Iranian Hydraulic Association
Journal of Hydraulics
Vol. 19, No. 4, PP. 31-45, Winter 2025

Research Article
https://doi.org/10.30482/jhyd.2024.453360.1701

A New Machine Learning Model for Predicting Suspended Sediment Load

Elham Ghanbari-Adivi[1*]

1- Department of Water Engineering, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran.

[*] **ghanbariadivi@sku.ac.ir**

Received: 20 April 2024, Accepted: 22 September 2024 J. Hydraul. Homepage: www.jhyd.iha.ir

Abstract

Predicting suspended sediment load is important for water resource management, water quality protection, erosion control, infrastructure planning, flood management, ecological conservation, pollution control, and environmental impact assessments. As a novel aspect of this study, the ANFIS-M5T model is introduced and used to predict suspended sediment loads. Our study combines the adaptive neuro-fuzzy inference system (ANFIS) and the M5T model to create a hybrid model for predicting suspended sediment load (SSL). The lagged rainfall, discharge, and SSL values were used to predict SSL. The results showed that the introduced ANFIS-M5T model performs better than other models so that it had the lowest mean absolute error (MAE: 525), the highest Nash Sutcliffe efficiency (0.98) and the lowest Percent bias (4).The ANFIS model had the second lowest MAE of 576, followed by MLP (586), RFN (682), and M5T (981). The ANFIS-M5T was a reliable tool for predicting SSL. By tackling the obstacles, assessing various methods, and showcasing the ANFIS-M5T model's efficiency, our research contributes to the continuous improvement and advancement of sediment measurement techniques and tools.

Key words: Hybrid models, Watershed management, Water resource management, Machine learning models.

1. Introduction

By carrying pollutants, nutrients, and heavy metals, suspended sediments can significantly impact water quality (Salih et al., 2020). A high sediment load can cause problems for dams, bridges, and water treatment facilities. During flood events, sediment loads often increase dramatically. The sediment load can also be a good indicator of soil erosion in the watershed. Suspended sediments can affect water clarity and quality, which is crucial for both aquatic life and human uses of water, such as drinking and recreation (Gupta et al., 2021). Predicting sediment loads can contribute to flood risk assessment and management strategies (AlDahoul et al., 2021).

Predicting suspended sediment load can be complex due to several factors. Numerous factors contribute to sediment transport, such as rainfall intensity, river flow rates, sediment characteristics (e.g., particle size, density), land use practices, topography, and vegetation cover (AlDahoul et al., 2021).

By examining the relationships between delayed rainfall, discharge, and suspended sediment values, researchers can create models to predict future sediment loads. Rainfall is a key factor in erosion and sediment transport. Considering delayed rainfall values allows researchers to evaluate the cumulative impact of rainfall events on sediment transport. Studies have shown that rainfall intensity and duration significantly influence erosion rates and sediment transport dynamics (e.g., de Almeida et al., 2021). Analyzing delayed discharge values helps capture the overall effect of past flow conditions on sediment movement, including transport and deposition. Suspended sediment concentration, which measures the amount of sediment particles in the water, is crucial for understanding sediment transport and erosion. Research indicates that historical discharge rates can affect sediment transport patterns, as higher flows often lead to increased sediment mobilization (e.g., Downs et al., 2021). The concentration of suspended sediment is critical for assessing water quality and ecosystem health, with various studies linking sediment loads to aquatic habitat degradation (e.g., Bainbridge et al., 2018). Models that incorporate both rainfall and discharge data have been developed to predict sediment loads more accurately, highlighting the importance of understanding the interplay between these factors (e.g., McMillan et al., 2010). Evaluating cumulative rainfall effects allows for a more comprehensive understanding of sediment transport over time, which is vital for effective watershed management (e.g., Bartley et al., 2014). These studies collectively emphasize the interconnectedness of delayed rainfall, discharge, and suspended sediment concentration in understanding and predicting sediment transport in aquatic systems.

River sediment dynamics involves complex feedback mechanisms and interactions. Sediment transport is a dynamic process affected by sedimentation and sediment resuspension. Due to the wide range and complexity of the effective parameters, the existing models and empirical relationships often do not have sufficient accuracy to predict the suspended sediment load. Advanced modeling techniques such as machine learning, data simulation, and numerical simulation are often used to address these complexities. (Ehteram et al., 2021; Ghanbari-Adivi et al., 2022). These methods can predict sediment load based on spatial and temporal variability, nonlinear relationships, and large datasets. A machine learning model can identify the most relevant features or input variables that contribute to sediment load prediction (Ehteram et al., 2021). By leveraging the advantages of machine learning models, researchers and practitioners can enhance their understanding of sediment transport processes and make more accurate predictions, supporting effective water resource management, erosion control, and ecosystem preservation (Nhu et al., 2020). Machine learning models can handle large and diverse datasets, making them suitable for predicting suspended sediment load at different spatial and temporal scales.

The M5Tree model is a machine learning model (Heddam and Kisi, 2018). It is an extension of the M5 model that uses decision trees for regression tasks. The M5T model is flexible and adaptable (Heddam and Kisi, 2018). It can handle different types of variables (categorical and continuous). By combining decision trees with linear regression, the M5T model can capture non-linear relationships between the input variable and the target variable. Researchers have widely used the M5T model for predicting sediment load.

Goyal et al. (2014) conducted a study to predict sediment yield within a watershed. The study specifically focused on the development of a decision tree-based M5 model tree and wavelet regression models for sediment yield modeling. A comparative evaluation revealed that the M5 model tree and wavelet regression models performed better than the ANN model. Rahgoshay et al. (2019) conducted a study on predicting suspended sediment load in watershed basins. They evaluated the support vector machine (SVM) and M5T models for predicting suspended sediment load. The study highlighted the superior performance of the optimized SVM model and recommended its use for hydrological simulations and sedimentary basin analysis. Salih et al. (2020) used M5P, attribute-selected classifier (AS M5P), and M5Rule to predict SSL data. The study examined various input scenarios based on river discharge and sediment load data. The performance of the data mining models was evaluated using statistical metrics and graphical analysis. Among the models tested, the M5P had the best performance. Achite et al. (2022) examined the performance of four machine learning methods: dynamic evolving neural-fuzzy inference systems (DENFIS), multivariate adaptive regression spline (MARS), and M5 model tree (M5Tree) in estimating suspended sediments. They aimed to determine the most effective combination of inputs for estimating suspended sediments. They constructed different input scenarios using streamflow (Q) and sediment (S) data. Study results showed that the DENFIS model performed better than the other models.

The M5 model tree (M5Tree) has limitations that can affect its ability to accurately predict suspended sediment load (SSL). However, the suspended sediment load is often influenced by complex nonlinear interactions between various factors, such as flow velocity, sediment characteristics, and channel morphology. M5Tree typically do not consider temporal dependencies explicitly, as they build independent decision trees for each prediction (Adnan et al., 2018). M5Tree are sensitive to outliers and can be affected by imbalanced data distributions (Adnan et al., 2018). M5Tree consist of a series of decision trees, where each leaf node represents a linear regression model. This structure may not be able to capture the complex nonlinear relationships.

In recent years, researchers have used hybrid methods to overcome the limitations of classical machine learning models (Ehteram et al., 2021). By integrating machine learning models into existing frameworks, researchers can improve accuracy, efficiency, and scalability of their analyses.

Adaptive neuro fuzzy interface system (ANFIS) is one of the robust machine learning algorithms (Ghenai et al., 2022). The ANFIS model is a hybrid approach that combines the adaptive learning capabilities of neural networks with the interpretability of fuzzy logic (Adnan et al., 2022). By utilizing fuzzy rules and membership functions, the ANFIS model can approximate complex nonlinear functions with good accuracy (Rahul et al., 2022). They can handle different types of data, including numeric and linguistic variables. ANFIS models can be trained efficiently, compared to more complex neural network architectures (Adnan et al., 2022). Thus, the ANFIS model is an ideal candidate to improve the performance of the M5T model.

By combining the advantages of ANFIS and M5T, the ANFIS-M5T model leverages fuzzy logic, adaptive learning, and nonlinear modeling to enhance the accuracy of predictions. By integrating ANFIS into the M5T model, the combined model benefits from adaptive learning, which helps represent and predict the underlying patterns and dynamics of the data.

The ANFIS-M5T model combines fuzzy logic and decision trees in order to improve the accuracy of suspended sediment load predictions. The ANFIS-M5T model helps to understand and manage water quality dynamics by accurately predicting suspended sediment load.

Accurately forecasting sediment load is crucial and significant for effective water resource management, as it aids in understanding and quantifying sediment transport in rivers, reservoirs, and other water bodies. The ANFIS-M5T model excels in predicting suspended load, which is essential for planning and optimizing reservoir operations. By leveraging these predictions, water resource managers can determine suitable sediment management practices, such as sediment flushing or periodic dredging, to ensure efficient and sustainable reservoir operations.

The ANFIS-M5T model offers reliable predictions that help evaluate the potential impacts of projects on water quality, aquatic habitats, and downstream ecosystems. It enables water resource managers to identify areas with high sediment loads and implement targeted erosion control measures. Additionally, the model contributes to environmental impact assessments by providing accurate predictions of suspended sediment load, supporting the evaluation of potential impacts on sensitive ecosystems and aiding in the development of appropriate mitigation strategies. Our study uses the ANFIS-M5T model to predict suspended sediment load in a reservoir basin.

The limitations of the methodologies present in such studies include restrictions in the accuracy of experimental models, as well as limitations in other machine learning methods like MT5, which are explained in the introduction of the paper. Regarding the limitations of the combined approach introduced in this research, one can point to constraints in optimizing the model structure, training time, and some complexities during implementation. However, considering that advanced computer systems are available today, these limitations can be addressed.

In summary, based on the above literature, considering the limitations mentioned about the M5T model, a new model that does not have these issues and limitations will be very helpful. In fact, the main goal of this study is to overcome the limitations of the M5T model and predict SSL data close to reality. The novelty of this study is the introduction of a new hybrid model for the mentioned purpose, which we called ANFIS-M5T.

2. Materials and Methods

Recent studies have increasingly concentrated on the use of machine learning (ML) models to forecast sediment loads, comparing these models against traditional empirical equations. For example, Bhattacharya et al. (2007) highlighted that artificial neural networks (ANN) and decision trees outperformed empirical formulations in accuracy. Similarly, Doğan et al. (2007) reported superior performance of their ANN model (3-2-1) in comparison to empirical equations. Furthermore, Yang et al. (2009) developed another ANN model (3-4-1) that utilized key factors and demonstrated that ANN is both a reliable and straightforward approach for predicting total sediment transport. Extending this, Ebtehaj and Bonakdari (2013) showed that their ANN model outperformed multiple empirical equations when predicting sediment transport in sewage systems.

Kitsikoudis et al. (2015) implemented three different ML methods and indicated that ML-based outcomes were generally better than commonly adopted empirical equations. Additionally, Sharafati et al. (2020) utilized gradient boosting regression (GBR), AdaBoost (AB), and random forest regression (RFR) to demonstrate that ML models excel in estimating daily suspended sediment loads.

Niazkar and Zakwan (2021) uncovered that models based on multi-gene genetic programming (MGGP) outshone traditional empirical models. Ghanbari-Adivi et al. (2022) discovered that integrating multiple radial basis function (RBF) neural network models in an ensemble approach significantly improved predictions of suspended sediment loads. Existing studies suggests that various machine learning techniques have considerable potential to enhance sediment load predictions compared to traditional empirical formulas, thereby underscoring the need for further exploration in this domain. In the following, the ML methods which used in this research were introduced.

2.1. M5Tree model

The M5Tree model, also known as the M5 model tree, is a hybrid modeling technique that combines decision trees and linear regression (Jia et al., 2023). As an extension of the M5 model, it improves the interpretability and accuracy of regression models. The M5Tree model consists of two main components: the tree structure and linear regression models at the leaf nodes. The M5Tree model consists of the following components:

Level 1: Root node: The root node represents the starting point of the decision tree. It contains the entire data set and serves as a basis for dividing the data into subsets based on different characteristics (Achite et al., 2022).

Level 2: Internal Nodes (Splitting Criteria): Internal nodes are decision nodes that split the data based on specific features or attributes. Each internal node represents a splitting criterion that divides the data into subsets based on the feature values. The decision at each internal node is made based on a specific condition, such as "if feature A is greater than X, go to the left child node, otherwise go to the right child node (Duarte et al., 2022).

Level 3: Leaf Nodes: A leaf node of the M5T Model is the most basic node of the model. The leaf node is typically the terminal node of the tree, and contains the predictions made by the decision trees (Duarte et al., 2022).

Level 4: Prediction and Model Selection: Once the M5Tree model is constructed, it performs predictions by traversing the decision tree. For a given data instance, it follows the splits and reaches the appropriate leaf node.

Level 5: Model Pruning and Simplification: The M5Tree model may undergo pruning and simplification processes to enhance its performance and interpretability. These techniques aim to remove unnecessary branches or reduce the complexity of the model.

2.2. ANFIS model

The ANFIS model operates based on five key components or layers, which can be explained as follows:

Layer 1: Input Layer: The first layer of the ANFIS model represents the input variables or features of the dataset. Each node in this layer corresponds to an input variable.

Layer 2: Rule Layer: Each node in this layer represents a specific rule, and it computes the firing strength or activation level of that rule based on the degree of membership of the input variables (Rahul et al., 2022).

Layer 3: Normalization Layer: The third layer, called the normalization layer, normalizes the firing strengths obtained from the rule layer. This layer combines the inputs based on the antecedent part of the fuzzy rules. The layer performs computations, such as weighted averages or other operations, to determine the output (Adnan et al., 2022).

Layer 4: Consequent Layer: The consequent layer computes the overall output of the ANFIS model based on the normalized firing strengths. The nodes of this layer combine the firing strengths and compute the weighted output contributions of each rule (Adnan et al., 2022).
Layer 5: Output Layer: The final layer of the ANFIS model is the output layer.

The ANFIS model uses a hybrid learning algorithm that combines gradient descent optimization and least squares estimation. During the learning process, the model adjusts the model parameters to minimize the difference between the predicted outputs and the actual outputs. This learning algorithm allows the ANFIS model to adapt and optimize its parameters based on the available input-output data.

2.3. Hybrid ANFIS-M5T model

The hybrid ANFIS-M5T is created based on the following levels:

- Train the M5Tree Model: Build and train the M5Tree model using the available dataset, considering the input variables and the target variable (e.g., suspended sediment load). The M5Tree model will provide initial predictions.
- Calculate Residual Values: Calculate residual values by subtracting the predictions from the actual values. These residuals represent the differences between the observed and predicted values.
- Train the ANFIS Model: Construct and train the ANFIS model using the residual values obtained from the M5Tree model as inputs.
- Combine Predictions: Use the trained ANFIS model to make predictions based on the original input variables and the residual values. The final output is computed based on the following levels (Eqs. 1-2):

$$\left(ANFIS_{res}\right) = f\left(re\left(t-1\right),...,re(t-n)\right) \tag{1}$$

$$final(output):\left(ANFIS_{res}\right) + M5T_{output} \tag{2}$$

Where , $ANFIS_{res}$: The outputs of the ANFIS model (residual values), and MLR_{output} : The outputs of the M5T model, $re\left(t-1\right)$: residual values at (t-1), n: lag times.

- Evaluation and Refinement: Evaluate the performance of the hybrid NAFIS-M5T model using appropriate evaluation metrics (e.g., mean squared error, R-squared)

The M5Tree model captures the primary patterns and relationships, while the ANFIS model learns to capture the remaining residuals, enhancing the overall prediction ability of the hybrid model. In this study, the new model is benchmarked against the ANFIS, M5T, multilayer perceptron, and radial basis function neural network models.

2.4. Benchmark models

The Multilayer Perceptron (MLP) model is a type of artificial neural network (ANN) that consists of multiple layers of interconnected artificial neurons. The input layer receives the input features of the data. The MLP model can have one or more hidden layers between the input and output layers. Each hidden layer contains multiple neurons (Fath et al., 2020). These hidden layers allow the MLP model to learn complex representations and capture non-linear relationships. The MLP model uses forward propagation to process the input data and generate predictions. Each neuron receives inputs from the previous layer and computes a weighted sum of those inputs. The weights represent the strength or importance of each input (Mohammadi et al., 2020). The neuron applies an activation function to the weighted sum. The output layer of the MLP model produces the final predictions or outputs. The MLP model is trained using an optimization algorithm, such as gradient descent, to minimize the difference between the predicted outputs and the actual target values (Mohammadi et al., 2021).
The Radial Basis Function (RBF) neural network is a type of artificial neural network that utilizes radial basis functions as activation functions (Fath et al., 2020). It is commonly used for various machine learning tasks, including pattern recognition, function approximation, and time series prediction. The input layer of the RBF neural network receives the input data. The RBF neural network typically has a single hidden layer. Each RBF neuron has a center point, which represents a prototype vector or a reference point in the input space. The width of each radial basis function determines the influence or the spread of the activation function. The output layer of the RBF neural network consists of one or more neurons that produce the final predictions or classifications.

3. Case Study

The Talar River watershed is located in the Mazandaran region of Iran. It covers an area of 2100 km^2. The watershed is characterized by the presence of the Talar River, which flows through the region and plays an important role in local hydrology and water resources. The watershed plays an important role in water supply, agriculture and ecosystem maintenance in Mazandaran. It is important to study and manage the Talar River watershed to ensure sustainable use of water resources and to maintain the ecological balance of the region. The climate of the basin is influenced by its geographical location and topography. The region experiences a Mediterranean climate, characterized by mild, wet winters and warm, dry summers. The annual rainfall of the basin is 552.7 mm.
In this study, lagged rainfall values, discharge values and sediment loads were used as inputs to the models. A prediction of suspended sediment load helps identify erosion-prone areas in the Talar Basin. High sediment loads can impact water quality, aquatic habitats, and the health of flora and fauna in the Talar River. By predicting suspended sediment load, appropriate mitigation measures can be implemented to protect and restore the ecosystem. Figure 1 shows location of case study. Figure 2 shows time series data. Daily data sets were gathered from 2000 to 2004. The models were used to predict one-day ahead SSL. Sediment samplers, such the US P-61 sampler, are designed to collect suspended sediment samples from specific depths of the water column. They can be lowered from a bridge or boat, and allow water and suspended sediment to enter the sampler. Once the sampler is in position, it can be opened to allow water and suspended sediment to enter the sampler. As water flows through the sampler, suspended sediments are trapped. After collecting the sample, it is taken to a laboratory for analysis. The sediment is separated from the water by filtering or centrifuging. It is then dried and weighed to determine the sediment concentration. Table 1 shows details of data. Using lagged values allows the models to incorporate the delayed effects of rainfall and discharge on sediment transport, as sediment movement often does not occur instantaneously.

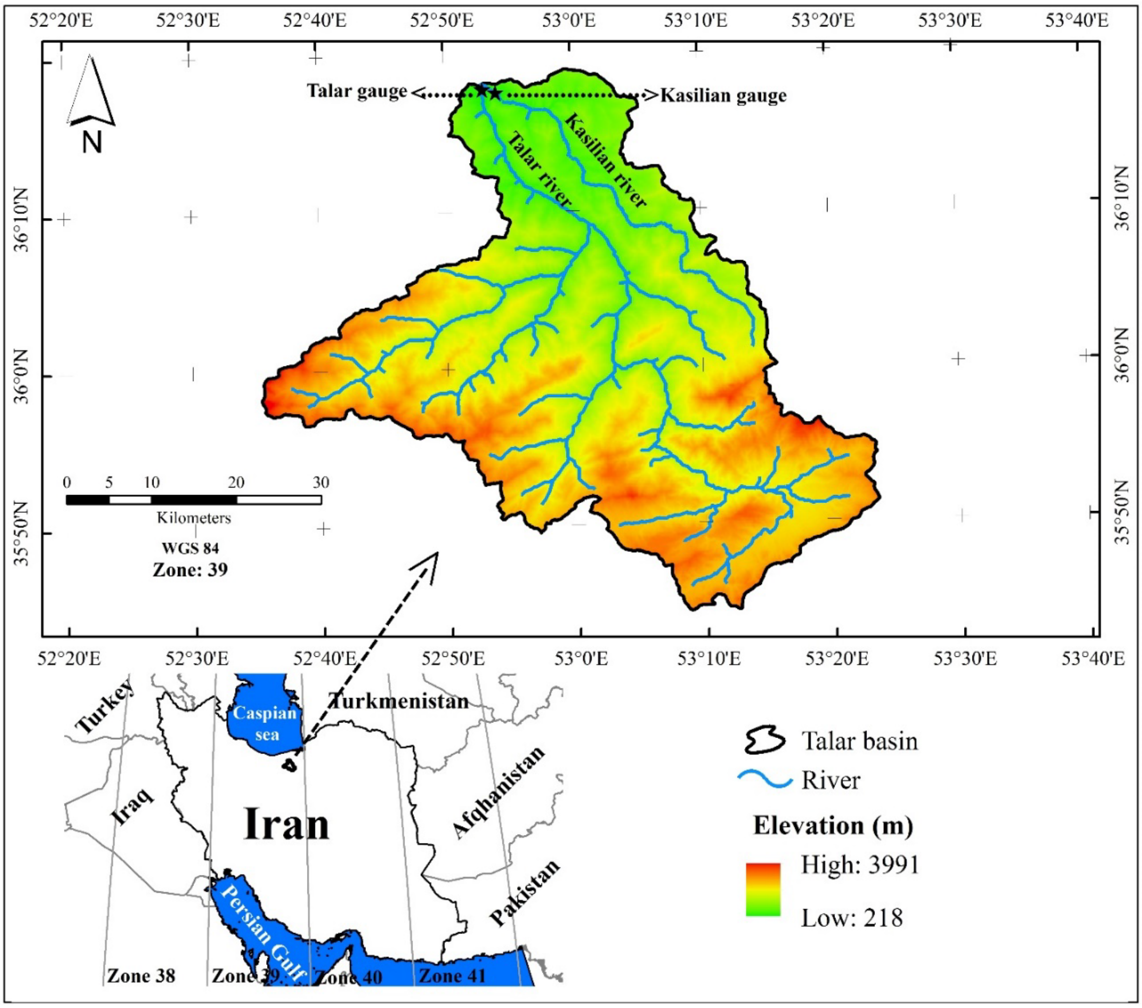

Fig. 1 Location of Case Study

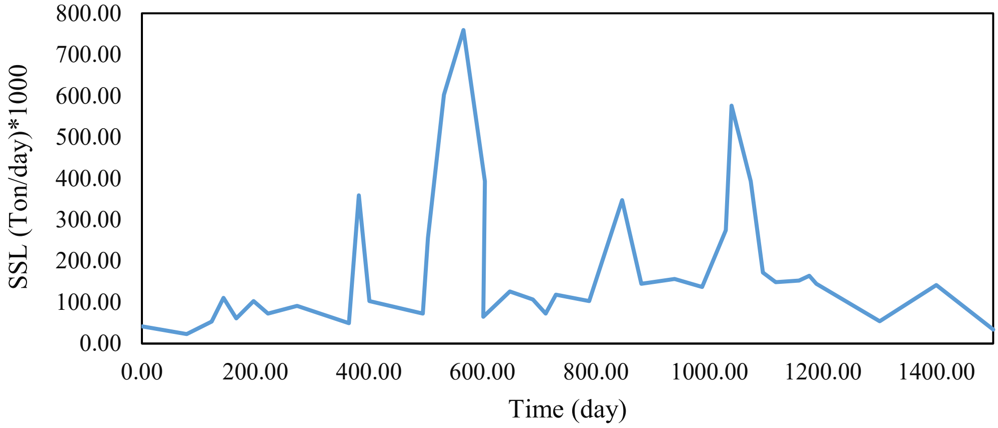

Fig. 2 Time series data of suspended sediment load

Table 1 Details of data points.

Parameter	Minimum	Average	Maximum
Discharge (m^3/s)	31	38	44
SSL (Ton/day)*1000	22.00	301.23	760
Rainfall (mm)	24	46	65

To calibrate and validate the ANFIS-M5T model, the following steps were taken:

First, the data was split into two sets: 70% of the data was used for the training set (for calibration), and 30% was used for the testing set (for validation). The ANFIS and M5T algorithms were configured. Their parameters were set up, and then the model was trained. The model was tested using the testing set. Finally, the model results were compared with the measured data. The model performance was measured using metrics like MAE (Mean Absolute Error), NSE (Nash Sutcliffe efficiency) and PBIAS (Percent bias).
Following equations (3-5), were used to assess the performance of the models:

1- Mean absolute error (MAE)

$$MAE = \frac{\sum_{i=1}^{N} \left| SSL_{es} - SSL_{ob} \right|}{N} \tag{3}$$

2- Nash Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{N} \left(SSL_{ob} - SSL_{es} \right)}{\sum_{i=1}^{N} \left(SSL_{ob} - \overline{SSL} \right)} \tag{4}$$

3- Percent bias (PBIAS)

$$PBIAS = \frac{\sum_{i=1}^{N} \left(SSL_{ob} - SSL_{es} \right)}{\sum_{i=1}^{N} SSL} \tag{5}$$

where N: number of data, SSL_{ob}: Observed data, SSL_{es}: Estimated data, and $\overline{SSL}$: Average data.

4. Results and Discussion

4.1. Determination of input parameters

By analyzing the correlations and relationships between lagged rainfall, discharge, and suspended sediment values, researchers can develop predictive models to estimate future suspended sediment loads. There is a positive correlation between river discharge and suspended sediment load. As river flow increases, it can mobilize and transport more sediment, leading to a positive correlation with suspended sediment load. Table 2 shows correlation values between SSL and input variables. There is a strong positive relationship between rainfall and suspended sediment load. The highest correlation is observed at shorter time lags (e.g., t-1, t-2, t-3). The discharge also shows a significant influence on the current sediment load, particularly for recent time periods. Similar to rainfall, the highest correlation is observed at shorter time lags (e.g., t-1, t-2, t-3). The correlation between the lagged sediment load values and the outputs is also strong. The correlation values range from 0.98 (t-1) to 0.12 (t-10). The highest correlation is observed at t-1, t-2 and t-3, respectively. We can see that the lagged suspended sediment load values have a large impact on the output prediction. If we consider a threshold of 0.8 for choosing parameters, the following parameters can be considered for predicting suspended sediment load based on the given correlation values. The 0.8 threshold is a common selection designed to enhance the model's predictive performance by leveraging robust connections between the input parameters and the target variable.
It should be noted that including delayed sediment load values in models increases our understanding of sediment transport dynamics and improves the prediction of future sediment loads. These lag values help identify temporal patterns in sediment transport. Rainfall significantly affects erosion and sediment transport by dislodging soil particles, increasing runoff, and creating overland flow. Runoff carries eroded sediments to streams and rivers and lifts suspended sediments.

4.2. Study of the accuracy of the models

Table 3a shows training results. Based on the provided metrics, the ANFIS-M5T model had the lowest MAE (525), highest NSE (0.98), and lowest PBIAS (4), indicating better performance compared to the other models. The ANFIS model had the second lowest MAE of 576, followed by MLP (586), RFN (682), and M5T (981). The ANFIS model had the second lowest PBIAS of 5, followed by MLP (7), RFN (8), and M5T (12). The MLP model had an MAE of 586, which was higher than that of the ANFIS-M5T (525) and ANFIS (576) models. The NSE value of the MLP model (0.95) was higher than that of the RFN and M5T models. The M5T model also had the highest PBIAS value (12), indicating a higher bias compared to ANFIS-M5T (4), ANFIS (5), and MLP (7). Table 3b shows testing results. The ANFIS-M5T model had the lowest MAE (545), highest NSE (0.96), and lowest PBIAS (6). The M5T model had the worst performance with the highest MAE (999), lowest NSE (0.86), and highest PBIAS (14). Compared to M5T, the RFN model was also superior in all three metrics. It had a lower MAE (694 vs. 999), higher NSE (0.90 vs. 0.86), and lower PBIAS (10 vs. 14). The MLP model also performed better than M5T. It had a lower MAE (612 vs. 999), a higher NSE (0.92 vs. 0.86), and a lower PBIAS (9 vs. 14). Therefore, MLP was more accurate than the M5T model.

Table 2 Correlation values between parameters affecting the amount of SSL

Parameter	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10
Rainfall	0.97	0.95	0.87	0.78	0.67	0.61	0.45	0.33	0.27	0.24
Discharge	0.93	0.90	0.84	0.78	0.63	0.56	0.34	0.22	0.19	0.14
Sediment load	0.98	0.92	0.83	0.76	0.65	0.54	0.32	0.23	0.20	0.12

These tables and their analysis provide a clear comparison of the models' performances in predicting suspended sediment load, allowing for informed decisions in model selection for future applications.

Table 3 Error function values for a: training results and b: testing results

(a)

Model	MAE (Ton/day)	NSE	PBIAS
ANFIS-M5T	525	0.98	4
ANFIS	576	0.97	5
MLP	586	0.95	7
RFN	682	0.92	8
M5T	981	0.89	12

(b)

Model	MAE	NSE	PBIAS
ANFIS-M5T	545	0.96	6
ANFIS	**589**	**0.94**	**8**
MLP	612	0.92	9
RFN	**694**	**0.90**	**10**
M5T	999	0.86	14

Tables 3a and 3b allows water resource managers to determine the most accurate model for predicting suspended sediment load. By comparing these models, we can identify their strengths and weaknesses. The results can be used to calibrate and validate these models for different scenarios or environments. The table serves as a benchmark for future model development. New models can be benchmarked against these existing models to evaluate their performance. If a new model performs better than those listed in the table, it can be considered the new standard for predicting suspended sediment load.

The ANFIS-M5T model seems to perform better than the other models (ANFIS, MLP, RFN, M5T). By combining ANFIS and M5T, the model may benefit from the strengths of both methods. The fusion of these models can result in a more robust and accurate predictor for complex systems. The M5T component of ANFIS-M5T is a type of decision tree that can model complex relationships between inputs and outputs. It can create multiple linear regression models at the leaves of the tree, which can provide more accurate predictions than a single global model. The adaptive aspect of ANFIS allows the hybrid model to learn and improve its performance over time. By combining multiple modeling techniques and utilizing adaptive learning mechanisms, the ANFIS-M5 reduces the risk of overfitting.

The ANFIS-M5 model benefits from the nonlinear modeling capability of ANFIS and M5. By combining the strengths of these models, the ANFIS-M5 model can capture and represent nonlinear relationships and interactions between input variables. The residuals contain valuable information that may not be fully captured by the original M5T model. By using them as inputs, the ANFIS-M5T model combines the advantages of M5T and ANFIS. By incorporating the residual values of M5T into ANFIS, the ANFIS-M5T model enhances the modeling capability by capturing and modeling the remaining complex relationships that may not have been fully captured by the M5T model. The residuals represent the discrepancies between the actual observed values and the predictions of M5T. By leveraging the residuals of the M5T model and incorporating them into the ANFIS model, the ANFIS-M5T model can refine and correct the predictions made by M5T, capture additional nonlinear relationships, and improve the overall accuracy of the suspended sediment load prediction.

The ANFIS-M5T model provides more accurate estimates of the suspended sediment load in water bodies. This information is crucial for understanding sediment transport dynamics, sedimentation patterns, and the impact of sediment on water quality. By accurately estimating suspended sediment load, the ANFIS-M5T model helps identify areas prone to erosion and sedimentation. By providing accurate suspended sediment load estimates, the ANFIS-M5T model contributes to informed decision-making, effective sediment management strategies, water quality improvement, infrastructure planning, and environmental impact assessment. By understanding sediment dynamics, managers can implement appropriate restoration activities, such as river bank stabilization, re-vegetation, and channel reconfiguration, to mitigate erosion, enhance habitat quality, and improve ecosystem functioning. By providing valuable insight into sediment load dynamics within a watershed, ANFIS-M5T helps prioritize conservation practices, land management strategies, and erosion control measures. By estimating suspended sediment load, the ANFIS-M5T model provides critical information for assessing environmental impacts on water bodies. The ANFIS-M5T model helps assess the potential environmental impacts of human activities that contribute to sedimentation, such as construction, land development, or agricultural practices.

Our study can provide valuable data for the validation and calibration of suspended sediment measurement devices. By comparing the measurements from our study with the readings from different devices, we can assess the accuracy and precision of those devices and identify potential biases or discrepancies. This information can help improve the calibration algorithms and enhance the performance of suspended sediment measurement devices. By proposing innovative approaches, our study can inspire researchers and manufacturers to explore new designs, sensor technologies, or measurement algorithms that enhance the accuracy, reliability, and efficiency of suspended sediment measurement devices.

Boxplots visually depict the distribution of a dataset, emphasizing important statistical measures like the median, quartiles, and potential outliers. Figure 3 shows boxplots of different models. The median of observed data, ANFIS-M5T, ANFIS, MLR, RFN, and M5T models was 225, 231, 237, 241.5, 247.5, and 252. 5. The ANFIS-M5T had the best performance among the other models.

Visualizing multiple models on a Taylor diagram simplifies the comparison of their performance against observed data. The reference point on the diagram denotes the reference dataset or observed data. A Taylor diagram graphically assesses and compares the performance of various models or simulations, with the radial axis representing the data's standard deviation. The angular axis represents the correlation coefficient between the model data and the reference data. Each model is represented by a data point on the diagram. The distance of a model data point from the origin represents the overall performance of the model. By examining the position of points on the Taylor diagram, we can assess the performance of models or datasets in terms of their correlation, standard deviation and centralized root mean square error (CRMSE). Figure 4 shows a Taylor diagram for comparing models. The CRMSE of the ANFIS-M5T, ANFIS, MLP, RFN, and M5T models was 186.12, 269.24, 359.14, 540.12, and 773.23, respectively. The correlation coefficient of the ANFIS-M5T, ANFIS, MLP, RFN, and M5T models was 0.99, 0.95, 0.90, 0.85, and 0.60, respectively.

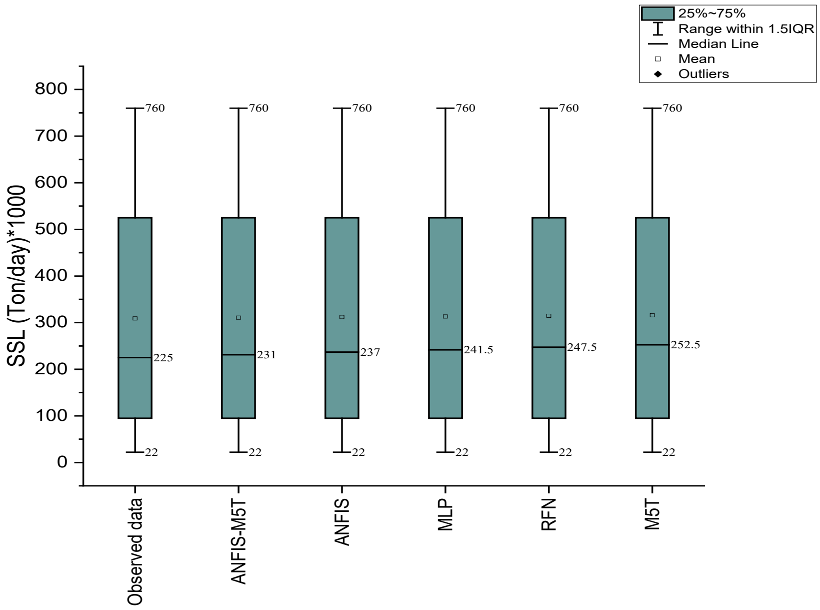

Fig. 3 Boxplots of models for comparing models

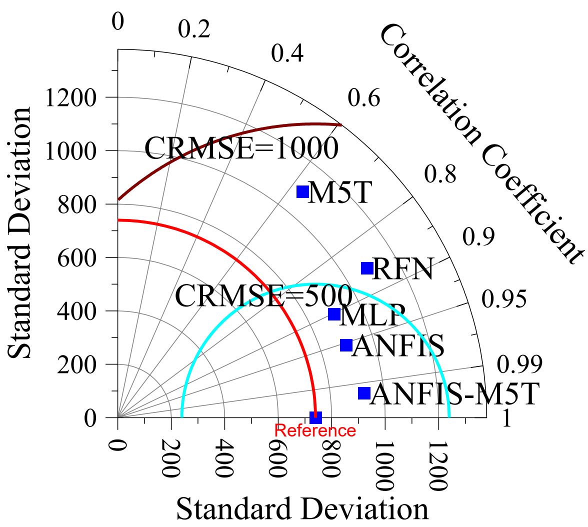

Fig. 4 Taylor diagram for comparing models

By introducing the ANFIS-M5T model and showcasing its superior performance compared to other models, our study offers an advanced method for predicting suspended sediment load. These accurate predictions can aid informed decision-making in various fields such as water resource management, erosion control, infrastructure planning, and environmental impact assessments. Our research provides a dependable tool for predicting sediment load, enabling engineers and decision-makers to design and implement effective erosion control strategies, sediment trapping systems, and sediment management practices. Sediment load predictions are crucial for infrastructure planning, particularly for water-related projects. This study assists in designing hydraulic structures, reservoirs, and flood management systems capable of managing sediment transport and deposition effectively. By addressing challenges, evaluating different approaches, and demonstrating the ANFIS-M5T model's effectiveness, our research supports the advancement of sediment monitoring and measurement techniques. Accurate prediction models are essential for enhancing the efficiency and reliability of sediment monitoring and measurement.

Sediment transport and erosion are significant contributors to natural hazards, particularly in areas prone to landslides and slope failures. By predicting suspended sediment load, the study helps to identify areas at risk of erosion and supports the implementation of erosion control measures. This information is valuable for land-use planning, slope stabilization, and erosion prevention initiatives. By providing accurate predictions and valuable insights, the ANFIS-M5T model enhances our ability to assess, monitor, and manage natural hazards related to sediment transport. The ANFIS-M5T model can serve as part of an early warning system for natural hazards related to sediment transport. Sediment transport can have adverse effects on aquatic ecosystems and biodiversity. By accurately predicting suspended sediment load, the ANFIS-M5T model can aid in the conservation and management of ecosystems affected by sediment-related hazards.

In summary, the ANFIS-M5T model contributes to natural hazard management by providing accurate predictions of suspended sediment load. This information enhances early warning systems, supports hazard mapping and risk assessment and aids in environmental conservation efforts. By incorporating sediment-related hazards into decision-making processes, the model helps reduce the impact of natural hazards and improves overall hazard management practices.

Table 4 shows MAE values of the ANFIS-M5T for one, three-, and nine-day head predictions.

Table 4 MAE values of the ANFIS-M5T model for predicting different horizons

Model	One day ahead	Three-day ahead	Nine-day ahead
	Training		
ANFIS-M5T	525	575	585
	Testing		
ANFIS-M5T	545	595	612

The ANFIS-M5T model achieved training MAE values of 525, 575, and 585 for one, three, and nine- day head predictions. The ANFIS-M5T model achieved testing MAE values of 545, 595, and 612 for one, three, and nine- day head predictions. Time series data often exhibit non-linear trends, seasonality, or other dynamic patterns that may change over time. As the forecast horizon expands, it can be difficult for models to capture and accurately predict these changing patterns. The behavior of the suspended matter load can change over time due to changing environmental conditions, natural processes or human activities. As a result, models may not be able to accurately capture and adapt to these changes, leading to reduced forecasting accuracy over longer time horizons. Natural processes and environmental factors affecting suspended sediment load show inherent variability and uncertainty. This variability increases over time, making it more challenging for the model to capture and predict the dynamics accurately. By considering the impact of prediction horizon on model accuracy, modelers can make informed decisions about model selection, data requirements, and model refinement. Increasing forecast horizons lead to additional uncertainties in the modeling process. Modelers should be aware of the sources of uncertainty associated with longer-term forecasts and consider methods of quantifying this uncertainty.

In essence, the declining accuracy of the ANFIS-M5T model as the prediction horizon extends offers valuable insights for modelers regarding model selection, uncertainty quantification, decision-making, model

refinement, and the advancement of suspended sediment load prediction. By delivering precise predictions and a deeper comprehension of suspended sediment load dynamics, our study aids decision-makers, water resource managers, and environmental agencies in making informed choices to safeguard the Talar basin from sediment-related challenges.
However, there are different methods to manage the suspended sediment load in the Telar basin. The planting and preservation of vegetation, especially along slopes, riverbanks, and critical areas, help stabilize soil, absorb rainfall, reduce surface runoff, and filter sediment.
Effective sediment management relies heavily on proper land use planning. By adopting suitable land use practices, soil disturbance is minimized, runoff is controlled, and sedimentation is reduced. Regular monitoring of sediment loads, water quality, and erosion rates offers critical insights into sediment dynamics in the Talar Basin. Involving local communities, farmers, landowners, and other stakeholders is vital for successful sediment management. It's important to recognize that sediment management demands a comprehensive and multidisciplinary strategy, requiring cooperation among government agencies, researchers, local communities, and other stakeholders. The specific measures and strategies may vary depending on the characteristics and challenges of the Talar Basin. Therefore, conducting site-specific assessments, considering local conditions, and incorporating adaptive management practices are vital for successful sediment management in the basin. Excessive sediment load can clog irrigation channels, reduce water storage capacity, and increase treatment costs for drinking water. Effective sediment control measures ensure a reliable and clean water supply and support agricultural productivity, industrial operations and economic growth. Recreational areas, natural attractions, and ecotourism potential exist within the Talar Basin. Managing the sediment load ensures the preservation of these ecosystems and their recreational value, attracting tourists and supporting the local tourism economies.

The major finding of this study is that using the capabilities of a hybrid model can improve the performance of standalone machine learning models in predicting complex variables such as suspended sediment loads. The M5T model is a simple machine learning model that cannot capture complex relationships in the data. As a result, the hybrid model enhances the performance of the M5T model by utilizing the advantages of the ANFIS model. The high capability of the new model enables it to simulate complex relationships in dynamic environments. The new model can improve the performance of the previous models. Darabi et al. (2021) combined ANFIS model with since cosine algorithm (SCA) and bat algorithm (BA) to predict SSL in the Talar basin. The testing MAE of the ANFIS-SCA and ANFIS-BA model was 1412.12 and 1423.14 while the testing MAE of the ANFIS-M5T model was 525. Thus, the new model improved the performance of the previous models. Thus, the new model can be used a key tool for managing watersheds by monitoring suspended sediment load. The new model's computer code can be used to monitor environmental pollutants. In addition, the new model can be used to predict other complex hydrological variables such as runoff, rainfall patterns and water quality parameters
However, there are important issues that should be considered as limitations of the new model. The new model cannot quantify the uncertainty values of the outputs. Thus, it should be combined with methods such as kernel density estimation to quantify the uncertainty values of outputs. Moreover, non-stationary time series may negatively impact the performance of the new model. Data processing techniques can address the problem of non-stationary data. These techniques can be combined with the new model to improve performance in predicting outputs. Furthermore, the training time of the new model may be high. Advanced computer systems can be used to address this issue.

5. Conclusion

By accurately predicting suspended sediment load, water managers can assess potential impacts on water quality, design appropriate treatment processes, and implement strategies to mitigate sediment-related issues. By implementing effective sediment control strategies, such as erosion control structures or vegetative buffers, erosion can be minimized, soil health can be preserved, and downstream sedimentation can be reduced. Our study introduces a new model for predicting SSL. We couple the ANFIS model with the M5T model to predict one- day ahead SSL. The lagged rainfall, discharge, and SSL values were used to predict SSL. The ANFIS-M5T model outperformed the other models. Based on the provided metrics, the ANFIS-M5T model had the lowest MAE (525), highest NSE (0.98), and lowest PBIAS (4), indicating better performance compared to the other models. The ANFIS model had the second lowest MAE of 576, followed by MLP (586), RFN (682), and M5T (981). By integrating fuzzy logic, adaptive learning, decision trees, and a hybrid modeling approach, the ANFIS-M5T model enhances the accuracy of sediment load predictions. Our model can capture the nonlinear relationships, adapt to changing conditions, and represent the complex interactions between input variables and sediment load. This study will provide insights into the strengths, limitations and applicability of different modeling approaches and will allow us to identify the best models for watershed management. By understanding the expected sediment load under different scenarios or land management practices, managers can identify erosion-prone areas, prioritize erosion control measures, and implement targeted sediment reduction strategies. By accurately estimating suspended sediment loads, our study enables the evaluation of potential impacts of land -use change, infrastructure development, or land management practices on sediment transport and water quality. The incorporation of the ANFIS-M5T model into watershed modeling frameworks contributes to more comprehensive and reliable watershed management. By combining sediment load predictions with hydrological models and other environmental parameters, managers can gain a better understanding of the interactions between sediment, water, and other elements of the watershed system. Next studies can focus on validating and comparing the performance of the ANFIS-M5T model with other existing sediment load prediction models.

6. Notation

ANFIS: Adaptive Neuro-Fuzzy Inference System
SSL: Suspended Sediment Load
MAE: mean absolute error (MAE
SVM: Support Vector Machine
ANN: Artificial Neural Network
MLP: MultiLayer Perceptron
RFN: Radial Basis Function Neural Network
RBF: Radial Basis Function
DENFIS: Dynamic Evolving Neural-Fuzzy Inference Systems
MARS: Multivariate Adaptive Regression Spline
PBIAS: Percent Bias
NSE: Nash Sutcliffe Efficiency
CRMSE: Centralized Root Mean Square Error

Author contribution:

Elham Ghanbari-Adivi: Conceptualization, Methodology, Formal analysis and investigation, Writing original draft preparation:

Funding:
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

7. References

Achite, M., Yaseen, Z.M., Heddam, S., Malik, A. & Kisi, O. (2022). Advanced machine learning models development for suspended sediment prediction: comparative analysis study. *Geocarto International*. *37*(21), https://doi.org/10.1080/10106049.2021.1933210.

Adnan, R.M., Yaseen, Z.M., Heddam, S., Shahid, S., Sadeghi-Niaraki, A. & Kisi, O. (2022). Predictability performance enhancement for suspended sediment in rivers: Inspection of newly developed hybrid adaptive neuro-fuzzy system model. *International Journal of Sediment Research*, *37*(3), 383-398.

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,* 7826, https://doi.org/10.1038/s41598-021-87415-4.

Bainbridge, Z., Lewis, S., Bartley, R., Fabricius, K., Collier, C., Waterhouse, J., Garzon-Garcia, A., Robson, B., Burton, J., Wenger, A. & Brodie, J. (2018). Fine sediment and particulate organic matter: A review and case study on ridge-to-reef transport, transformations, fates, and impacts on marine ecosystems. *Marine Pollution Bulletin*, *135*, 1205-1220.

Bartley, R., Bainbridge, Z.T., Lewis, S.E., Kroon, F.J., Wilkinson, S.N., Brodie, J.E. & Silburn, D.M. (2014). Relating sediment impacts on coral reefs to watershed sources, processes and management: A review. *Science of the Total Environment*, *468*, 1138-1153.

Darabi, H., Mohamadi, S., Karimidastenaei, Z., Kisi, O., Ehteram, M., ELShafie, A. & Torabi Haghighi, A. (2021). Prediction of daily suspended sediment load (SSL) using new optimization algorithms and soft computing models. *Soft Computing*, *25*, 7609-7626.

de Almeida, W.S., Seitz, S., de Oliveira, L.F.C. & de Carvalho, D.F. (2021). Duration and intensity of rainfall events with the same erosivity change sediment yield and runoff rates. *International Soil and Water Conservation Research*, *9*(1), 69-75.

Downs, P.W. & Soar, P.J. (2021). Beyond stationarity: influence of flow history and sediment supply on coarse bedload transport. *Water Resources Research*, *57*(2), e2020WR027774, https://doi.org/10.1029/2020WR027774.

Duarte, V.B.R., Viola, M.R., Giongo, M., Uliana, E. M. & de Mello, C.R. (2022). Streamflow forecasting in Tocantins river basins using machine learning. *Water Supply*, *7*, 6230-6244.

Ehteram, M., Ahmed, A.N., Latif, S.D., Huang, Y. F., Alizamir, M., Kisi, O., ... & El-Shafie, A. (2021). Design of a hybrid ANN multi-objective whale algorithm for suspended sediment load prediction. *Environmental Science and Pollution Research*, *28*, 1596-1611.

Fath, A.H., Madanifar, F. & Abbasi, M. (2020). Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems. *Petroleum*, *6*(1), 80-91.

Ghanbari-Adivi, E., Ehteram, M., Farrokhi, A. & Sheikh Khozani, Z. (2022). Combining radial basis function neural network models and inclusive multiple models for predicting suspended sediment loads. *Water Resources Management*, *36*(11), 4313-4342

Ghenai, C., Al-Mufti, O.A.A., Al-Isawi, O.A.M., Amirah, L.H.L. & Merabet, A. (2022). Short-term building electrical load forecasting using adaptive neuro-fuzzy inference system (ANFIS). *Journal of Building Engineering*, *52*, https://doi.org/10.1016/j.jobe.2022.104323.

Goyal, M.K. (2014). Modeling of Sediment Yield Prediction Using M5 Model Tree Algorithm and Wavelet Regression. *Water Resources Management*, *7*, 1991-2003.

Gupta, D., Hazarika, B.B., Berlin, M., Sharma, U. M. & Mishra, K. (2021). Artificial intelligence for suspended sediment load prediction: a review. *Environmental Earth Sciences*, *80*, 346. https://doi.org/10.1007/s12665-021-09625-3.

Heddam, S. & Kisi, O. (2018). Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, *559*, 499-509.

Jia, Y., Wang, H., Li, P., Su, Y., Wang, F. & Huo, S. (2023). Particle swarm optimization algorithm with Gaussian exponential model to predict daily and monthly global solar radiation in Northeast China. *Environmental Science and Pollution Research*, *30*(5), 12769-12784.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T. & Clark, M. (2010). Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes: An International Journal*, *24*(10), 1270-1284.

Mohammadi, B., Guan, Y., Moazenzadeh, R. & Safari, M.J.S. (2021). Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. *Catena*, *198*, 105024, https://doi.org/10.1016/j.catena.2020.105024.

Nhu, V.H., Khosravi, K., Cooper, J.R., Karimi, M., Kisi, O., Pham, B.T. & Lyu, Z. (2020). Monthly suspended sediment load prediction using artificial intelligence: testing of a new random subspace method. *Hydrological Sciences Journal*, *65*(12), 2116-2127.

Rahgoshay, M., Feiznia, S., Arian, M. & Hashemi, S.A.A. (2019). Simulation of daily suspended sediment load using an improved model of support vector machine and genetic algorithms and particle swarm. *Arabian Journal of Geosciences*, *12*, 277, https://doi.org/10.1007/s12517-019-4444-7.

Rahul, A.K., Shivhare, N., Kumar, S., Dwivedi, S. B. & Dikshit, P.K.S. (2022). Modelling suspended sediment concentration and discharge relationship using neural network and adaptive neuro-fuzzy inference system. *Arabian Journal of Geosciences*, *15*, 493, https://doi.org/10.1007/s12517-022-09744-6.

Salih, S.Q., Sharafati, A., Khosravi, K., Faris, H., Kisi, O., Tao, H., Ali, M. & Yaseen, Z.M. (2020). River suspended sediment load prediction based on river discharge information: application of newly developed data mining models. *Hydrological Sciences Journal*, *65*(4), 624-637.