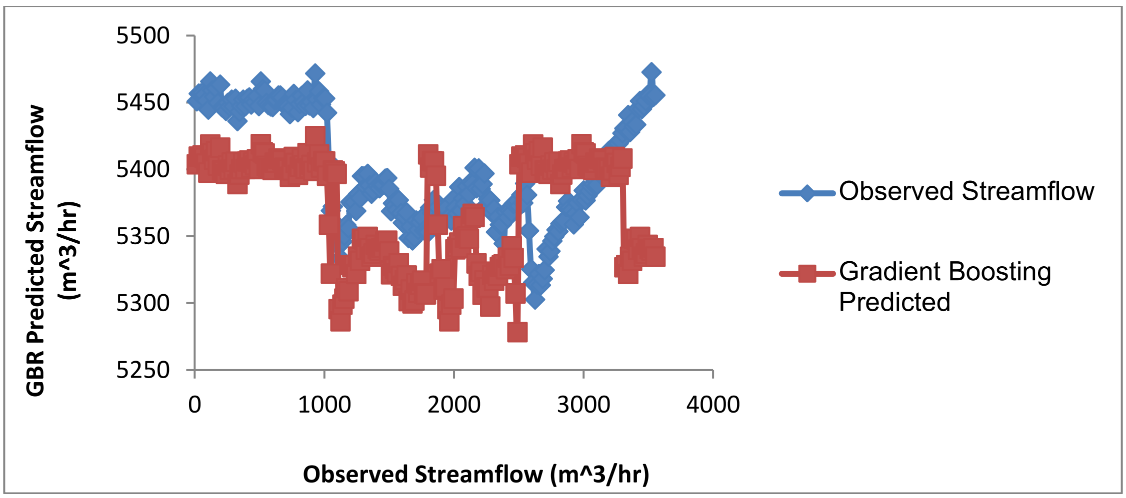

Figure 15. Observed versus GB predicted streamflow

Each model's overall goodness of fit was assessed using a ranking system that compared the five test criteria categories according to the relative importance of the statistical test results. A score of 4 was given to the model with the lowest MAE, lowest MSE, lowest RMSE, highest R^2 and highest EVS; a score of 3 was given to the next model, and a score of 1 to the worst. The model with the highest total point score was chosen as the best-fit model. With the highest total score of 19, RFR was chosen as the best-fit model, and was followed by DTR, GB, and SVR. In a similar study by Ates and Dadaser-Celik, 2020 on streamflow prediction using machine learning techniques. The authors used RF, DT, and SVR to predict daily streamflow in the Karasu River in Turkey using meteorological and hydrological variables as inputs. With an R^2 value of 0.93, RF was found to perform the best by the authors, followed by SVR (0.89), DT (0.85), and SVR (0.89). In a related study, Sahoo et al., (2020) compared the efficacy of RF, DT, and SVR for streamflow prediction in three Indian river basins using hydrological and meteorological data as inputs. The study focused on streamflow prediction using machine learning techniques. With an R^2 value of 0.91, RF proved to be the most effective, according to the authors, ahead of SVR (0.87) and DT (0.81).

5. CONCLUSIONS

This study compares, using various evaluation metrics, the streamflow prediction performance of four well-known machine learning algorithms: RF, DT, SVR, and GB. RF demonstrated superior predictive performance when compared to DT, SVR, and GB algorithms. RF had the lowest RMSE of 0.02 and the highest R^2 of 0.98. With an RMSE value of 0.02, DT has the lowest RMSE value, suggesting that it predicts individual data points more accurately. The study's conclusions regarding RF's superior performance in streamflow prediction are in line with those of earlier research, which discovered RF to be a useful algorithm for hydrological modeling. The results of the study also emphasize the significance of taking into account a variety of algorithms and assessment metrics when choosing a prediction model, since the effectiveness of an algorithm can differ based on the particular context and objectives of the modeling exercise.

REFERENCES

Adnan, R. M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B., & Kisi, O. (2019). Daily streamflow prediction using optimally pruned extreme learning machine, *Journal of Hydrology*, *577*, 123981.

Anna, C.H., Sareh, M., & Geoffrey, R. B. (2021). Climate change transformation: A definition and typology to guide decision making in urban environments, *Sustainable Cities and Society*, *70*, 1-6.

Ates, A., & Dadaser-Celik, F. (2021). Streamflow Prediction Using Machine Learning Approaches: A Case Study of the Karasu River in Turkey. *Water*, *13*(7), 932.

Cacal, J.C., Austria, V.C. A., & Taboada, E.B. (2023). Extreme event-based rainfall-runoff simulation utilizing GIS techniques in Irawan Watershed Palawan Philippines, *Civil Engineering Journal*, *9* (1), 220–232.

Chui, C.K., & Han, N. (2021). Wavelet thresholding for recovery of active sub-signals of a composite signal from its discrete samples, *Appl. Comput, Harmon, Anal.*, *52*, 1–24.

Dalkiliç, H.Y., & Hashimi, S.A. (2020). Prediction of daily streamflow using artificial neural networks (ANNs), wavelet neural networks (WNNs), and adaptive neuro-fuzzy inference system (ANFIS) models, water science and technology. *Water Supply*, *20*(4), 1396–1408.

Fayaz, H., & Goswami, M. (2019). Streamflow forecasting using machine learning techniques: A review. *Journal of Hydrology*, *575*, 631-642.

Feng, B.F., Xu, Y. S., Zhang, T., & Zhang, X. (2022). Hydrological time series prediction by extreme learning machine and sparrow search algorithm, *Water Supply*, *22*(3), 3143–3157.

Ghobadi, F., & Kang, D. (2022). Improving long-term streamflow prediction in a poorly gauged basin using geo-spatiotemporal mesoscale data and attention-based deep learning: A comparative study, *Journal of Hydrology*, *615*, 128608.

Guo, H., Wang, X., Liu, Q., Yang, C., & Du, X. (2021). Generalization performance of deep neural networks on overlapping datasets, *Neural Networks*, *145*, 295-306.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Khullar, S., & Singh, N. (2021). Machine learning techniques in river water quality modelling: a research travelogue, *Water Supply*, *21*(1), 1–13.

Kumar, A., Ramsankaran, R. A. A. J., Brocca, L., & Muñoz-Arriola, F. (2021). A simple machine learning approach to model real-time streamflow using satellite inputs: demonstration in a data scarce catchment, *Journal of Hydrology*, *595*, 126046.

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W., & Dai, Y. (2021). A hybrid deep learning algorithm and its application to streamflow prediction, *Journal of Hydrology*, *601*(6), 1–10.

Liu, Z., & Hsieh, C. (2021). A learning curve-based method for analyzing the generalization performance of machine learning models. *Journal of Computational Science*, *51*, 31-44.

Noori, N., & Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction, *Journal of Hydrology*, *533*, 141–151.

Petty, T. R., & Dhingra, P. (2018). Streamflow hydrology estimate using machine learning (SHEM), *Journal of the American Water Resources Association*, *54*(1), 55–68.

Qin, Y., & Huang, C. (2021). Evaluating the generalization performance of deep neural networks using learning curves, *Pattern Recognition*, *119*, 81-93.

Sayed, B.T., Al-Mohair, H.K., Alkhayyat, A., Ramírez-Coronel, A.A. & Elsahabi, M. (2023). Comparing machine-learning-based black box techniques and white box models to predict rainfall-runoff in a northern area of Iraq, the Little Khabur River. *Water Science and Technology 87*(3), 812–822.

Shah, L. A., Khan, A. U., Khan, F. A., Khan, Z., Rauf, A. U., Rahman, S. U., Iqbal, M. J., Ahmad, I., & Abbas, A. (2021). Statistical significance assessment of streamflow elasticity of major rivers, *Civil Engineering Journal*, *7*(5), 893–905.

Shijun, C., Qin, W., Yanmei, Z., Guangwen, M., Xiaoyan, H., & Liang, W. (2020). Medium- and long-term runoff forecasting based on a random forest regression model. *Water Supply*, *20*(8), 3658–3664.

Shrestha, R., Lu, J., Plank, B., & Mandel, M. (2021). Learning curve approach to performance evaluation of clinical natural language processing systems, *Journal of Biomedical Informatics*, *116,* 37-48.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12), 2635–2670.

Sohn, W., Kim, J. H., Li, M. H., Brown, R. D., & Jaber, F. H. (2020). How does an increasing impervious surface affect urban flooding in response to climate variability. *Ecol. Indic.*, *118*, 67-74.

Victor, R., & Dickson, D. T. (1985). Macro-benthic invertebrates of a perturbed stream in Southern Nigeria, *Environmental Pollution (Series A)*, *38*, 99-107.

Wu, J., & Wang, Z. (2022). A hybrid model for water quality prediction based on an artificial neural network, wavelets transform, and long and short-term memory. *Water*, *14*, 610.

Xu, W., Chen, J., Zhang, X.J., Xiong, L., & Chen, H. (2022). A framework of integrating heterogeneous data sources for monthly streamflow prediction using a state-of-the-art deep learning model. *Journal of Hydrology*, *614*, 85-99.