ISSN: 3050-9696 E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Machine Learning-Based Computation of Transient Bed Profiles for cohesive channel bed

Hazi Mohammad Azamathulla[1*]; Kailas Banerjee[2]; Umesh K. Singh [3]; Ketaki H. Kulkarni[4;] Rhowena J Ramdeen[5]; Anton Ali[6]; Ryan Rampair[7]; Sanjeet Kumar[8]; Mehdi Fuladipanapah[9]; and Upaka Rathnayake[10]

[1*] Professor and Head, Department of Civil & Environmental Engineering, The University of the West Indies, St. Augustine Campus, Trinidad, E-mail: hazi.azamathulla@uwi.edu

[2] Associate Professor, Department of Civil and Environmental Engineering, University of West Indies, St. Augustine Campus, Trinidad and Tobago. Email: kailas.banerjee@uwi.edu

[3]Water Resources Systems Division, National Institute of Hydrology Roorkee, Roorkee, India, umesh.ais@gmail.com.

[4] Associate Professor, School of Civil Engineering, Program Coordinator, B. Tech Civil Engineering, Dr Vishwanath D Karad MIT-World Peace University, Pune, India. PIN 411038. E-mail: ketaki.kulkarni@mitwpu.edu.in

[5] Department of Civil & Environmental Engineering, The University of the West Indies, St. Augustine Campus, Trinidad, Email : rhowena.ramdeen@my.uwi.edu

[6]Department of Civil & Environmental Engineering, The University of the West Indies, St. Augustine Campus, Trinidad, E-mail: anton.ali@uwi.edu

7 Department of Civil & Environmental Engineering, The University of the West Indies, St. Augustine Campus, Trinidad, E-mail :ryan.rampair@uwi.edu

[8] Department of Civil Engineering, Koneru Lakshmaiah Education Foundation Deemed to be University, Guntur-522502, Andhra Pradesh, India email: Sanjeetiitkgp@gmail.com

[9] Department of Civil Engineering, Ramh. C., Islamic Azad University, Ramhormoz, Iran. Email: mehdi.fuladipanah@iau.ac.ir, ORCID: 0000-0001-5305-2718)

[10]Atlantic Technological University, Sligo, Ireland, Upaka.rathnayake@atu.ie

Corresponding author*: Hazi Mohamamd Azamathulla

Abstract

This study utilizes machine learning models, including Support Vector Machine (SVM), Gene Expression Programming (GEP), and Artificial Neural Network (ANN), to predict transient bed

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

profiles in cohesive channel beds composed of clay-silt-sand-gravel mixtures with clay content ranging from 10% to 50%. By leveraging 4,200 experimental data points (70% for training and 30% for testing), dimensional analysis and Γ-test for sensitivity analysis were employed to identify key dimensionless parameters: clay percentage (P_c), normalized time ($\frac{t}{t_e}$), relative density ($\frac{\gamma_b}{\gamma_w}$), void ratio (e), moisture content (w), and normalized distance ($\frac{x}{L_{max}}$). GEP demonstrated superior performance, achieving the highest testing accuracy (R^2=0.9955, RMSE=1.8456), followed by SVM (R^2 = 0.9870, RMSE = 3.2480) and ANN (R^2 = 0.9838, RMSE = 3.9143). Taylor diagrams and DDR analysis validated the robustness of Gene Expression Programming (GEP), demonstrating optimal correlation and minimal variance compared to observed data. GEP's symbolic regression yielded interpretable equations using operators (+, -, *, /, x^2, $\sqrt[3]{x}$), while SVM and ANN exhibited higher deviations for extreme values. This study emphasizes GEP's effectiveness in capturing nonlinear erosion dynamics, making it a reliable tool for predicting bed degradation in hydraulic engineering applications. The results highlight the significance of model interpretability and parameter selection, offering valuable insights for soil erosion management and sediment transport modeling.

Keywords: *Machine Learning Models, Transient Bed Profiles, Cohesive Sediments, Gamma-test Sensitivity Analysis, Hydraulic Erosion Modeling*

Introduction

Soil aggregates serve as the fundamental units of soil structure, with their particle size distribution and stability significantly affecting the onset and progression of erosion, as well as the particle size characteristics of the resulting eroded sediment. The erosion profile of a mixture containing clay, silt, and sand is shaped by several factors, including the proportions of each component and the shear stress applied by moving water (Ahmad et al., 2018; Jacob et al., 2011). A comprehensive understanding of these factors is essential for hydraulic engineering applications, such as the design of stable channels and the prevention of soil erosion. Among these critical

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

factors, the composition of the mixture plays a vital role. The relative amounts of clay, silt, and sand play a crucial role in determining erosion characteristics, as mixtures with a higher proportion of sand tend to erode more rapidly than those with a greater clay content (Singh et al., 2017; Jain and Kothyari, 2009). This variation can significantly affect erosion patterns across different environments, including coastal regions, river channels, agricultural lands, and urban drainage systems (Yang et al., 2024).

Sediment load simulation models are typically grouped into six main categories. Empirical models rely on simplified equations derived from laboratory or field data, while conceptual (or semi-empirical) models combine physical principles with empirically determined coefficients. Physics-based deterministic models solve the governing equations of fluid and sediment dynamics, and stochastic models incorporate uncertainty in sediment transport using probabilistic frameworks. Data-driven models (DDMs) utilize machine learning (ML) and artificial intelligence (AI) to identify patterns in observational data, whereas hybrid models—also known as process-guided ML models—integrate physical laws with the adaptability of ML techniques (Hamizak et al., 2024; Fuladipanah et al., 2024). Compared to traditional approaches, data-driven and hybrid models offer superior performance by effectively capturing complex sediment-flow interactions without relying on restrictive empirical assumptions or incurring the high computational costs associated with physics-based simulations. Hybrid models are particularly robust due to their incorporation of physical constraints, such as conservation laws, and the use of explainable AI (XAI) helps address the interpretability challenges often faced by purely data-driven methods (Baranwal & Das, 2024; Moradinejad et al., 2024; Bezak et al., 2024; Liu et al., 2023; Majedi-Asl et al., 2020; Fatahi & Lashkar-Ara, 2017). Numerous studies have investigated the application of data-driven and hybrid models in sediment load prediction, with several examples presented in Table 1.

Table 1 A summary of literature review of MLMs application for sediment transport

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Reference	Models included	Model application	Recommendations
Baranwal et al. (2025)	MARS, ETR	Bridge pier scour	ETR had better results.
Sammen et al. (2025)	ANN, GTB, CBR	Scour depth prediction of bridge pier	The results of ANN were more reliable.
Murtaza et al. (2025)	ANN, ANFIS, GEP, SVM, ELM	Scour depth prediction of abutment	The results showed satisfactory performance of the ELM technique.
Kheimi (2024)	ANN, GEP, GMDH, MARS,	Sediment rate in river	GMDH and GEP were more accurate than the MARS and ANN.
Fuladipanah et al. (2024)	GEP, SVM, MARS, MLP	Sediment load in river	GEP had more accurate outcomes.
Fardoost et al. (2024)	GEP, ANN	Suspended load prediction of river flow	ANN showed better results.
Moradinejad (2024)	ANFIS, SVR, GEP, GMDH	Suspended load modeling	Results' GEP were more accurate,
Fuladipanah et al. (2023)	SVM, GEP, MARS, MLP	Scour depth downstream of flip bucket	MARS was the best model.
Shalini and Roshni (2023)	GEP, M5-tree, MARS, ANFIS	Scour depth around bridge pier	ANFIS was selected as the best model.
Liu et al. (2023)	CNN	Soil erosion in coastal area	The model improved scour prediction.
Singh et. al. (2022)	XGBoost, Adaboost, RF	Cohesive sediment transport	XGBoost has better performance.
Achite et al. (2022)	MARS, M5-tree, DENFIS, ANFIS-FCM	Suspended sediment load prediction	DENFIS performed superior that the others.
Majeed et al. (2022)	MARS	Regolith geochemical grade prediction	MARS has outstanding results.
Gupta et al. (2021)	ANN, GRNN, NF, GA, CART, LR, CHAID, ELM, GEP,	Predict suspended load	ANN, SVM, and NF had acceptable results than the others.
Lashkar-Ara et al. (2021)	GP, ANFIS	Transverse Shear Stress Distribution in a Rectangular Channel	MLMs were more accurate than empirical ones.
TAO et al. (2021)	SVM, ANFIS, BN	Predict suspended load	These Ais improved sediment load prediction.
Lashkar-Ara et al. (2020)	ANN, ANFIS	Scour Depth Caused by the Aerated Vertical Jet	ANN had more accurate.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Generalized regression neural network (GRNN); Neuro-fuzzy (NF), Genetic algorithm (GA); Extreme gradient boosting (XGBoost), Adaptive boosting (Adaboost), Random Forest (RF); Classification and regression tree (CART); Convolutional neural network (CNN); Chi-squared automatic interaction detection (CHAID); Multivariate adaptive regression splines (MARS); Extra tree regression (ETR); Gradient Tree Boosting (GTB); CatBoost Regression (CBR); Aadaptive neuro-fuzzy inference systems (ANFIS): Bayesian networks (BNs); Extreme learning machines (ELM); group method of data handling (GMDH); Multivariate adaptive regression splines (MARS); Multi-layer perceptron (MLP); Support vector regression (SVR); Classical method of sediment rating curve (SRC); Dynamic evolving neural-fuzzy inference systems (DENFIS), Fuzzy c-means based adaptive neuro fuzzy system (ANFIS-FCM),

This study aimed to develop a predictive relationship for estimating transient bed profiles in mixtures of clay, silt, sand, and gravel, with clay content ranging from 10% to 50%. To model the erosion behavior of these sediment mixtures, three machine learning techniques—Support Vector Machine (SVM), Gene Expression Programming (GEP), and Artificial Neural Networks (ANN)—were employed. Additionally, the research was extended to formulate a generalized approach for computing transient bed profiles, offering a potential solution to this complex sediment transport problem.

Materials and methods

In this study, the bed erosion profile, composed of a clay-silt-sand mixture, was predicted using three machine learning models: SVM, GEP, and ANN. For this purpose, experimental data measured by Singh and Ahmad (2019) were utilized. In the first step, dimensionless parameters were extracted using the dimensional analysis method, and then the available data were divided into training and testing sets for use in the aforementioned models. Based on the proposed step-by-step procedure for each model, the optimal tuning parameters were determined, and the longitudinal profile of the eroded bed was predicted and compared with the measured values. Finally, using performance evaluation metrics, the accuracy of each model was assessed, and the best model for profile prediction was identified.

Data preparing

Experiments were carried out in a tilting flume at the Hydraulic Engineering Laboratory of the Civil Engineering Department, Indian Institute of Technology Roorkee, India. The flume's overall

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

dimensions are 16 m in length, 0.75 m in width, and 0.50 m in depth, with a designated test section starting 7.0 m from the entrance that measures 6.0 m long, 0.75 m wide, and 0.18 m deep (see Fig. 1). Discharge was measured volumetrically using a tank at the flume's end. A rectangular trap immediately following the tail gate collected the bed load, while a depth-integrated sampler positioned just before the tail gate captured the suspended load. The channel bed profile was recorded using both a two-dimensional profiler (with a 1.0 mm least count) and a flat gauge (with a 0.10 mm least count). In addition, the water surface profile was measured using a pointer gauge with a 0.10 mm least count. Measurements of both the bed and water surface were taken at 0.50 m intervals along the center line of the flume (Sing and Ahmad, 2019). Fig. 2 illustrates bed profile of texture 30% clay.

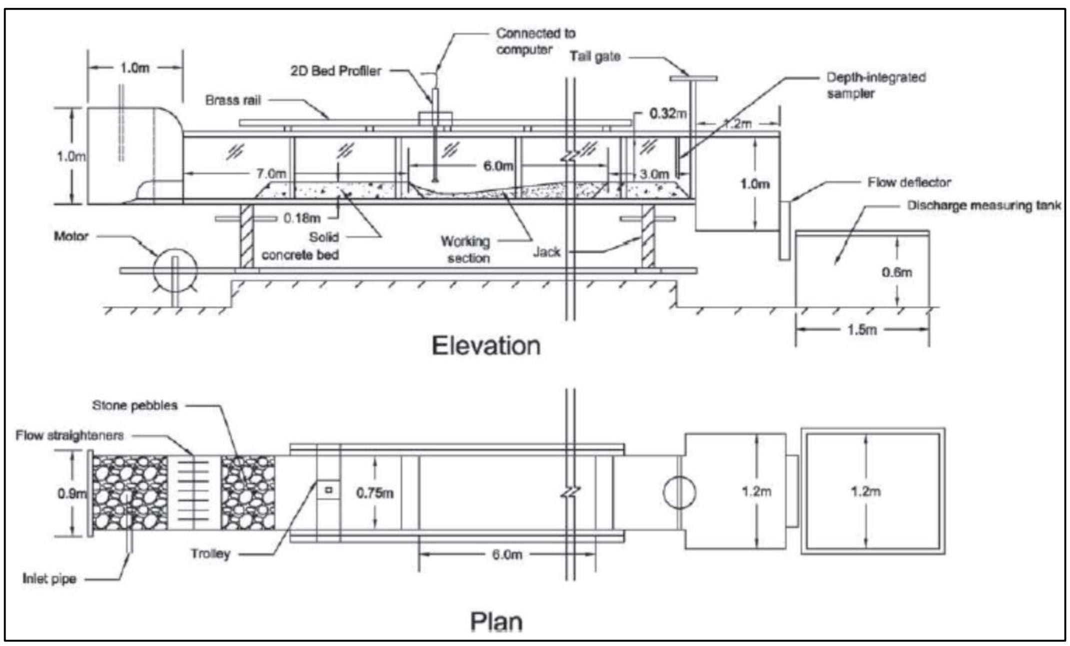


Figure 1 Schematic view of flume (Sing and Ahmad, 2019**)**

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Figure 2 Erodible bed profile of containing 30% clay during test (Sing and Ahmad, 2019**)**

Dimensional analysis

Channel bed degradation begins when the shear stress exerted on the bed becomes sufficient to mobilize and transport sediment particles. This process continues until a stable bed configuration is achieved. Transient bed profiles were recorded at the center of the flume width, with measurements taken every 50 cm along the flow direction. Initially, the rate of degradation is high but gradually decreases over time. Bed profiles were plotted relative to the original bed level, which served as the reference point for zero degradation. Visual inspection of the channel bed at the end of the experiment confirmed more pronounced degradation in the upstream section (Fig. 2). Figure 2 shows the degradation profile for 30% of clay content in clay–silt–sand-gravel mixture. Singh and Ahmad (2019), using dimensional analysis, stated that the maximum scour depth and its temporal variations can be expressed in the following form in terms of dimensionless parameters:

$$\frac{Z_{max}}{\sqrt[3]{\frac{q^2}{g}}} = f\left(P_c, \frac{\tau - \tau_c}{(\rho_s - \rho) g d_s}\right) \qquad (1)$$

$$\frac{Z}{Z_{max}} = f\left(\frac{t}{t_e}, \frac{X}{L_{max}}\right) \qquad (2)$$

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Where P_c is the clay percentage in fraction, $\tau_e = \tau - \tau_c$ is the excess shear stress (N/m^2) developed due to the incoming flow, ρ_s, and ρ are particle and fluid densities (Kg/m^3), respectively, d_s is arithmetic mean size (m) of the cohesive sediment mixture, g is gravitational acceleration (m/s^2), q is flow discharge per unit width (m^2/s), Z represents the bed degradation (m), Z_{max} denotes the maximum degradation occurring within the test section of the channel bed (m), X refers to the horizontal distance in the longitudinal direction measured from the beginning of the test section (m), and L_{max} indicates the distance from the start of the test section where the maximum degradation takes place (m). In this study, based on the experimental data obtained by Singh and Ahmad (2019), three machine learning models, namely SVM, GEP, and ANN have been employed to simulate the bed profile using the dimensionless parameters measured during the experiments (as presented in Table 2).

Table 2 Statistical Characteristics of the Measured Parameters Utilized in This Study

	Parameter	Max	Min	Mean	Standard deviation
Dependent variable	$\frac{Z}{\sqrt[3]{\frac{q^2}{g}}}$	3.129	181.777	33.415	34.507
Independent variable	P_c	0.500	0.100	0.290	0.143
	$\frac{t}{t_e}$	1.000	0.013	0.381	0.300
	$\frac{\gamma_b}{\gamma_w}$	2.146	1.729	1.957	0.115
	e	0.727	0.415	0.528	0.075
	w	0.182	0.079	0.126	0.025
	σ_{cc}	7.49	4.186	5.524	1.164
	$\frac{X}{L_{max}}$	12.00	1.000	6.500	3.452

Sensitivity analysis

Sensitivity analysis (SA) is a critical methodology in computational modeling to quantify how variations in input variables influence model outputs, enabling researchers to identify key drivers of system behavior, reduce model complexity, and enhance decision-making robustness.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Traditional SA methods, often require significant computational resources or assumptions about model structure. In contrast, the Gamma-test, a nonparametric technique, offers an efficient, data-driven alternative for assessing input relevance, particularly in nonlinear systems. Implemented via Win-Gamma software, this approach bypasses the need for explicit model formulation by directly analyzing input-output relationships from empirical data. The Gamma-test operates on the principle that for a smooth function

$$\mathbf{y}=G\mathbf{X}+\Gamma \tag{3}$$

where $\mathbf{X}=(x_1, x_2, x_3, \cdots, x_N)$ are input. Here, G represents the gradient, and Γ denotes the intercept of the regression line when x = 0, while **y** is the output variable. Lower values of G and Γ indicate that the corresponding input variables are more suitable for the model. Besides these two criteria, an additional metric called the V-Ratio is used to determine the most effective input parameters (Koncar, 1997). The V-Ratio is defined as $\frac{\Gamma}{\mathrm{var}\,(y)} = \frac{\Gamma}{\sigma^2\,(y)}$. This ratio ranges between 0 and 1, with values closer to zero signifying a stronger contribution of the respective input variable to the model. The different input variable combinations have been represented according to the format established by the Mask column. Win-Gamma software automates this analysis through a structured workflow. Deciding effective parameters hinges on three criteria (Koncar, 1997):

1. Gamma Statistic (Γ): Lower Γ values indicate inputs that collectively explain more variance in y.
2. V-ratio: Subsets with V-ratio < 0.3 are typically considered predictive, while >0.5 suggests negligible utility.
3. Standard Error (SE): Win-Gamma reports SE for Γ; smaller SE implies stable estimates.

According to the first scenario, the following equation is written to describe the bed profile:

$$\frac{Z}{\sqrt[3]{\frac{q^2}{g}}}=f\left(P_c, \frac{t}{t_e}, \frac{\gamma_b}{\gamma_w}, e, w, \sigma_{cc}, \frac{X}{L_{max}}\right)$$

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

In sensitivity analysis using the gamma test, each independent variable is represented by a binary digit, where 1 signifies its inclusion and 0 denotes its exclusion. For example, if all independent variables are accounted for in the analysis, the Mask column in Table 2, which presents a part of the sensitivity results, will show the sequence “1111111”. Conversely, the entry “1111001” in the second row of Table 3 indicates that the analysis in this case was conducted without considering the variables w and σ_{cc}. According to the Gamma-test criteria, the optimal combination of independent variables for describing the model output is determined by the lowest values of Γ, gradient, SE, and V-ratio. Based on this criterion, the combination in the first row meets all conditions, suggesting that σ_{cc} does not play a significant role in predicting the bed profile. Therefore, modeling $\frac{Z}{\sqrt[3]{\frac{q^2}{g}}}$ using machine learning models (MLMs) is carried out with the following variables: P_c, $\frac{t}{t_e}$, $\frac{\gamma_b}{\gamma_w}$, e, w and $\frac{X}{L_{max}}$.

Table 3 A brief cut of the sensitivity analysis using Γ-test

Gamma	Gradian	Standard Error	V-ratio	MASK
22.144	3849.867	2.954	0.01860	1111101
22.247	3903.485	2.659	0.01868	1111001
23.325	4657.948	3.715	0.01959	1101101
27.727	1724.151	3.734	0.02329	1111111
24.932	4103.020	3.104	0.02094	1110101
27.727	1724.151	3.734	0.02329	0111111
24.280	4883.525	3.976	0.02039	1101001
29.337	1573.356	3.397	0.02464	1111011
29.646	1689.469	3.662	0.02490	1110111
25.804	4081.698	3.211	0.02167	1110001
31.494	1648.274	3.182	0.02645	1101111
29.337	1573.356	3.397	0.02464	0111011
29.646	1689.469	3.662	0.02490	0110111
30.206	1669.150	3.450	0.02537	1110011
31.494	1648.274	3.182	0.02645	101111

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Figure 3 illustrates the variation of the Gamma statistic across different combinations of input variables, highlighting their influence on the accuracy of the model. A lower Gamma value suggests that the selected input variables contribute more effectively to predicting the bed profile by reducing the model's mean squared error. Figure 4 presents the V-ratio values for different input sets, providing insight into the predictability of each combination. Since the V-ratio measures the proportion of unexplained variance, lower values indicate a stronger relationship between the input variables and the model output, whereas higher values suggest that noise dominates the predictions. Figure 5 displays the gradient values of the regression line for various input combinations, reflecting the sensitivity of the model to changes in input variables. A lower gradient value implies that the chosen input variables are more appropriate for modeling bed profile changes, ensuring a more stable and reliable predictive model.

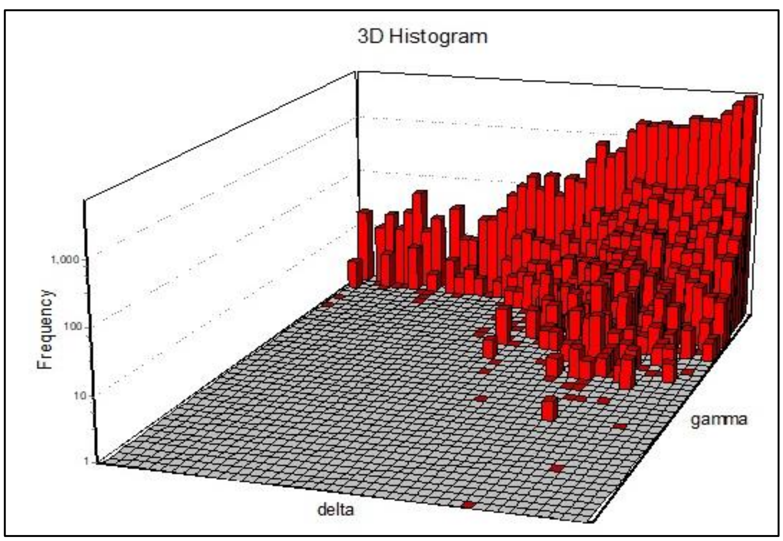


Figure 3 Γ variation for different input combinations, identifying key variables influencing model accuracy.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

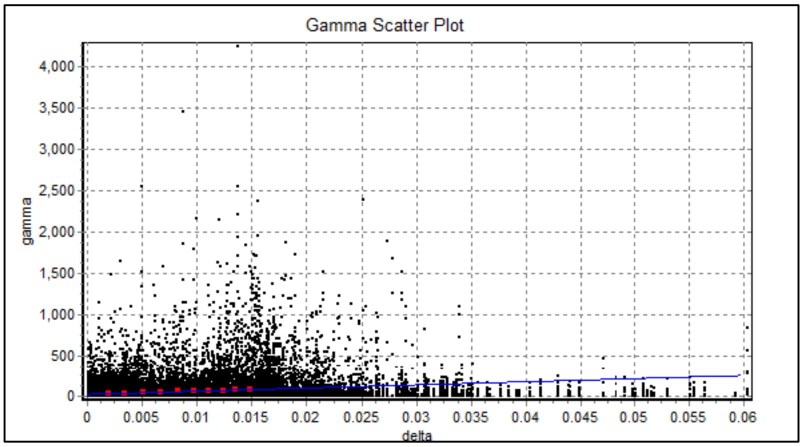


Figure 4 V-ratio analysis demonstrating the predictability strength of input variable sets

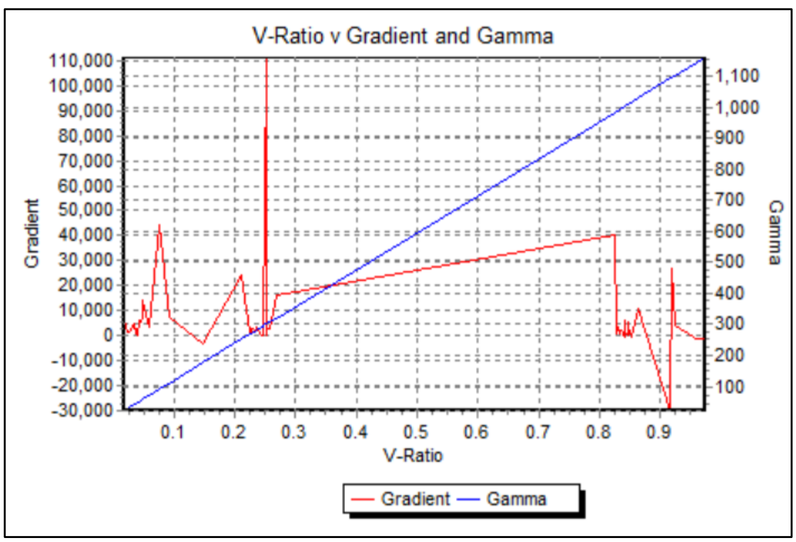


Figure 5 Gradient values highlighting the sensitivity of model predictions to input variations.

Overview of SVM

Support Vector Machines (SVMs) are a class of supervised learning algorithms commonly employed for tasks such as classification, regression, and outlier detection. Grounded in statistical learning theory and the principle of structural risk minimization, SVMs aim to identify the optimal

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

hyperplane that maximizes the margin between classes within the feature space. This margin is defined by the distance between the hyperplane and the closest data points from each class, referred to as support vectors. Due to their strong generalization capabilities, even in high-dimensional spaces, SVMs have become a widely adopted tool in various machine learning applications, including bioinformatics and image recognition (Cortes & Vapnik, 1995). Given a training dataset $\{(x_i, y_i)\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d$ represents the feature vector and $y_i \in \{-1,1\}$ denotes the class label, the goal of an SVM is to solve the following primal optimization problem:

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \tag{4}$$

subject to the constraints:

$$y_i(\omega . x_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0, \quad \forall i = 1, 2, \cdots, n \tag{5}$$

Here ω is the weight vector orthogonal to the hyperplane, b is the bias term, which shifts the hyperplane, ξ_i are slack variables that allow for soft-margin classification, enabling the model to handle misclassified or noisy data, and C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing classification errors. A larger C penalizes misclassifications more heavily, leading to a narrower margin (Cortes and Vapnik, 1995). The term $\frac{1}{2}\|\omega\|^2$ represents the margin maximization objective, as the margin is inversely proportional to ω. The constraints in Support Vector Machines (SVMs) ensure that data points are correctly classified, with slack variables (ξ_i) introduced to accommodate instances of misclassification or deviations. For datasets that are not linearly separable, SVMs utilize kernel functions to project the input data into a higher-dimensional feature space, where a linear separation becomes feasible. This transformation is efficiently handled using the kernel trick, which allows the computation of inner products in the transformed space without explicitly calculating the mapping function $\varphi(x)$. Instead, the kernel function $K(x_i, x_j)$ computes the inner product $\varphi(x_i).\varphi(x_j)$. The dual

ISSN: 3050-9696 E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

formulation of the SVM optimization problem, leveraging Lagrange multipliers α_i, is given by (Cortes and Vapnik, 1995):

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{6}$$

subject to:

$$0 \leq \alpha_i \leq C \tag{7}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{8}$$

The dual formulation is often preferred because it simplifies the optimization problem and allows for the use of kernel functions. The decision function for a new data point x is then expressed as:

$$f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b) \tag{9}$$

Kernel functions play a crucial role in SVMs by enabling the model to capture non-linear relationships in the data. Some commonly used kernel functions include:

- Linear Kernel: $K(x_i, x_j) = x_i . x_j$. This is equivalent to a standard linear SVM.
- Polynomial Kernel: $K(x_i, x_j) = (x_i . x_j + c)^d$ where c is a constant and d is the degree of the polynomial. This kernel is useful for capturing polynomial relationships.
- Radial Basis Function RBF Kernel: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, where γ controls the influence of individual data points. The RBF kernel is particularly effective for capturing complex, non-linear patterns.
- Sigmoid Kernel: $K(x_i, x_j) = \tanh(\kappa x_i . x_j + c)$, which resembles the activation function of a neural network.

The choice of kernel and its parameters (e.g., γ in the RBF kernel) significantly impacts the performance of the SVM and must be carefully tuned using techniques such as cross-validation.

The SVMs can be adapted for regression tasks by introducing an ε-insensitive loss function. The goal is to find a function f(x) that deviates from the true values by at most ε, while still maintaining the margin maximization principle. While the SVMs are inherently binary classifiers, they can be extended to multi-class problems using strategies such as one-vs-one or one-vs-all.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Nu-SVM variant introduces a parameter ν that controls the number of support vectors and margin errors, providing a more intuitive way to tune the model. Following are the steps of modeling using SVM (Cortes and Vapnik, 1995):

1- Define the Problem:
- Determine whether the task is classification, regression, or outlier detection.
- Collect and preprocess the dataset, ensuring it is clean and normalized.

2- Choose a Kernel:
- Select an appropriate kernel function based on the data structure:

3- Split the Data:
- Divide the dataset into training and testing sets (e.g., 80% training, 20% testing) to evaluate model performance.

4- Set Hyperparameters:
- Define key hyperparameters:
- Regularization parameter C: Controls the trade-off between margin size and classification errors.
- Kernel parameters (e.g., ν for RBF, degree for polynomial).
- ε for regression tasks (in SVR).

5- Train the Model:
- Use the training data to solve the optimization problem:
  - ✓ For classification: Maximize the margin while minimizing classification errors.
  - ✓ For regression: Fit a hyperplane within an ε-insensitive tube.

6- Validate and Tune:
- Perform cross-validation to assess model performance.
- Tune hyperparameters using techniques like grid search or random search.

7- Evaluate the Model
- Test the model on the testing dataset.
- Use evaluation metrics.

Overview of GEP

The GEP is an evolutionary algorithm introduced by Ferreira (2001) as an extension of Genetic Programming (GP). Unlike GP, which represents solutions as tree structures, GEP encodes

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

solutions in linear chromosomes that are later expressed as nonlinear trees. This feature allows GEP to efficiently evolve mathematical models, making it a powerful tool for symbolic regression, function approximation, and predictive modeling in various scientific and engineering applications. The GEP consists of a population of individuals, each represented by a chromosome composed of genes. Each gene consists of a head and a tail, where the head contains function and terminal symbols, and the tail consists only of terminal symbols. The expression of these genes results in an expression tree (ET), which represents the mathematical model or decision structure. The general form of a symbolic regression model evolved using GEP can be expressed as (Ferreira, 2001):

$$Y = f(X_1, X_2, \cdots, X_n) \tag{10}$$

Where Y is the target variable, X_i (for i=1, 2, ..., n) are the input variables, and ff is the function discovered by GEP through evolutionary processes. The fitness of an individual in GEP is commonly evaluated using an error function. The evolutionary process in GEP follows a sequence of steps to generate, evaluate, and evolve models iteratively. The key operators include (Ferreira, 2001):

- Selection: Individuals with higher fitness have a greater chance of passing their genes to the next generation.
- Recombination: Genetic material is exchanged between chromosomes to create new offspring.
- Mutation: Random changes are introduced in the genes to maintain diversity.
- Transposition: Certain segments of genes are moved within the chromosome to explore new solutions.

Steps of GEP model simulation procedure are as following (Ferreira, 2001):

1. Define the problem and dataset: Select the dependent and independent variables.
2. Initialize the population: Generate an initial set of chromosomes randomly.
3. Encode individuals: Represent solutions using fixed-length linear chromosomes.
4. Express chromosomes as expression trees (ETs): Convert linear genes into parseable mathematical expressions.
5. Evaluate fitness: Compute the error function to assess model performance.
6. Apply genetic operators: Perform selection, mutation, recombination, and transposition.
7. Generate a new population: Replace the existing population with evolved individuals.
8. Repeat steps 4-7: Continue the iterative process until convergence criteria (e.g., error threshold, max iterations) are met.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

9. Extract the best model: Select the highest-performing individual as the final predictive model.

Overview on ANN

The ANNs are computational models inspired by the biological neural networks in the human brain. ANNs consist of interconnected processing units, known as neurons, arranged in layers to learn and approximate complex nonlinear relationships in data. Due to their ability to capture intricate patterns, ANNs are widely used in classification, regression, and forecasting tasks (Azmathullah et al., 2005). An ANN is composed of three main layers:

- Input Layer: Receives the feature vectors.
- Hidden Layer(s): Performs intermediate computations through weighted connections and activation functions.
- Output Layer: Generates the final prediction based on the transformed data from the hidden layers.

Each neuron processes the weighted sum of inputs and passes it through an activation function f(x)f(x) such as ReLU, sigmoid, or tanh:

$$O_i = f\left(\sum_{j=1}^{n} \omega_{ij} x_j + b_i\right) \tag{11}$$

where O_i is the neuron output, ω_{ij} represents connection weights, x_j are input signals, and b_i is the bias term. The ANN learning process involves tuning the weights using an optimization algorithm, most commonly Backpropagation and Gradient Descent. The loss function, quantifies the model error. During training, weight updates are performed using:

$$\omega_{ij}^{t+1} = \omega_{ij}^{t} - \eta \frac{\partial L}{\partial \omega_{ij}} \tag{12}$$

Where η is the learning rate and L is the loss function. Steps of ANN model simulation procedure are as following (Azmathullah et al., 2005):

1. Define the problem and dataset: Select input features and output targets.
2. Preprocess the data: Normalize and split into training, validation, and test sets.
3. Initialize the network: Define architecture (number of layers, neurons per layer, activation functions).

ISSN: 3050-9696 E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

4. Forward propagation: Compute outputs layer by layer using weighted sums and activation functions.
5. Compute the loss: Evaluate error using a loss function (e.g., MSE, cross-entropy).
6. Backpropagation: Compute gradients and update weights using an optimization algorithm.
7. Iterate training process: Continue forward and backward passes until convergence or stopping criteria.
8. Evaluate performance: Assess accuracy on validation and test datasets.
9. Deploy the trained model: Use the trained ANN for real-world predictions.

Evaluation metrics

Comparing measured and predicted data using statistical metrics is essential to objectively assess model accuracy, identify systematic errors or biases, and quantify prediction uncertainty. This evaluation ensures reliable, actionable insights for refining models and supporting data-driven decisions. Coefficient of Determination, R^2, Root Mean Square Error, RMSE, Mean Absolute Error, MAE, Scatter Index, SI, and Developed Discrepancy Ratio (DDR) are included in this paper to evaluate models' outcome accuracy, as following (Fuladipanah et al., 2024):

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (13)$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \quad (14)$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \quad (15)$$

$$SI=\frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n}y_i} \quad (16)$$

$$DDR=\frac{\text{Predicted value}}{\text{Observed value}}-1 \quad (17)$$

Where y_i represents the actual value for the i^{th} data point $\hat{y}_i$ is the predicted value and n is the number of observations. R^2 explains variance proportion; higher is better. RMSE measures average error magnitude, sensitive to outliers. MAE gives mean absolute error, robust to outliers. Scatter Index normalizes error, enabling scale-independent comparison. R^2 assesses fit quality;

ISSN: 3050-9696 E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

RMSE/MAE quantify prediction errors; and SI compares relative accuracy across datasets. The DDR index measures prediction precision by analyzing how prediction errors relate to a specified benchmark distribution (Noori et al., 2010). This metric provides insights into the agreement between predicted and real outcomes at different error levels, highlighting the model's consistency across varying magnitudes of deviation. To improve clarity and visual analysis, adjusting DDR values to a standard normal distribution is essential. This involves two key stages. First, DDR data (variable x) is adjusted through a Gaussian normalization process to produce x_{DDR}. Next, a comparative plot is generated between x_{DDR} and their standardized equivalents Z_{DDR}. In this plot, clustering of errors around the central trend and elevated x_{DDR} scores reflect enhanced model precision.

Results and discussion

In this study, using 4,200 experimentally measured data points, the simulation of the eroded bed profile composed of clay-silt-sand was conducted with the assistance of three machine learning models. The proportions of training and testing processes for the models were set at 70% (including 3,000 data points) and 30% (including 1,200 data points), respectively. By categorizing the data into two classes—training and testing—the hyperparameters of each machine learning model were optimized to achieve the most accurate output. Table 4 presents the performance evaluation of each of the three models during the training and testing phases.

Table 4 Statistical metrics to assess MLMs involved during training and testing phases

	Training phase				
	RMSE	MAE	R^2	SI	DDR_{max}
SVM	7.46612	4.36531	0.97123	0.19802	2.968
GEP	4.60538	2.67282	0.98722	0.126	4.924
ANN	10.30862	5.88411	0.96491	0.25775	2.503
	Testing phase				
	RMSE	MAE	R^2	SI	DDR_{max}

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

SVM	3.2480	2.0193	0.9870	0.11192	4.970
GEP	1.8456	1.1759	0.9955	0.06449	8.536
ANN	3.9143	2.4641	0.9838	0.13184	4.311

The first evaluated model is the Support Vector Machine (SVM). As mentioned earlier, the hyperparameters of this model include C, γ, and ε. For different kernel functions, various values were considered for each of these parameters, and after performing trial and error, the optimal values for C, γ, and ε were determined as 43, 1.6, and 0.35, respectively. The accuracy of the model's output, based on performance evaluation indices, is presented in Table 3. The values of the performance indices (RMSE, MAE, R^2, SI, DDR_{max}) for the training and testing phases were (7.42, 4.36, 0.971, 0.198, 2.968) and (3.25, 2.02, 0.987, 0.111, 4.97), respectively. As observed, the model's performance in the testing phase is superior to that in the training phase, indicating the correctness of the hyperparameter optimization process. The next model utilized in this study for simulating the eroded bed profile is the Gene Expression Programming (GEP) model. Based on the defined hyperparameters for this model, the performance evaluation indices during the training and testing phases are provided in Table 3. The values of (RMSE, MAE, R^2, SI, DDR_{max}) in the training phase were (4.61, 2.68, 0.987, 0.126, 4.92), while in the testing phase, they were (1.8456, 1.1759, 0.9955, 0.0644, 8.54). It is noteworthy that the hyperparameter values of the GEP model, which resulted in these outcomes, are presented in Table 5. Figure 6 presents the tree-based representation of the GEP model output. As observed, the mathematical operators used in this model include +, -, /, * while the utilized functions consist of x^2 and $\sqrt[3]{x}$. The symbols d_0, d_1, d_2, d_3, d_4, and d_5 represent the variables P_c, $\frac{t}{t_e}$, $\frac{\gamma_b}{\gamma_w}$, e, w and $\frac{X}{L_{max}}$, respectively. The numerical constant values for the first gene are G1C0 = 9.639923 and G1C1 = -1.126495; for the second gene, they are G2C0 = 6.402008 and G2C1 = 0.793915; and for the third gene, they are G3C0 = 6.747925 and G3C1 = -9.800293.

Table 5 Tuning parameters of the GEP model's outcome

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Parameters	Value
Head size	8
Chromosomes numbers	30
Number of genes	3
Mutation rate	0.044
Inversion rate	0.1
One-point recombination rate	0.3
Two-point recombination rate	0.3
Gene recombination rate	0.1
Gene transposition rate	0.1
IS transposition rate	0.1
RIS transposition rate	0.1
Fitness function error type	RMSE
Linking function	+

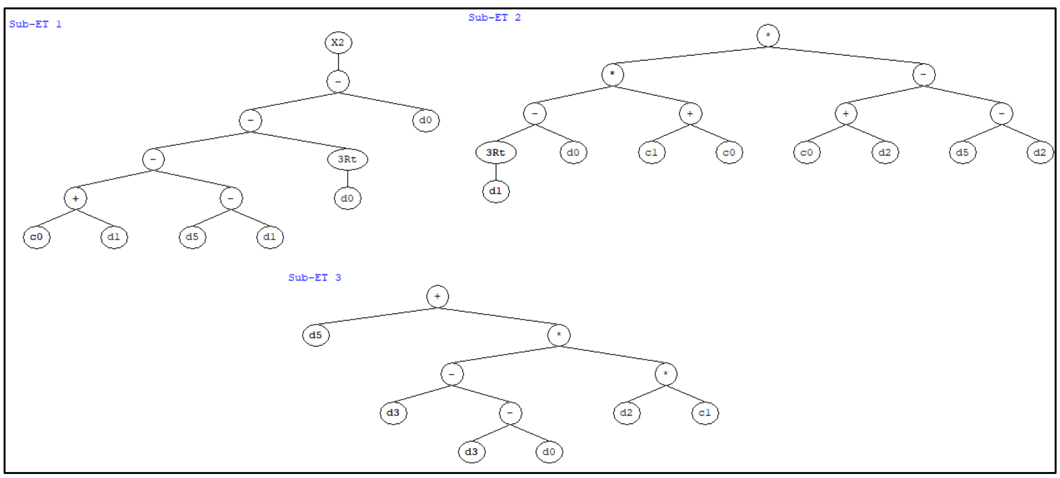


Figure 6 Tree expression of the GEP's outcome

The ANN is the third model implemented on the experimental data. To achieve the most optimized output, various network structures were considered, with a summary of their results presented in Table 6. As observed, the MLP 6-9-1 model exhibits the highest accuracy compared to other models, which have relatively lower precision. Figure 7 illustrates the simulated values of the scour depth along with the residual error.

Table 6 Performance of ANN models of different structures

Net name	Training perf.	Testing perf.	Validation perf.	Algorithm	Error function	Hidden act.
MLP 6-9-1	0.96491	0.9838	BFGS 356	SOS	Tanh	Tanh
MLP 6-11-1	0.9582	0.9765	BFGS 344	SOS	Tanh	Logistic

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

MLP 6-8-1	0.9489	0.9724	BFGS 243	SOS	Exponential	Tanh
MLP 6-5-1	0.9478	0.9658	BFGS 287	SOS	Exponential	Tanh
MLP 6-7-1	0.9467	0.9645	BFGS 228	SOS	Tanh	Logistic

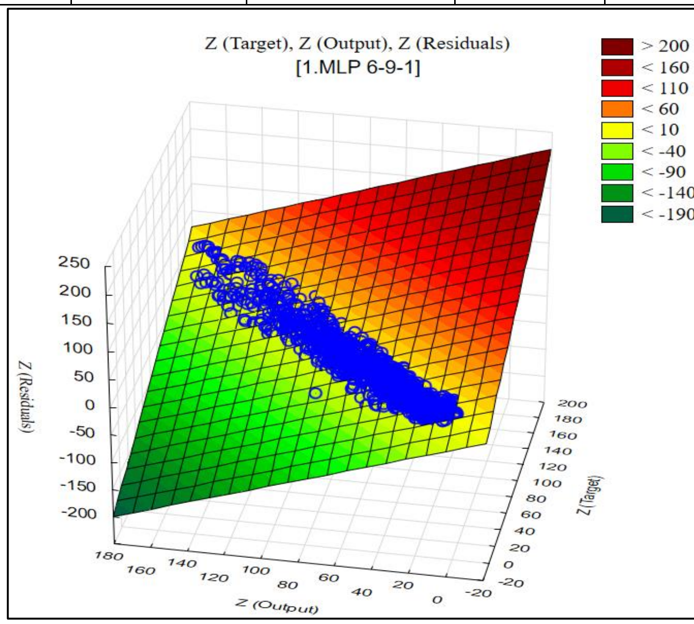


Figure 7 Performance of MLP 6-9-1 model along with residuals

Figure 9 presents a scatter plot comparing predicted and measured values of the bed profile for all three machine learning models. The closeness of the points to the 45-degree diagonal line demonstrates the accuracy of the models. GEP model exhibited the highest concentration of points near the diagonal reference line, confirming strong agreement between predicted and actual values. This is supported by its highest R^2 value (0.9955) among the models. SVM model showed slightly greater deviation, particularly for higher scour depths, indicating some limitations in capturing extreme values**.** ANN model also performed well but displayed slight dispersion around the diagonal, suggesting some residual errors in prediction, especially for mid-range bed profile values**.** Figure 10 illustrates the Developed Discrepancy Ratio (DDR) distribution for the training and testing phases across the machine learning models. The DDR metric evaluates the degree of prediction error**,** where higher values indicate better agreement between predicted and actual results**.** GEP model exhibited the highest DDR values, particularly in the testing phase, which aligns with its lowest RMSE and highest R^2 values. The higher DDR in testing compared to training further confirms model robustness and generalization ability. SVM and ANN models

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

showed lower DDR values, particularly in the training phase, indicating a higher level of residual discrepancies. However, DDR values increased in the testing phase, suggesting an improvement in generalization performance.

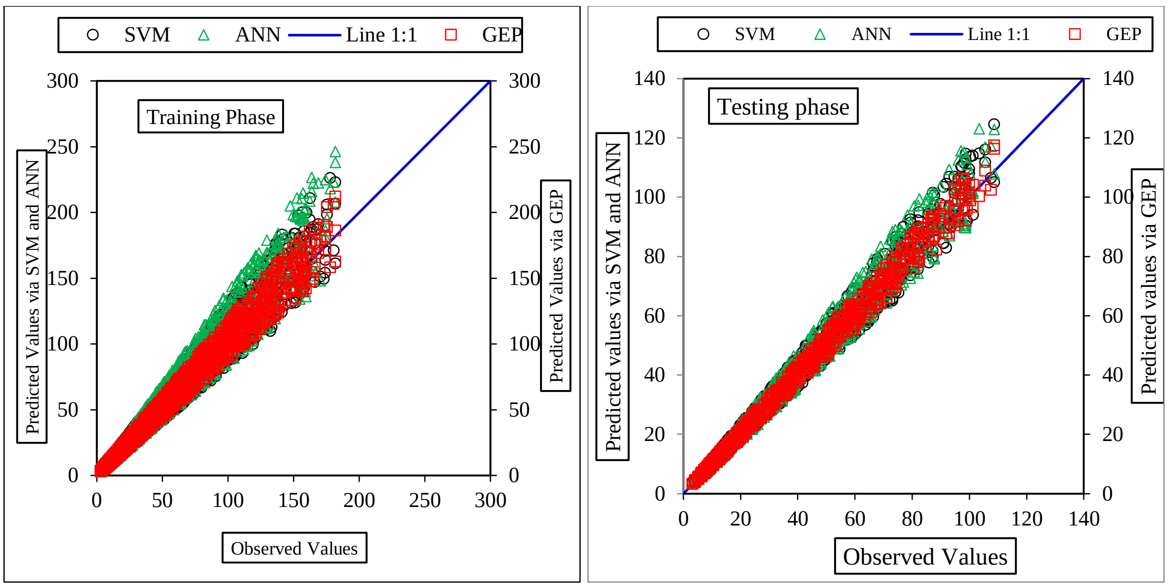
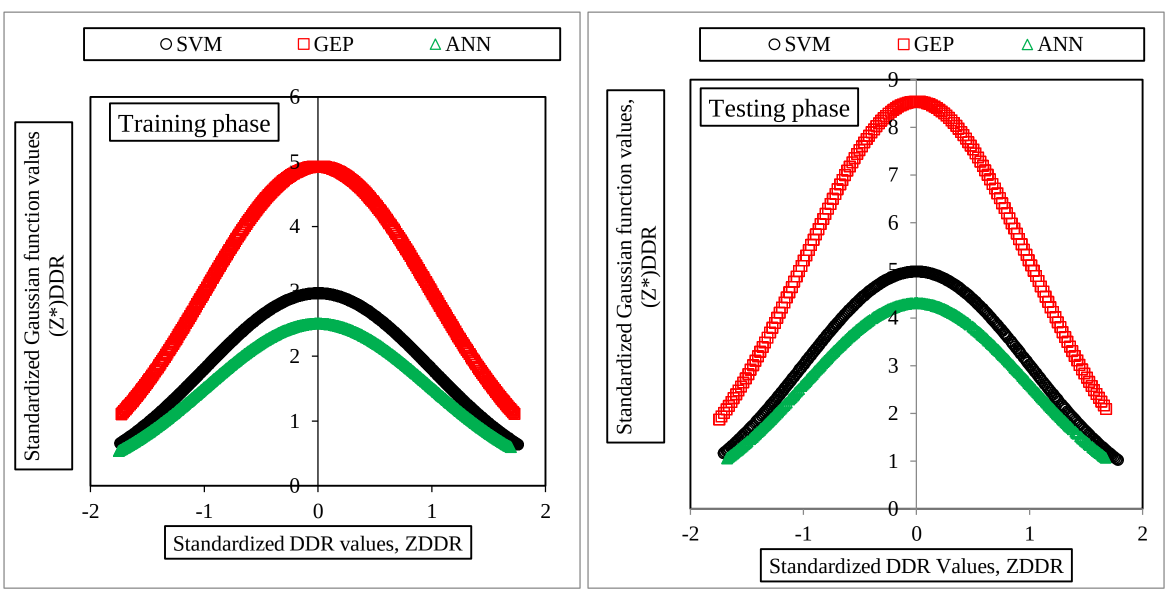


Figure 8 Scatter plot of predicted vs measured values of bed profile



ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

Figure 9 Distribution of DDR metric for training and testing phases for MLMs involved

The Taylor diagram, introduced by Taylor (2001), is a graphical performance evaluation method against observational data or benchmark datasets. This polar plot synthesizes three statistical metrics—correlation coefficient, standard deviation, and root-mean-square error- into a single visualization, enabling efficient comparison of multiple models or simulations. At its core, the diagram plots the correlation coefficient between modeled and observed data along the azimuthal angle, ranging from 0 (no correlation) to 1 (perfect correlation) on the x-axis. The radial distance from the origin represents the normalized standard deviation, quantifying data variability. A reference point denoting observations is typically placed on the x-axis at a radial distance equal to its standard deviation (often normalized to 1 for simplicity). Model outputs are plotted as points relative to this reference. Crucially, the RMSE, which measures the average deviation between model and observation, is geometrically inferred as the Euclidean distance between the model point and the reference. This relationship derives from the law of cosines, linking correlation, standard deviation, and RMSE. Thus, models closer to the reference point exhibit lower errors, realistic variance, and higher correlation. The Taylor diagram's strength lies in its ability to condense multidimensional statistical information into an intuitive format, facilitating rapid assessment of model skill. It highlights trade-offs between metrics: a model may match observed variance but lack correlation, or vice versa. Applications span climate model intercomparisons, remote sensing validation, and hydrological simulations. Figure 11 presents a Taylor diagram that evaluates the performance of the three applied MLMs (SVM, GEP, and ANN) in predicting the eroded bed profile. From the diagram, the GEP model exhibits the highest correlation coefficient, nearing 1, indicating that it closely aligns with the measured data. Additionally, its standard deviation is closest to the reference standard deviation of observed values, confirming that the model maintains realistic variability in predictions. Furthermore, the Euclidean distance between the GEP model's point and the reference point is the shortest, suggesting the lowest RMSE among the three models. The SVM model, while still exhibiting strong performance, shows a slightly lower correlation coefficient and higher RMSE compared to GEP. This indicates that although it captures the general trend of bed degradation, it introduces some deviations, particularly for extreme values. The ANN model, although effective, demonstrates the lowest correlation coefficient among the three, with a slightly higher standard deviation and RMSE. This suggests that ANN predictions have greater dispersion around measured values, making it relatively less reliable compared to GEP and SVM.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

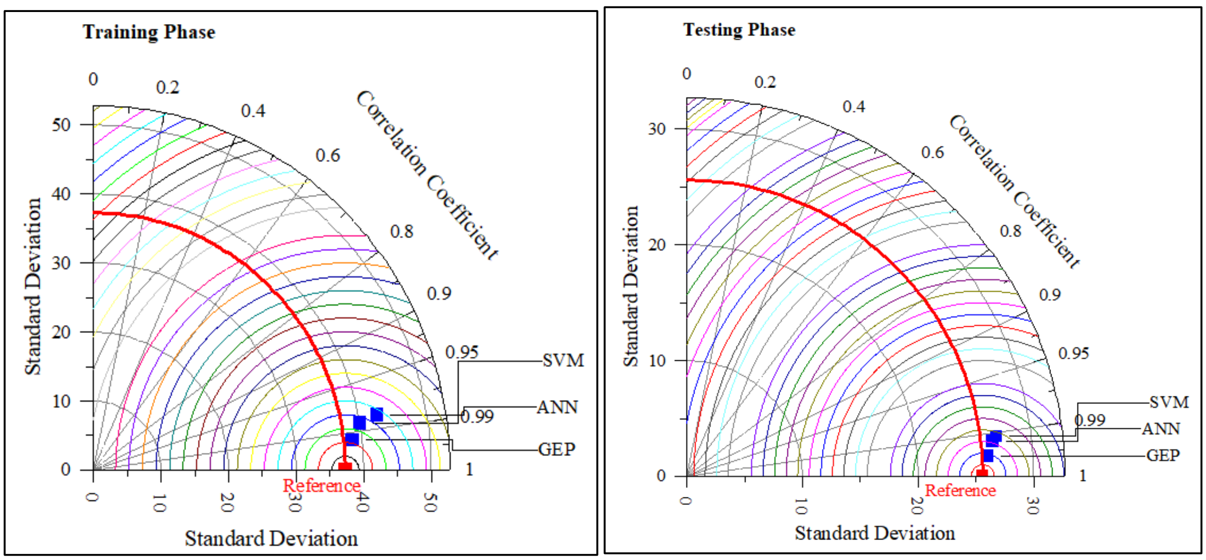


Figure 10 Performance assessment of MLMs via Tylor diagram

Conclusion

This study investigated the capability of MLMs including SVM, GEP, and ANN, in predicting the longitudinal bed profile of an eroded channel composed of a clay-silt-sand mixture. The models were trained and validated using experimental data, with key input parameters extracted through dimensional analysis. The results demonstrated that all three machine learning models effectively simulated the bed profile, but with varying degrees of accuracy. Among the models, the GEP model exhibited the highest performance, achieving the best agreement with measured values. It attained the lowest RMSE and highest R-squared value (0.9955), confirming its superior predictive capability. The Developed Discrepancy Ratio (DDR) analysis also highlighted GEP's robustness, particularly in the testing phase, where it exhibited a strong generalization ability. The SVM model followed in accuracy, with a high correlation coefficient and reasonable error metrics, though it showed slightly greater deviation in extreme values. The ANN model, while still effective, displayed more variability in predictions and relatively lower performance compared to GEP and SVM. The Taylor diagram provided further validation, indicating that the GEP model maintained the highest correlation with observed data while keeping standard deviation and RMSE values in an optimal range. This highlights its suitability for complex erosion modeling tasks. In conclusion, the findings suggest that GEP is the most reliable model for simulating the bed degradation process, making it a valuable tool for sediment transport studies and hydraulic engineering

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

applications. Future research could explore hybrid modeling approaches and incorporate additional influencing parameters to further refine prediction accuracy.

Author contributions : Hazi Mohammad Azamathulla :Supervision and Software, corresponding : Kailas Banerjee and Ketaki Kulkarni; Balaji Adusupalli ; Srinivasarao Paleti and Mahesh Racharla: Software modelling Umesh Singh ,Sanjeet Kumar and Mehdipa: Original draft, writing, data curation, modelling; Upaka Rathnayake : Supervision and Editing

Funding The author(s) received no specific funding for this work.

Data availability Data can be available with a request.

Declarations

Competing interests The authors declare no competing interests

Clinical Trial Number in the manuscript. : Not Applicable

Consent to Publish declaration: Not Applicable.

Consent to Participate declaration: Not Applicable.

Ethics declaration: Not Applicable..

References

1. 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), 6116-6140.
2. Ahmad, Z., Singh, U. K., & Kumar, A. (2018). Incipient motion for gravel particles in clay-silt-gravel cohesive mixtures. J Soils Sediments, 18, 3082–3093. doi: 10.1007/s11368-017-1869-z.
3. Amistapuram, K., Pandiri, L., Raju, V. R., Paleti, S., Singireddy, S., & Sheelam, G. K. (2025). AI-Based Cloud Infrastructure and MLOps Frameworks for Scalable Data Engineering Across Banking and Insurance. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 186–192). IEEE. 2025 IEEE International Conference on Communication Networks and Computing (CNC). https://doi.org/10.1109/cnc68716.2025.11484532

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

4. Azmathullah, HM, Deo, MC and Deolalikar, PB (2005). Neural Networks for Estimation of Scour Downstream of a Ski-Jump Bucket. Journal of Hydraulic Engineering 131 (10), 898-908
5. Baranwal, A., & Das, B. S. (2024). Live-bed scour depth modelling around the bridge pier using ANN-PSO, ANFIS, MARS, and M5Tree. Water Resources Management, 38(12), 4555-4587.
6. Baranwal, A., Kumar, S., Saurabh, S., & Das, B. S. (2025). Estimation of Bridge Pier Scour Depth Under Clear Water and Live Bed Scouring Conditions Using MARS and ETR Machine Learning Technique. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 1-27.
7. Mattaparthi, R. (2025). GenAI-Augmented Diagnostic Reasoning for Diesel Engine Fault Triage: A Large Language Model Framework for Technician Decision Support at Scale. Journal of Material Sciences & Manufacturing Research, 6(12), 1. https://doi.org/10.47363/jmsmr/2025(6)226
8. Bezak, N., Borrelli, P., Mikoš, M., Auflič, M. J., & Panagos, P. (2024). Towards multi-model soil erosion modelling: An evaluation of the erosion potential method (EPM) for global soil erosion assessments. CATENA, 234, 107596, doi: 10.1016/j.catena.2023.107596.
9. Cortes C., Vapnik V. (1995) Support-vector networks. Machine learning, 20, 273-297.
10. Fardoost, E., Dastgahi, M., Nourali, R., & Ayati, E. (2024). Assessment of GEP and ANN for Predicting Suspended Sediment Load: A Case Study of Ghatoor and Aland Rivers, West Azerbaijan, Iran. Journal of Hydraulic and Water Engineering, 1(2), 11-26.
11. Fatahi, M. and Lashkar-Ara, B. (2017). Estimating scour below inverted siphon structures using stochastic and soft computing approaches. *Journal of AI and Data Mining*, *5*(1), 55-66.
12. Ferreira, C. (2001) Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. Complex System, 13(2), 87–129.
13. Fuladipanah, M., Azamathulla, H. M., Kisi, O., Kouhdaragh, M., & Mandala, V. (2024). Quantitative forecasting of bed sediment load in river engineering: an investigation into machine learning methodologies for complex phenomena. Water Supply, 24(2), 585-600. https://doi.org/10.2166/ws.2024.017
14. Kolla, S. K. (2025). Next-Generation Precision Healthcare: AI-Driven Clinical Intelligence, Predictive Analytics, and Adaptive Decision Support Systems. International

Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 8(4), 12539-12552.

15. Fuladipanah, M., Azamathulla, H. M., Tota-Maharaj, K., Mandala, V., & Chadee, A. (2023). Precise forecasting of scour depth downstream of flip bucket spillway through data-driven models. Results in Engineering, 20, 101604.
16. 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(9), 346.
17. Hamizak, S. N., Suifa, Z., Jelania, J., & Ahmada, N. (2024). Dynamic of Sediment Transport Capacity by Overland Flow: A Review. Jurnal Kejuruteraan, 36(5), 2019–2034. doi: 10.17576/jkukm-2024-36(5)-21.
18. Siva Hemanth Kolla, Narendra Mangala. (2025). DESIGNING AUTONOMOUS LLM AGENT FRAMEWORKS USING GEN AI PIPELINES TO ENHANCE CUSTOMER SERVICE MANAGEMENT AND KNOWLEDGE WORKFLOWS. Lex Localis - Journal of Local Self-Government, 23(S6), 9719–9733. https://doi.org/10.52152/rhxpbz87
19. Jacobs, W., Le Hir, P., Van Kesteren, W., & Cann, P. (2011). Erosion threshold of sand–mud mixtures. Cont. Shelf Res. 31(10): S14–S25. doi: 10.1016/j.csr.2010.05.012.
20. Jain, R. K., & Kothyari, U. C. (2009). Cohesion influences on erosion andbed load transport. Water Resour. Res. 45(6), W06410. doi: 10.1029/2008WR007044.
21. Kheimi, M. (2024). Data-driven approaches for estimation of sediment discharge in rivers. Earth Science Informatics, 17(1), 761-781.
22. Mattaparthi, R. (2025). GenAI-Augmented Diagnostic Reasoning for Diesel Engine Fault Triage: A Large Language Model Framework for Technician Decision Support at Scale. Journal of Material Sciences & Manufacturing Research, 6(12), 1. https://doi.org/10.47363/jmsmr/2025(6)226
23. Koncar N. 1997. Optimisation methodologies for direct inverse neurocontrol. (Publication No. SW72BZ) [Doctoral dissertation London University Englan].
24. Lashkar-Ara, B., Baharvand, S., & Najafi, L. (2020). The Study of the Performance of Data-Driven Models to Predict the Scour Depth Caused by the Aerated Vertical Jet. *Irrigation Sciences and Engineering*, *43*(4), 79-89.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

25. Lashkar-Ara, B., Kalantari, N., Sheikh Khozani, Z., & Mosavi, A. (2021). Assessing machine learning versus a mathematical model to estimate the transverse shear stress distribution in a rectangular channel. Mathematics, 9(6), 596.
26. Mangalampalli, B. M., Kolla, S. K., Bandi, V. D. V. K., Yandamuri, U. S., & Rani, P. S. (2025). Designing Intelligent Healthcare Ecosystems through Adaptive Data Integration and Autonomous Learning Systems. Vascular and Endovascular Review, 8(20s), 330-347.
27. Li, K-Q., Liu, Y., & Kang, Q. (2022). Estimating the thermal conductivity of soils using six machine learning algorithms. International Communications in Heat and Mass Transfer, 136, 106139. doi: 10.1016/j.icheatmasstransfer.2022.106139.
28. Liu, C., Han Li, H., Xu, J., Gao,, W., Shen, X., & Miao, S. (2023). Applying Convolutional Neural Network to Predict Soil Erosion: A Case Study of Coastal Areas. Int. J. Environ. Res. Public Health, 20(3), 2513. doi: 10.3390/ijerph20032513.
29. Majedi-Asl, M., Daneshfaraz, R., Fuladipanah, M., Abraham, J., & Bagherzadeh, M. (2020). Simulation of bridge pier scour depth base on geometric characteristics and field data using support vector machine algorithm. Journal of Applied Research in Water and Wastewater, 7(2), 137-143.
30. Kolla, S. H. (2025). Autonomous Agentic Frameworks for Enterprise Service Operations and Intelligent Process Automation. International Journal of Science, Research and Technology, 8(4), 14643-14655.
31. Majeed, F., Ziggah, Y. Y., Kusi-Manu, C., Ibrahim, B., & Ahenkorah, I. (2022). A novel artificial intelligence approach for regolith geochemical grade prediction using multivariate adaptive regression splines. Geosystems and Geoenvironment, 1(2), 100038.
32. Moradinejad, A. (2024). Suspended load modeling of river using soft computing techniques. Water Resources Management, 38(6), 1965-1986.
33. Moradinejad, A., Khosrobeigi, S., Akbari, M., & Hosseini, S. A. (2024). Investigating soft calculation methods in river suspended sediment estimation (Hassan Abad station of Tirah river). Water and Soil Management and Modeling, 4(2), 241-260.
34. Murtaza, N., Khan, D., Rezzoug, A., Khan, Z. U., Benzougagh, B., & Khedher, K. M. (2025). Scour depth prediction around bridge abutments: A comprehensive review of artificial intelligence and hybrid models. Physics of Fluids, 37(2).
35. Raghunath Loganathan. (2025). AGENTIC AI FRAMEWORKS FOR AUTONOMOUS RISK DETECTION AND COMPLIANCE REMEDIATION IN ENTERPRISE DATA

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 03 Issue: 03
RECEIVED: JUNE 25 REVISED: JULY 09 ACCEPTED: JULY 23 PUBLISHED: AUGUST 13

CENTER OPERATIONS. Lex Localis - Journal of Local Self-Government, 23(S6), 9672–9697. https://doi.org/10.52152/3f90ak91

36. Noori R., Khakpour A., Omidvar B., Farokhnia A. 2010. Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistics. Expert Systems with Applications 37: 5856-5862.
37. Sammen, S. S., Amini, A., Othman Ahmed, K., Sadeghifar, T., Pushparaj, J., & Naganna, S. R. (2025). Improving Bridge Safety: A Spider Monkey Optimization-based ANN Model for Scour Depth Prediction. Water Resources Management, 1-23.
38. Shalini, S. H. A. L. I. N. I., & Roshni, T. (2023). Application of GEP, M5-TREE, ANFIS, and MARS for predicting scour depth in live bed conditions around bridge piers. Journal of Soft Computing in Civil Engineering, 7(4), 24-49.
39. Lebcir, I., Mageswari, S. U., Bhosale, Y. H., Nagubandi, A. R., & Mahabooba, M. M. Agile Strategic Management in the Age of Disruption: Leveraging AI and Data Analytics for Competitive Advantage.
40. Singh, U. K., & Ahmad, Z. (2019). Transport rate and bed profile computations for clay–silt–gravel mixture. Environ Earth Sci., 78, 432. doi: 10.1007/s12665-019-8419-5.
41. Singh, U. K., Ahmad, Z., & Kumar, A. (2017). Formulation for critical shear stress of cohesive sediment mixture. Current Science, 113(11), 2105-2111.
42. Singh, U. K., Jamei, M., Karbasi, M., Malik, A., & Pandey, M. (2022). Application of a modern multi-level ensemble approach for the estimation of critical shear stress in cohesive sediment mixture. Journal of Hydrology, 607, 127549, doi: 10.1016/j.jhydrol.2022.127549.
43. Tao, H., Al-Khafaji, Z. S., Qi, C., Zounemat-Kermani, M., Kisi, O., Tiyasha, T., ... & Yaseen, Z. M. (2021). Artificial intelligence models for suspended river sediment prediction: state-of-the art, modeling framework appraisal, and proposed future research directions. Engineering Applications of Computational Fluid Mechanics, 15(1), 1585-1612.
44. Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res-Atmos 106:7183–7192.
45. Yang, J., Jia, D., Zhai, B., Chen, X, & Wang, J. (2024). Bank erosion under the impacts of fluvial erosion, frost heaving/freeze-thaw process of rivers in seasonal frozen regions. Heliyon, 10(17), e37448. doi: 10.1016/j.heliyon. 2024.e37448.