PREDICTIVE MODELING OF SUSPENDED SEDIMENT LOAD IN THE SURMA RIVER USING MACHINE LEARNING APPROACHES

Saifullah Mansur Akil*[1], Abrar Mahmood Chowdhury[2],Gulam Md. Munna[3],Ahmad Hasan Nury[4] ,Md. Misbah Uddin[5] and Tajmunnaher[6]

*[*1] Saifullah Mansur Akil, Shahjalal University of Science and Technology, Bangladesh,*
e-mail: shaifullahmansur12345@gmail.com
[2] Abrar Mahmood Chowdhury, Shahjalal University of Science and Technology, Bangladesh,
e-mail: gravitydestroyer333@gmail.com
[3] Gulam Md. Munna, Shahjalal University of Science and Technology, Bangladesh,
e-mail:gmunna192-cee@sust.edu
[4] Ahmad Hasan Nury, Shahjalal University of Science and Technology, Bangladesh,
e-mail: hasan-cee@sust.edu
[5] Md. Misbah Uddin, Shahjalal University of Science and Technology, Bangladesh,
e-mail: mun-cee@sust.edu
[6] Tajmunnaher, Shahjalal University of Science and Technology, Bangladesh,
e-mail: moon_cee@yahoo.com

***Corresponding Author**

ABSTRACT

Accurate prediction of suspended sediment load is crucial for sustainable river management and the safe design of hydraulic structures. Above all, it is vital for flood risk reduction in monsoon-affected regions like northeastern Bangladesh. This study presents a data-driven framework for estimating the total suspended sediment load in the Surma River using machine learning techniques. For this study, data was used from two stations: SW 267 located at Sylhet Sadar, and SW 273 located at Bhairab Bazar. The study considered six key variables: discharge, cross-sectional area, precipitation, temperature, water level, and maximum depth. Correlation analysis has been applied to treat multicollinearity and the best predictors have been selected for each station. Preprocessing is performed by transforming data into monthly values and filling up the missing data using appropriate statistical methods. A total of five machine learning models have been implemented for SW 267 and three models for SW 273. The models were trained, fine-tuned, and evaluated using standard performance metrics. Feature importance analysis was also done to find the main hydrological factors which control the sediment behavior at each site. The results indicated that out of the different machine learning models used in this study, the Multi-Layer Perceptron (MLP) model showed the best performance at SW 267, giving an R^2 test score of 0.86 and the lowest RMSE test score of 133.027. For station SW 273, Random Forest model gave the best with an R^2 test score of 0.74 and an RMSE test score of 32.965. Feature importance analysis shows the most influential variable for most of the models at station SW 267 is the discharge. In the case of station SW 273, the top features have varied by model and included max depth, cross-sectional area, and precipitation. The study provides a practical modeling approach for sediment prediction and helps to improve data-based river management.

Keywords: *Suspended sediment load, Machine learning, Neural networks, Feature importance, River management.*

1. INTRODUCTION

Sediment load is the total amount of solid material (such as soil, silt, sand, gravel, and organic matter) carried by a river or stream. It includes all particles carried by the flow of water, whether along the bed, in suspension, or dissolved. The suspended sediment load includes that part of the sediment load which is carried within the water column (and is not in contact with the riverbed) by turbulence and flow velocity (Li et al., 2018). The quantitative evaluation of sediment load is important for ecosystem analysis and management as well as hydraulic structure design, operation and maintenance, particularly for dams and channels (Nourani & Andalib, 2015). For instance, in the Bengal delta system, the Meghna-Surma rivers combined drain on the order of 1,060 million tons per year of fluvial sediment (Rashid et al., 2024). High sediment loads are of concern for river management. Extremely high siltation can raise the flood level, reduce the reservoir capacity, degrade water quality, and destroy aquatic habitat. On the other hand, reduced sediment supply due to, for instance, upstream dams, can starve the delta of nourishment and thus worsen coastal erosion and wetland loss. Thus, the accurate estimation and prediction of sediment transport is of prime importance in hydraulic engineering, such as designing dams, levees, and channels, in watershed and delta management, and in control of flooding and erosion.

Historically, sediment transport models fall into two classes. First, physical/empirical equations (e.g., those of Meyer-Peter–Müller or Einstein) derive sediment flux from hydraulic parameters: flow, shear stress, grain size. These formulas often require detailed calibration and assumptions about bed composition and can be quite inaccurate when applied outside of their original context. Second, the sediment rating curve is a purely empirical power-law relation of discharge to the suspended sediment load or concentration. The SRC is a simple way to assess the suspended sediment load, whatever the SS and discharge data are available (Horowitz, 2003; Warrick, 2015). But their simplicity comes at a cost. In practice, SRCs often underestimate sediment load at high flows and overestimate at low flows (Boukhrissa et al., 2013; Horowitz, 2003). Over recent decades, machine learning and artificial intelligence methods have shown great promise for the prediction of sediment load. These data-driven models (among them, artificial neural networks, support vector machines, ensemble trees, deep learning) are able to learn complex nonlinear relationships between inputs (flow, rainfall, watershed features) and sediment output with no need for explicit physical equations. Machine learning and deep learning methods are being popular for the prediction of suspended sediment load (AlDahoul et al., 2021; Samantaray et al., 2024).

This study aims to develop a predictive framework for estimating the total suspended sediment load of the Surma River at the SW 267 Sylhet Sadar station using twelve years of data (2013-2024) and at the SW 273 Bhairab Bazar station using eight years of data (2016-2023). Since the Surma River supports water supply and navigation and requires yearly dredging, understanding its sediment load is important for local decision-making. (Rony et al., 2022). This study offers novel contributions for the Surma River context. Since sediment load measurement is often costly and irregular in Bangladesh, particularly for rivers like the Surma, predictive modeling provides a practical and efficient alternative for reliable estimation.

2. STUDY AREA AND DATA DESCRIPTION

This study considers two hydrological monitoring stations along the Surma River: SW267(Sylhet Sadar) and SW273(Bhairab Bazar), both operated by the Bangladesh Water Development Board (BWDB). Station SW 267, situated in the upper reach of the river, is located at 24.88794° N, 91.86801° E, while station SW 273, positioned further downstream, lies at 24.04548° N, 90.99107° E. These stations were selected due to their consistent data availability and strategic positioning along the river.

The geographic distribution of the stations along the Surma River is illustrated in Figure 1, with detailed visual representations for SW267 and SW273. This figure has been prepared to serve as the primary reference for the spatial extent of the study area.

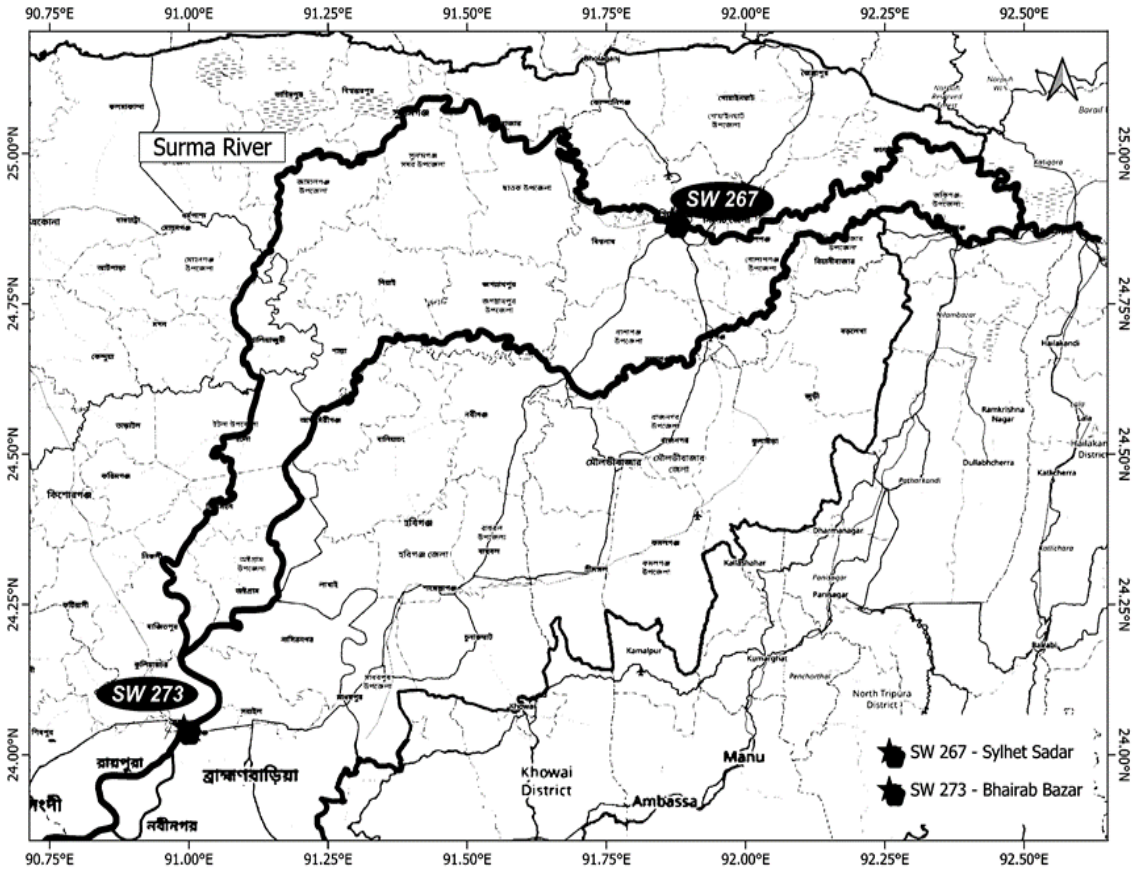

Figure 1: Study Area Map.

To understand the river's hydraulics and how sediment moves, we chose four important hydrological parameters: discharge, water level, maximum depth, and cross-sectional area. We got this information from the Bangladesh Water Development Board (BWDB) for two locations. For station SW267, we have data from 2013 to 2024, and for SW273, from 2016 to 2023. We also included two meteorological factors: rainfall (in mm) and temperature (in °C) - to see how they affect sediment transport. The Bangladesh Meteorological Department (BMD) provided this weather data for the same timeframes. Finally, the study aims to predict the suspended sediment load (measured in kg/s), which we also obtained from the BWDB for both stations.

3. METHODOLOGY

The overall methodology workflow is presented in Figure 2, which illustrates the sequential phases from data acquisition to model comparison.

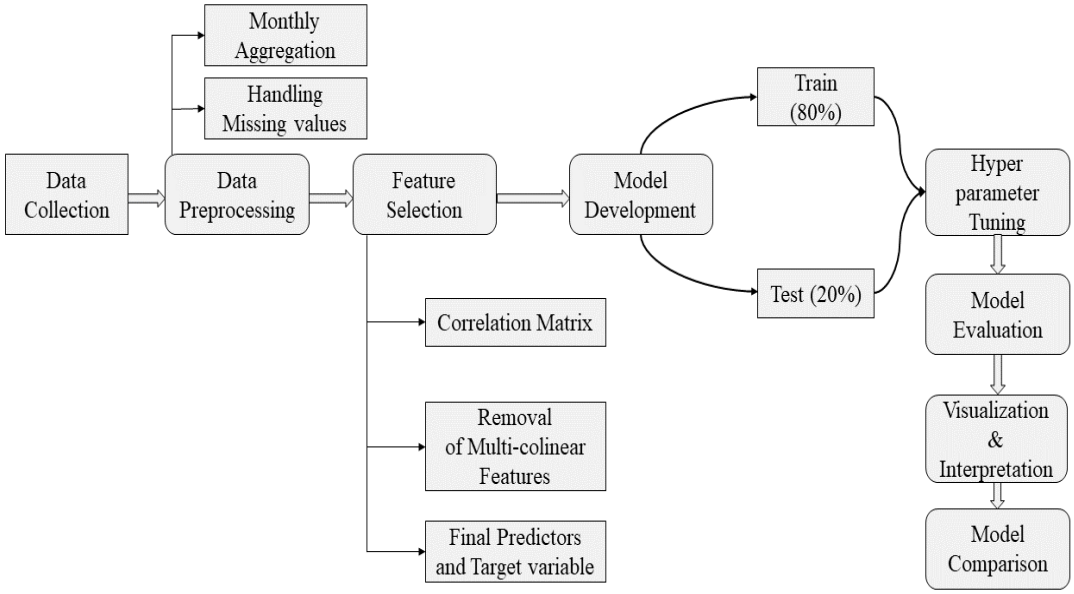

Figure 2: Methodology Framework.

3.1 Data Pre-processing

To maintain temporal consistency, all datasets were compiled into monthly values, using the average for each month. We dealt with missing data in the predictor variables by calculating the monthly average, which kept seasonal trends intact by filling in gaps with the long-term average for that specific month. For the target variable, suspended sediment load, we used the K-Nearest Neighbors (KNN) method to fill in missing values. To find the best 'k' value for KNN, we tested various values through cross-validation and chose the one that resulted in the lowest imputation error.

3.2 Predictor Selection and Correlation Analysis

Choosing the right predictors is essential for building reliable machine learning models. This step guarantees that only the most important and non-overlapping variables are used in the modeling process (Meyer et al., 2019). Our initial dataset included six predictors: Discharge (m^3/s), Cross-sectional area (m^2), Precipitation (mm), Temperature (°C), Maximum depth (m), and Water level (m). For better selection of predictors, a correlation matrix is generated for both stations for understanding the correlation among the variables. A correlation analysis is an important step in understanding the existence of multicollinearity among the predictors with high correlation values among two or more predictors (Shrestha, 2020).

At station SW 267, the analysis showed that cross-sectional area, maximum depth, and water level were strongly related. To find the best variable from this group, we used Pearson correlation. We chose cross-sectional area because it best matched the suspended sediment load, and we excluded maximum depth and water level. The final variables used to predict sediment load at this station were discharge, cross-sectional area, precipitation, and temperature. Figure 3 shows the correlation matrix for station SW 267.

At station SW 273, we didn't see any multicollinearity among the six variables. The correlations between all pairs of variables were within acceptable ranges, so we kept all six variables for the model. Figure 4 shows the correlation matrix for station SW 273.

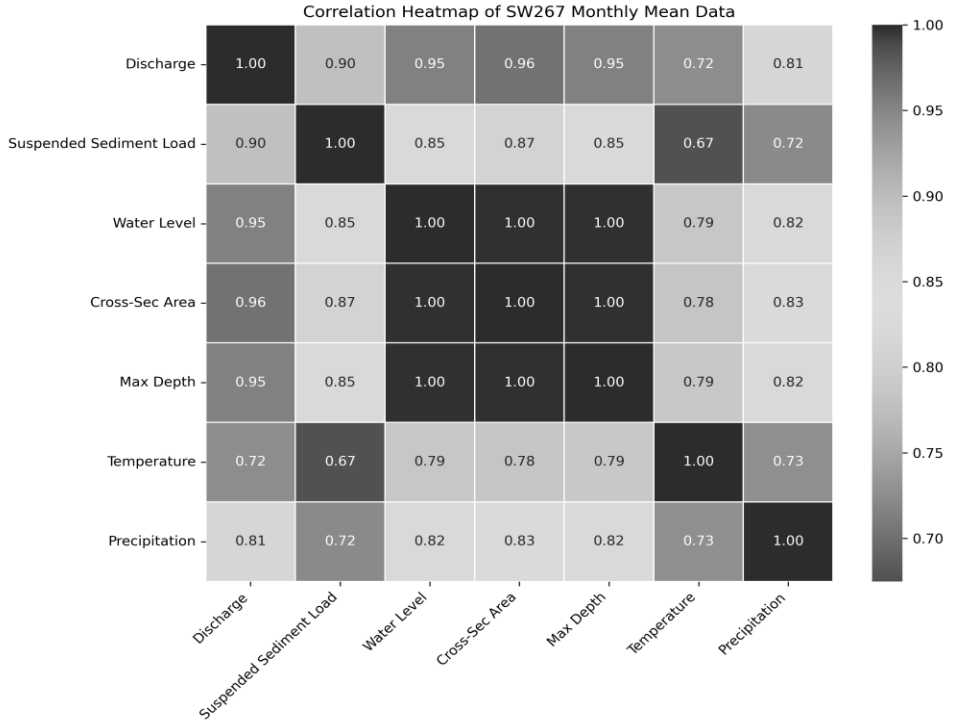

Figure 3: Correlation Matrix Heatmap of SW 267.

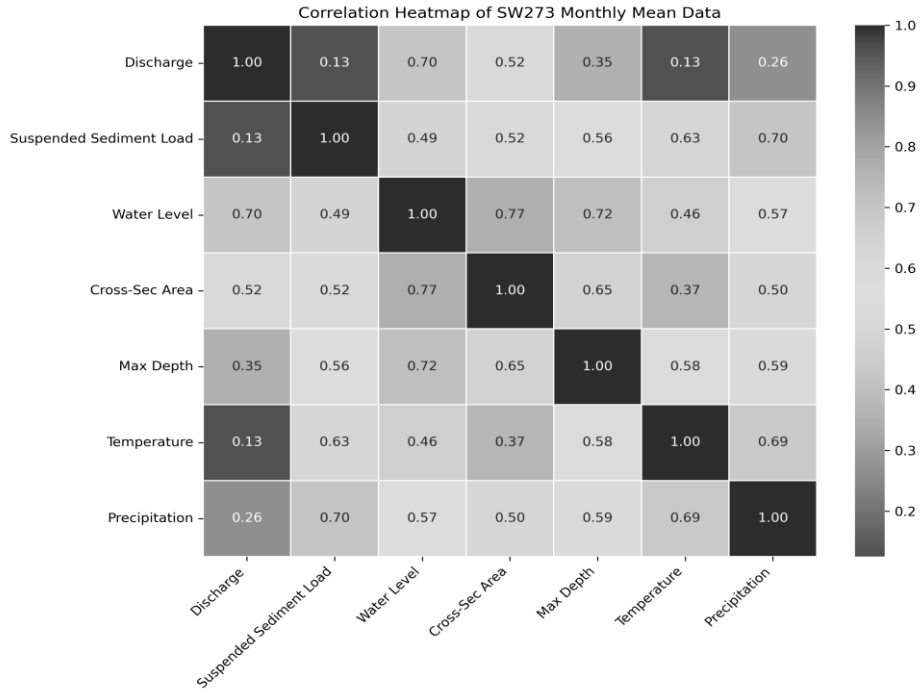

Figure 4: Correlation Matrix Heatmap of SW 273.

3.3 Model Training and Selection

The core of our approach is training and selecting the best model to accurately predict sediment load. This involves splitting the data into training and testing sets, using various machine learning algorithms, and evaluating their performance to find the most accurate model.

3.3.1 Data Partitioning

To preserve temporal dependencies present in hydrological data, the dataset was split chronologically:

1) Training set (80%) - Used for model fitting and parameter optimisation.
2) Testing set (20%) - Used exclusively for performance evaluation.

3.3.2 Model Candidates

For station SW 267, we tested five different machine learning models, covering linear, ensemble, boosting, and neural network methods to make a thorough comparison. These included Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), XGBoost (XGB), Multi-Layer Perceptron (MLP).

For station SW 273, we used a reduced set of three models because its data record was shorter and this station served mainly to assess the modeling framework's performance at a downstream location, rather than to allow detailed comparisons of all models. The selected models were Random Forest (RF), XGBoost (XGB), and Support Vector Regression (SVR).

3.4 Hyperparameter Tuning

Hyperparameter tuning is crucial for optimizing machine learning models and achieving the best predictive results (Weerts et al., 2020). In this study, we tuned hyperparameters for all models except Linear Regression (LR), as this model don't need extensive configuration. The tuning process involved a grid search combined with k-fold cross-validation on the training data to systematically test various hyperparameter combinations and pinpoint the optimal settings for each model. The list of tuned hyperparameters for all models applied at station SW 267 is presented in Table 1 to highlight the configuration used for performance optimization.

Table 1: Optimal Parameters for SW267

Model	Best Parameters
Random Forest	max_depth=7, max_features='sqrt', min_samples_leaf=3, min_samples_split=12, n_estimators=400
XGBoost	colsample_bytree=1.0, learning_rate=0.1, max_depth=1, n_estimators=250, reg_alpha=0.01, reg_lambda=5, subsample=1.0
SVR	C=50, gamma=0.01, epsilon=0.1, kernel='rbf'
MLP	activation='relu', alpha=0.0001, hidden_layer_sizes=(64,32), learning_rate='constant', solver='adam'

The corresponding set of optimized hyperparameters for station SW 273 is also provided in Table 2 to support reproducibility and comparative evaluation.

Table 2: Optimal Parameters for SW273

Model	Best Parameters
Random Forest	n_estimators=50, max_depth=3, min_samples_split=6, min_samples_leaf=1, max_features='sqrt'
XGBoost	n_estimators=50, max_depth=5, learning_rate=[0.05], subsample=0.6, colsample_bytree=0.6, reg_alpha=0, reg_lambda=10
SVR	C=5, gamma=0.1, epsilon=0.01, kernel='rbf'

3.5 Evaluation Metrics

Model evaluation is vital for assessing predictive accuracy and how well the models generalize (Emmert-Streib & Dehmer, 2019). In this study, three statistical indicators were used: the coefficient of determination (R2), the root mean square error (RMSE), and the mean absolute error (MAE).

3.5.1 Coefficient of Determination (R^2)

The R^2 value measures the proportion of variance in the observed data explained by the model (Filho et al., 2011).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

Where:
y_i is the observed sediment load,
$\hat{y_i}$ is the predicted sediment load,
$\bar{y}$ is the mean of observed sediment load values,
n is the number of observations.

An R^2 value closer to 1 indicates better model performance, with higher proportions of variance explained.

3.5.2 Root Mean Square Error (RMSE)

RMSE quantifies the square root of the average squared differences between observed and predicted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \quad (2)$$

Where:
y_i is the observed sediment load,
$\hat{y}_i$ is the predicted sediment load,
n is the number of observations.

A lower RMSE value indicates smaller overall errors and higher predictive accuracy.

3.5.3 Mean Absolute Error (MAE)

MAE represents the average absolute differences between observed and predicted values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

Where:
y_i is the observed sediment load,
$\hat{y}_i$ is the predicted sediment load,
n is the number of observations.

A lower MAE value indicates smaller average errors and therefore higher predictive accuracy.

3.6 Feature Importance Analysis

Feature importance analysis was conducted to determine how much each predictor contributed to the model's predictions. This helped us understand the key factors influencing sediment load changes. We used different methods to assess importance depending on the model type. For tree-based models like Random Forest and XGBoost, we used their built-in importance measures which use the structure of decision trees to calculate each feature's contribution (Zhou & Hooker, 2021). For Linear Regression, we used model coefficients to determine importance, standardizing the data when needed (Effrosynidis & Arampatzis, 2021). For Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP), we used permutation importance (Yang et al., 2009).

3.7 Model Comparison and Best Model Selection

To identify the best model for predicting the Surma River's total sediment load, we used a structured comparison. We evaluated all the trained models using an independent testing dataset, looking at statistical performance indicators like the coefficient of determination (R2), root mean square error (RMSE), and mean absolute error (MAE). We assessed each model's performance on both the training and testing datasets to see how well they generalized and to check for overfitting.

4. RESULTS AND DISCUSSION

4.1 Model Comparison Result

For station SW 267, we assessed the performance of five models to determine the best approach for predicting sediment load. The Multi-Layer Perceptron (MLP) achieved the highest performance with a test R^2 of 0.86, the lowest RMSE of 133.027, and an MAE of 90.614. Its small R^2 gap of 0.03 suggests excellent generalization without overfitting. Other models performed less well. Tree-based and ensemble models like Random Forest and XGBoost performed well, with R^2 values of 0.84 and 0.83, respectively, but they didn't outperform the MLP. The remaining models, including SVR and Linear Regression, showed slightly lower R^2 values and higher prediction errors. The model performance comparison for station SW 267 is shown in Table 3.

Table 3: Model Performance Comparison Table for SW267

Model	R^2 Test	R^2 Train	RMSE Test	RMSE Train	MAE Test	MAE Train	R^2 Gap
MLP	0.86	0.89	133.027	117.762	90.614	77.877	0.03
Random Forest	0.84	0.85	116.947	113.103	88.289	75.548	0.01
XGBoost	0.83	0.86	121.74	108.24	85.02	78.41	0.03
SVR	0.82	0.83	122.874	134.282	84.487	85.833	0.01
Linear Regression	0.80	0.81	132.351	136.944	90.359	87.395	0.01

For station SW 273, the Random Forest model performed best, with a test R^2 of 0.74 and a relatively low RMSE of 32.965. An R^2 Gap of 0.02 demonstrates good generalization. The SVR model performed similarly, with a test R^2 of 0.73 and a slightly higher RMSE of 33.610. While the XGBoost model also showed decent generalization, it yielded a lower test R^2 of 0.66 and a higher RMSE of 38.323. The model performance comparison for station SW 273 is shown in Table 4.

Table 4: Model Performance Comparison Table for SW273.

Model	R2 Test	R2 Train	RMSE Test	RMSE Train	MAE Test	MAE Train	R2 Gap
Random Forest	0.74	0.76	32.965	47.328	23.897	29.663	0.02
SVR	0.73	0.75	33.610	48.340	22.229	31.328	0.02
XGBoost	0.66	0.71	38.323	49.956	27.921	18.868	0.05

4.2 Model Plot of the best performing models

To further evaluate model performance beyond numerical metrics, visual analysis was carried out for the best models implemented at station SW267 and SW273. The predicted vs. actual plot also known as model plot for the MLP model (Figure 5) at station SW267 illustrates how closely the model's predictions align with the observed total sediment load values. Similarly the model plot for the best performing model at station SW273(Random Forest) is also shown in Figure 6.

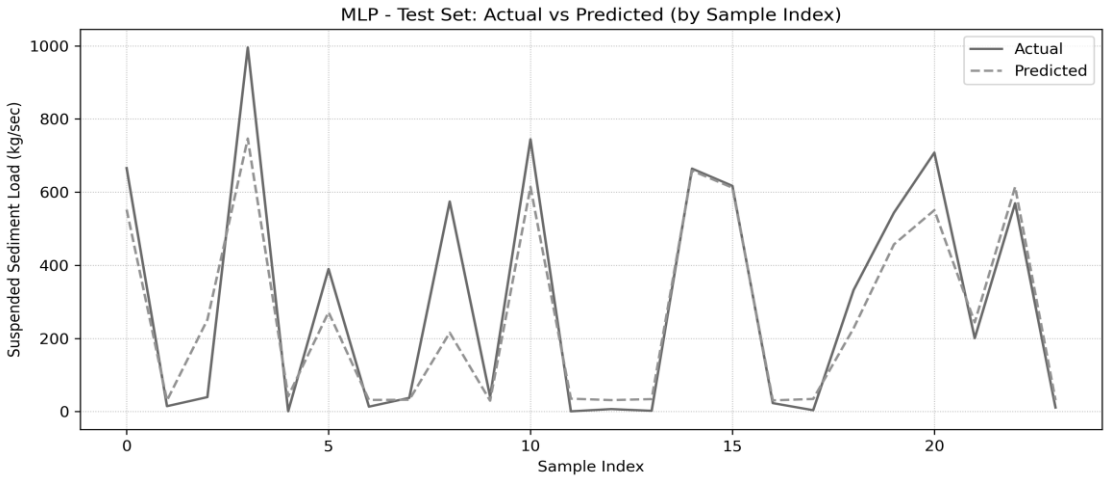

Figure 5: MLP Model Plot (SW 267).

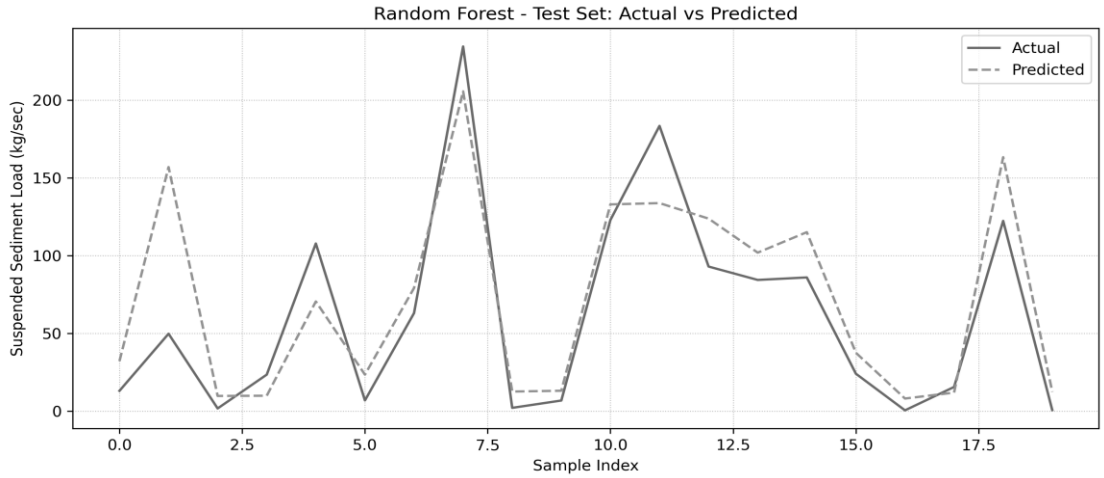

Figure 6: RF Model Plot (SW 273).

4.3 Feature Importance Analysis Result

At station SW 267, several evaluation techniques pinpointed discharge as the most impactful predictor across most models, such as MLP, Random Forest and XGBoost. SVR prioritized precipitation. Linear regression, on the other hand, considered temperature most significant. Table 5 below shows the top features and corresponding methods used by each model for station SW267.

Table 5: Feature Importance Analysis for SW267

Model	Top Feature	Method Used
MLP	Discharge	Permutation Importance (PI)
Random Forest	Discharge	Mean Decrease in Impurity (MDI)
XGBoost	Discharge	Built-in Gain Importance
Linear Regression	Temperature	Model Coefficients
SVR	Precipitation	Permutation Importance (PI)

For station SW 273, the feature importance analysis showed varied influential predictors. Random Forest indicated Max Depth as most important, using Mean Decrease in Impurity (MDI). XGBoost identified Cross Sec. Area as dominant, utilizing its built-in gain importance method. On the other hand, SVR considered Precipitation the most influential via Permutation Importance. This variation hints at a complex interaction between river characteristics, meteorological factors, and sediment dynamics at this location. Table 6 below presents a summary of the top features and the respective methods used by each model for station SW273.

Table 6: Feature Important Analysis for SW273

Model	Top Feature	Method Used
Random Forest	Max Depth	Mean Decrease in Impurity
XGBoost	Cross Sec. Area	Built-in Gain Importance
SVR	Precipitation	Permutation Importance

4.4 Comparative Analysis of Both Stations

A comparative summary of model performance and key predictor influence between station SW267 and station SW273 is presented in the following Table 7.

Table 7: Comparative Analysis of Both Stations.

Criteria	SW 267	SW 273
Data Period	2013-2024	2016-2023
Number of Models Applied	5 models	3 models
Best Performing Model	MLP	Random Forest
Best Test R^2 Score	0.86	0.74
Top Influential Feature	Discharge	Maximum Depth

5. CONCLUSIONS

This study aimed to create a predictive model using machine learning for total suspended sediment load in the Surma River. The results showed that these models could effectively represent the complex, non-linear connections between hydrometeorological factors and sediment load. At station SW 267, models using a neural network, like the Multi-Layer Perceptron (MLP), performed best, achieving the highest R^2 score of 0.86 and the lowest RMSE of 133.027. It demonstrates their strong ability to learn from the data. At station SW 273, ensemble models such as Random Forest excelled, with an R^2 score of 0.74 and an RMSE of 32.965. Feature analysis showed that discharge was most important at SW 267, while at SW 273, maximum depth was key for the Random Forest model and cross-sectional area for XGBoost. This result suggests that sediment behavior changes spatially. The results emphasize the flexibility and reliability of machine learning models for predicting river sediment, even when data is noisy or limited.

DECLARATION OF USE OF AI

Artificial Intelligence (AI) tools were used during the preparation of this manuscript to assist with refining the language, correct grammar, and enhance clarity. However, the scientific substance, including data analysis, result interpretation, and final conclusions, was solely the work of the authors.

REFERENCES

AlDahoul, N., Essam, Y., Kumar, P., Ahmed, A. N., Sherif, M., Sefelnasr, A., & Elshafie, A. (2021). Suspended sediment load prediction using long short-term memory neural network. *Scientific Reports*, *11*(1), 7826. https://doi.org/10.1038/s41598-021-87415-4

Boukhrissa, Z. A., Khanchoul, K., Le Bissonnais, Y., & Tourki, M. (2013). Prediction of sediment load by sediment rating curve and neural network (ANN) in El Kebir catchment, Algeria. *Journal of Earth System Science*, *122*(5), 1303–1312. https://doi.org/10.1007/s12040-013-0347-2

Effrosynidis, D., & Arampatzis, A. (2021). An evaluation of feature selection methods for environmental data. *Ecological Informatics*, *61*, 101224. https://doi.org/10.1016/j.ecoinf.2021.101224

Emmert-Streib, F., & Dehmer, M. (2019). Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error. *Machine Learning and Knowledge Extraction*, *1*(1), 521–551. https://doi.org/10.3390/make1010032

Filho, D. B. F., Júnior, J. A. S., & Rocha, E. C. (2011). What is R2 all about? *Leviathan (São Paulo)*, *3*, 60–68. https://doi.org/10.11606/issn.2237-4485.lev.2011.132282

Horowitz, A. J. (2003). An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. *Hydrological Processes*, *17*(17), 3387–3409. https://doi.org/10.1002/hyp.1299

Li, Y., Jia, J., Zhu, Q., Cheng, P., Gao, S., & Wang, Y. P. (2018). Differentiating the effects of advection and resuspension on suspended sediment concentrations in a turbid estuary. *Marine Geology*, *403*, 179–190. https://doi.org/10.1016/j.margeo.2018.06.001

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Nourani, V., & Andalib, G. (2015). Daily and monthly suspended sediment load predictions using wavelet based artificial intelligence approaches. *Journal of Mountain Science*, *12*(1), 85–100. https://doi.org/10.1007/s11629-014-3121-2

Rashid, M. B., Habib, M. A., Sultan-Ul-Islam, M., Khan, R., & Islam, A. R. M. T. (2024). Synthesis of drainage characteristics, water resources and sediment supply of the Bengal basin. *Quaternary Science Advances*, *16*, 100244. https://doi.org/10.1016/j.qsa.2024.100244

Rony, P. K., Uddin, Md. M., & Amin, G. Md. R. (2022). *Assessment of Dredging Impact on Hydrodynamics of Surma River using Hydrodynamic Model: HEC-RAS Approach.* Preprints. https://doi.org/10.22541/au.167249885.55039385/v1

Samantaray, S., Sahoo, A., Satapathy, D. P., Oudah, A. Y., & Yaseen, Z. M. (2024). Suspended sediment load prediction using sparrow search algorithm-based support vector machine model. *Scientific Reports*, *14*(1), 12889. https://doi.org/10.1038/s41598-024-63490-1

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*, 39–42. https://doi.org/10.12691/ajams-8-2-1

Warrick, J. A. (2015). Trend analyses with river sediment rating curves. *Hydrological Processes*, *29*(6), 936–949. https://doi.org/10.1002/hyp.10198

Weerts, H. J. P., Mueller, A. C., & Vanschoren, J. (2020). *Importance of Tuning Hyperparameters of Machine Learning Algorithms* (No. arXiv:2007.07588). arXiv. https://doi.org/10.48550/arXiv.2007.07588

Yang, J.-B., Shen, K.-Q., Ong, C.-J., & Li, X.-P. (2009). Feature Selection for MLP Neural Network: The Use of Random Permutation of Probabilistic Outputs. *IEEE Transactions on Neural Networks*, *20*(12), 1911–1922. https://doi.org/10.1109/TNN.2009.2032543

Zhou, Z., & Hooker, G. (2021). Unbiased Measurement of Feature Importance in Tree-Based Methods. *ACM Trans. Knowl. Discov. Data*, *15*(2), 26:1-26:21. https://doi.org/10.1145/3429445