

RESEARCH ARTICLE



Hybrid Machine Learning for Predicting Particle Froude Number in Auto-washout Drainage Systems with Sedimented Beds

Sanjit Kumar¹, Mayank Agarwal^{1,*} and Upaka Rathnayake^{2,*}

¹Department of Computer Science and Engineering, Indian Institute of Technology Patna, India

²Department of Civil Engineering and Construction, Atlantic Technological University, Ireland

Abstract: For proper utilization of the channel of the sewage system, an auto-washout method should be used to keep the deposited bed of the channel clean. Particle Froude number (PFR) is important in auto-washout methods. The prediction of PFR values becomes a very complex problem due to its multiple dependencies. In this study, five heterogeneous datasets collected from existing literature, covering a wide range of hydraulic and sediment conditions, were used to develop and validate the models. We used the Multilayer Perceptron Regressor (MLPR) as the base model and the Random Committee (RC-MLPR) as a hybrid machine learning (ML) model to predict the PFR value. Several performance measures were employed to assess the suggested models, including the agreement index, and other commonly used error criteria available in the published literature. The RC-MLPR model outperformed other proposed ML models, state-of-the-art ML models, and existing empirical equations. We also perform sensitive analysis that found the volumetric sediment concentration (C_{sed}) is the most sensitive variable to predict PFR value by the hybrid RC-MLPR model.

Keywords: auto-washout drainage, Froude number, Random Committee (RC), Multilayer Perceptron Regressor (MLPR), volumetric sediment concentration

1. Introduction

When sediment builds up in the sewer channel, it can cause blockages that lead to overflows that could pollute the environment [1–3]. Blockages and overflows in sewer systems can put people's health at risk because raw sewage can get into drinking water sources and spread disease [2, 4, 5]. A sewer system that cleans itself can last longer because it does not get clogged up and keeps its hydraulic capacity [6]. This means that it does not need as many expensive repairs or replacements. Therefore, an auto-washout sewer system helps to maintain its hydraulic capacity, ensuring that it can continue to transport wastewater efficiently and effectively. It also gets rid of the need to clean often by hand, which can take time and cost money.

Bertrand-Krajewski et al. [7] provided a brief overview of the understanding of the sediment transport model and comparison with experimental field data for sewer systems. It covers the different stages of the process, including buildup, wash off, transfer, transport, sedimentation, and erosion, and the various modeling approaches for each stage. Sewers are built on the idea that sediments move around and do not build up, which is called the

"auto-washout" concept. However, due to the intermittent nature of the flow, particularly at low flows when the flow is slowing down or migrating during droughts, particle deposition in sewers might still happen [8–12]. Therefore, both scenarios must be taken into consideration in order to completely comprehend the transit of sediment in sewers.

Sewers have been used to move sediment for a long time, but traditional minimum flow velocity criteria did not give a good picture of how well sewers could move sediment [13]. Sediment transport theory was used in sewer system design. However, in 1992, the Construction Industry Research and Information Association initiated a research project to develop a new design methodology for sewers based on the existing studies on sediment entrainment and erosion.

The alignment, shape, size, and bottom slope of a canal are all decided during its design process. The Shields diagram is widely used to assess sediment mobility in alluvial streams by defining the critical threshold for the initiation of sediment motion [14, 15]. Therefore, critical shear stress has a direct effect on the self-cleansing of a canal. The following Table 1 showcases some of the related literature found in discussing various types of analyses carried out for sewer canal sedimentation.

Generally, machine learning (ML)-based prediction models are very popular in different fields of research areas. As it was stated, the literature showcased some attempts to apply ML algorithms to predict the particle Froude number (PFR) in the

*Corresponding authors: Mayank Agarwal, Department of Computer Science and Engineering, Indian Institute of Technology Patna, India. Email: mayank265@iitp.ac.in and Upaka Rathnayake, Department of Civil Engineering and Construction, Atlantic Technological University, Ireland. Email: upaka.rathnayake@atu.ie

Table 1
Summary of related literature

Reference	System/context	Methodology	Key variables/focus	Main findings/contributions
Safari et al. [16]	Sewer and drainage channels	Empirical modeling	Channel geometry, auto-washout conditions	Broadly applicable sedimentation reduction models
Safari and Danandeh Mehr [17]	Large drains	Multigene genetic programming	PFr, sediment accumulation	Accurate prediction under non-sedimenting conditions
Tian et al. [18]	Fluidized bed with Geldart D particles	Numerical simulation using densimetric Froude number-based drag model	Particle Reynolds number, densimetric PFr, density ratio, liquid velocity, solid volume fraction	Proposed drag model improves prediction accuracy over existing models; densimetric Froude number effectively characterizes fluidization quality; transition to heterogeneous regime occurs at $PFr \approx 0.411$; drag coefficient decreases with increasing Fr and liquid velocity
Yi et al. [19]	Supercavitation flow in cavitation tunnel	Inviscid modeling based on potential theory (Logvinovich) with experimental validation	PFr, cavity shape, buoyancy effects	Model accurately predicts cavity shapes at high PFr (100–400); at low PFr (15–30), prediction accuracy is ~5–10% up to 2/3 cavity length; buoyancy significantly affects cavity shape at lower PFr
Song et al. [20]	Debris flow in steep mountainous regions (South Korea, Mt. Umyeon event)	Finite element modeling with Coulomb-viscous framework + SHapley Additive exPlanations (SHAP) analysis	Debris volume, liquid index, flow velocity, depth, PFr, viscosity, friction angle	Model accurately predicts impact pressure and flow behavior; $PFr > 10$ indicates high-energy debris flow; new empirical equation extends PFr range up to 40; debris volume and liquid index are most influential factors
Wan Mohtar et al. [21]	Drainage channels	Artificial Neural Network (ANN), (Radial Basis Function(RBF), Feedforward Neural Network (FFNN))	Particle diameter, flow depth	FFNN achieved highest accuracy
Roushangar et al. [22]	Sewage pipes	SVM and Gaussian Processes Regressor (GPR)	Hydraulic and sediment parameters	Kernel-based ML predicted roughness effectively
Luo et al. [23]	Open channels under ice cover	Analytical and experimental study	Incipient motion velocity, ice roughness	Derived equation incorporating ice cover effects
Mahdi et al. [24]	Wire-on-tube condenser with multi-diameter tubes (3.25–6.29 mm) using refrigerants R-134a and R-600a	Theoretical modeling using Engineering Equation Solver with dimensionless analysis	Mass velocity, Weber number (We), PFr, vapor quality, condensing temperature (35–54.4 °C)	Identified flow pattern transitions based on PFr and We; established operating ranges for different refrigerants; showed that flow regime strongly depends on vapor quality and mass velocity, with distinct transition thresholds for R-134a and R-600a
Zounemat-Kermani et al. [25]	Sanitary engineering systems	Soft computing and nature-inspired algorithms	Sediment concentration, flow parameters	Hybrid nature-inspired models improved prediction accuracy

(Continued)

Table 1
(Continued)

Reference	System/context	Methodology	Key variables/focus	Main findings/contributions
Wahl [26]	Hydraulic engineering; smooth chutes and stepped chutes	Formulations and applications of the roughness Froude number	Roughness Froude number, aeration inception, friction factor	Explains roughness Froude number and parameter. It highlights its role in predicting aeration inception and flow regime transition in stepped chutes. Also explains how one form of the parameter relates to friction effects and roughness submergence
Moradi and Mojra [27]	Cylinder flow near free surface at high Froude number	Parametric study of free-surface effects	Froude number, drag, lift, Strouhal number	High Froude number strongly affects cylinder flow: drag reduces by 59%, lift and Strouhal number increase. Near the free surface, vortex shedding is suppressed
Svantesson et al. [28]	Submerged gas blowing in compressible and incompressible systems	Computational Fluid Dynamics (CFD) simulation validated with experiments and empirical equations	Modified Froude number, gas jetting	CFD results matched experiments with a small error of 7–8%. A sufficiently high modified Froude number is required for stable jet formation. Below the limit, gas forms large pulsating bubbles, causing poor penetration-length prediction
Hřebřina et al. [29]	Boulder movement on smooth bedrock/artificial beds	Controlled flume tests with different slopes, discharge, and Froude number	Froude number, discharge threshold, drag, lift, bed friction	Entrainment was mainly controlled by Froude number. Near-critical flow showed lower resistance due to instability effects
Ma et al. [30]	Horizontal pipe with wet gas in low gas-phase Froude number region	Experimental study using circular and slotted multi-orifice plates	Pressure drop, FrG, FrL, stratified flow	Pressure drop changed strongly during transition from stratified to intermittent flow. Existing models failed in both low-FrG regions, so modified correlations using both FrG and FrL were proposed
Gul et al. [31]	Open-channel flow with sediment transport	M5RT and M5RGT compared with regression equations	Flow velocity, sediment concentration, bed slope, Froude number	ML models performed better than traditional regression equations. M5RT and M5RGT gave reliable results, with Root Mean Squared Error (RMSE) values of 1.184 and 1.071

sedimentation process. Non-sedimentation auto-washout models have been created recently to calculate the minimal velocity or sediment concentration required to meet the non-sedimented bed (NSB) condition. However, the following research gaps have been found based on the literature survey, and this paper proposes the associated solutions to the gaps that were identified.

- 1) Most of the studies usually apply a specific model for the prediction of the PFr value under NSB conditions. However, in this presented study, a standalone (Multilayer Perceptron Regressor (MLPR)) and hybrid ML approaches (Random Committee (RC)-MLPR) were utilized and compared their performance with existing empirical equations stated in the literature.
- 2) Studies often use homogeneous datasets or readings collected from similar experiments. Although modeling is challenging, when the readings are carried out in a similar environment, the prediction model usually yields good results. In this study, we have used five heterogeneous datasets, which can be found in the literature. These datasets include a variety of variables that are used to predict the value of PFr in the NSB condition. The dimensionless form of these variables was used to enhance the prediction results. Different model evaluation criteria, such as agreement index (AI), Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), and mean squared error (MSE), have been used to evaluate the accuracy of proposed models.
- 3) Existing studies usually consume all the available variables for prediction purposes. Although this may prove beneficial, a thorough analysis of how the choice of input variables affects prediction accuracy is missing. In this study, the following variables—volumetric sediment concentration (C_{sed}), the median sediment size (d_{med}), the dimensionless grain size (d_p), the hydraulic radius (R_h), and the friction factor (λ_f)—were used, and their effects on the estimation of PFr were analyzed using various input combinations. This may help the end user to ascertain the most accurate variable sequence for the estimation of PFr.
- 4) Sensitivity analysis is often missing in most studies. The sensitivity analysis describes which individual variable out of the best input combination has the highest impact on the

prediction accuracy. The most sensitive variable should be included while modeling the PFr to provide accurate results.

2. Materials and Methodology

2.1. Dataset

Five available literature datasets were used in this study for the NSB condition, which were obtained from the studies by El-Zaemey [32], Perrusquía Gonzalez [33], Ghani [34], May et al.¹, and Montes et al. [35] (refer to footnote). The total dataset consists of 454 rows in a data matrix. The major statistical values of all the variables used in this dataset are shown in Table 2. In addition, some experimentally obtained empirical equations by Montes et al. [35], Nalluri et al. [36], and Safari and Danandeh Mehr [17] were utilized to show their performance in the estimation of the PFr for the NSB scenario.

El-Zaemey [32] investigated sediment deposition in sewer systems, with particular emphasis on how sediment properties evolve over time and the conditions that promote sediment consolidation, especially during dry weather flow. Experimental studies conducted in circular channels revealed that flow behavior is strongly influenced by factors such as bed roughness, sediment thickness, and flow depth. Based on these findings, new approaches were proposed for estimating sediment transport and flow resistance, which have contributed to improved design strategies for non-deposition sewers and mitigation of sediment accumulation in existing networks.

Perrusquía Gonzalez [33] examined bedload transport over permanently deposited sediment beds in partially filled concrete pipes. The study provided valuable insights into predicting flow resistance using established methods such as Engelund–Hansen and van Rijn, as well as estimating critical shear stress through the Shields diagram. In addition, the work explored vertical velocity distribution patterns and introduced a new empirical relationship for estimating sediment transport rates. Although this formulation performed well under specific conditions, such as particular pipe sizes, sediment characteristics, slopes, and bed thicknesses, it requires further validation for broader applicability.

Despite being designed for self-cleansing, sewer systems can still experience sediment buildup under low-flow conditions. Ghani [34] analyzed available data on sediment transport in both clean and sediment-laden pipes and highlighted limitations in existing design approaches, particularly for pipes with diameters exceeding 300 mm. The study suggested that larger

sewer systems may need to accommodate a certain level of sediment deposition, even though pipes up to 1.0 m in diameter may achieve self-cleansing under appropriate conditions. Updated design equations and charts were proposed to improve system performance.

May et al.¹ focused on understanding flow resistance and sediment transport rates through controlled laboratory experiments involving non-cohesive sediments in sewer pipes. Using a 450 mm concrete pipe and different sand sizes, the study identified key hydraulic conditions required to prevent sediment deposition. The research also introduced improved theoretical models and equations for predicting bedload transport and flow resistance, demonstrating that sediment transport rates tend to be lower in rough pipes compared to smooth ones.

Many earlier sediment transport models were developed using data from relatively small pipes, typically with diameters below 500 mm, which limits their applicability to larger systems. To overcome this limitation, Montes et al. [35] conducted experiments using a 595 mm pipe and developed two new self-cleansing models. These models are applicable to both small and large sewer systems, under conditions with or without deposited beds, and they show improved predictive accuracy compared to earlier approaches.

It is important to note that the compiled datasets originate from different experimental and observational studies, which help ensure a baseline level of reliability. Additionally, to mitigate the impact of inconsistencies across datasets, all input variables were transformed into dimensionless forms, which reduces scale-related discrepancies and improves comparability across different experimental conditions.

These data include volumetric sediment concentration (C_{sed}), circular canal diameter/pipe diameter (d_{pipe}) (mm), sediment size (d) (mm), dimensionless sediment size (d_p), sediment relative density (S_g), canal bed slope (%) , volumetric sediment concentration (ppm), flow depth (Y) (mm), hydraulic radius (R_h) (mm), non-deposition mean flow velocity (V_{mean}) (m/s), canal/pipe friction factor (λ_f), deposited bed thickness (t_{bed}) (mm), deposited bed width (W_{bed}) (mm), kinematic viscosity (ν) (m²/s), fluid density (ρ_f) (kg/m³), sediment density (ρ_{sed}) (kg/m³).

2.2. Model formulation

A mathematical expression can be proposed to determine the relationship between the properties that affect the auto-washout capability of a sewerage system, as showcased in Equation (1). The expression can be used to predict the mean velocity (V_{mean}) of flow required for the sedimentation-free conditions based on the values of the other properties. It can also be used to determine the

¹May, R., Brown, P., Hare, G., & Jones, K., Self-cleansing conditions for sewers carrying sediment (Report No. SR221, Hydraulics Research Wallingford, 1989).

Table 2
Statistical properties of variables

	d_p	d_{med}/R_h	C_{sed}	t_{sed}/d_{pipe}	λ_f	PFr
Mean	73.355	0.057	0.001	0.213	0.027	4.424
Standard deviation	65.981	0.05	0.001	0.113	0.01	2.651
Minimum	11.81	0.005	0	0.001	0.003	1.263
25%	22.661	0.019	0	0.154	0.019	2.348
50%	49.739	0.041	0	0.213	0.026	3.656
75%	138.388	0.086	0	0.252	0.032	5.758
Maximum	209.777	0.233	0.01	0.4	0.065	17.453

Note: 25%, 50%, and 75%—data values are less than or equal to this value.

effect of changes in the properties on the V_{mean} and auto-washout ability of the sewage system.

$$V_{\text{mean}} = f\left(\rho_f, \rho_{\text{sed}}, g, d_{\text{med}}, v, R_h, C_{\text{sed}}, W_{\text{bed}}, t_{\text{bed}}, \lambda_f, t_{\text{sed}}, d_{\text{pipe}}, \lambda_f\right) \quad (1)$$

Equation (1) can be made dimensionless as shown in Equation (2).

$$\frac{V_{\text{mean}}}{\sqrt{g d_{\text{med}} (S_g - 1)}} = f\left(d_p, \frac{d_{\text{med}}}{R_h}, C_{\text{sed}}, \frac{W_{\text{bed}}}{t_{\text{bed}}}, \frac{t_{\text{sed}}}{d_{\text{pipe}}}, \lambda_f\right) \quad (2)$$

where the dimensionless grain size (d_p) is given by Equation (3) and S_g is the specific gravity and can be expressed as Equation (4).

$$d_p = \left[\frac{(S_g - 1)g (d_{\text{med}})^3}{v^2} \right]^{1/3} \quad (3)$$

$$S_g = \frac{\rho_f}{\rho_{\text{sed}}} \quad (4)$$

This dimensionless form is helpful because it makes it easier to understand the relative importance of each variable. This can be used to design and improve auto-washout sewage systems. The variable $W_{\text{bed}}/t_{\text{bed}}$ on the auto-washout performance is negligible and may be regarded as a constant. In addition, the left-hand side of Equation (2) represents PFr, which is a dimensionless number that represents the ratio of the inertial forces to the gravitational forces acting on the particles. It can be defined as Equation (5).

$$\text{PFr} = \frac{V_{\text{mean}}}{\sqrt{g d_{\text{med}} (S_g - 1)}} \quad (5)$$

2.3. Model description

2.3.1. Random Committee (RC)

An RC is a hybrid ML model that combines the predictions of multiple individual models to make a final prediction [37, 38]. The individual models that make up the RC are typically trained on the same dataset but may use different models, hyperparameter settings, or initializations. The predictions of the individual models are then combined in some way, such as by taking the mean or mode of the predictions, in order to make the final prediction. The idea behind using an RC is that the combination of the predictions of multiple models can lead to a more accurate overall prediction than any one of the individual models alone. RCs are often used in situations where the goal is to reduce the variance of the predictions made by a single approach, to improve the stability and robustness of the algorithm [39, 40]. They can also be used to improve the performance of a model on tasks with high noise or complexity.

2.3.2. Multilayer Perceptron Regressor (MLPR)

An MLPR is a type of ANN that uses one or more input variables to predict a continuous output value [41–44]. An MLPR has one or more hidden layers in addition to an input layer and an output layer [45, 46]. Each layer is composed of a set of artificial neurons, which are connected to the neurons in the adjacent layers via weighted connections. The weights of these connections are

adjusted during the training process in order to minimize the error between the predicted output and the actual output. The output of an MLPR is a continuous value rather than a discrete class label [47].

2.3.3. Empirical equations

Safari and Danandeh Mehr [17] proposed a new optimal model created by the multigene genetic programming (MGGP) technique to predict PFr in large drainage systems under the NSB situation. Here, normalized data values, including the bias and weights of the model's component genes, were expressed mathematically using a simplified version of the model (refer to Equation (6)).

$$\begin{aligned} Fr = 1.96 - 0.61\lambda - 0.51C_v + 1.18D_{gr}^{0.5}\lambda^{1.5} \\ + 0.61\left(2C_v + \frac{d}{R}\right)^{0.5} - 2.45\left(\frac{d}{R}\right)^{\frac{1}{8}} \end{aligned} \quad (6)$$

Where:

- λ - channel friction factor
- C_v - volumetric sediment concentration
- D_{gr} - dimensionless grain size parameter
- d - median sediment size
- R - hydraulic radius
- Fr - Particle Froude number
- (as per Safari and Danandeh Mehr notations)

Nalluri et al. [36] found that existing equations in May et al. [1] generally overestimated the transport velocity for sewers with larger fixed bed deposits. Therefore, they made modifications to the equations to make them applicable to sewers with large, fixed bed deposits (refer to Equation (7)).

$$Fr = 1.98C_v^{0.13}D_{gr}^{0.10}\left(\frac{d_{50}}{R}\right)^{-0.60}\lambda_s^{0.14} \quad (7)$$

Where:

- λ_s - channel friction factor
- C_v - volumetric sediment concentration
- D_{gr} - dimensionless grain size parameter
- d_{50} - median diameter of particles in a mixture
- R - hydraulic radius
- Fr - Particle Froude number
- (as per Nalluri et al. notations)

Montes et al. [48] developed sediment transport models in sewers using data from smaller pipes with diameters less than 500 mm. They collected experimental data on a 595 mm pipe and developed auto-washout equations. The developed equation avoids overfitting when it is considered in designing of small and large sewerage systems under the NSB conditions. The developed equation is given in Equation (8).

$$Fr = 4.79C_v^{0.209}\left(\frac{d}{R}\right)^{-0.593}\lambda^{0.058} \quad (8)$$

Where:

- λ - Darcy friction factor
- C_v - volumetric sediment concentration
- d - mean particle diameter
- R - hydraulic radius
- Fr - Particle Froude number
- (as per Montes et al. notations)

2.4. Selection of input variable combinations

Parameter selection for the ML prediction model development is highly important [49–51]. Therefore, coefficients of correlation (CC) were tested for the number of variables, and the heatmap was generated. Figure 1 depicts CC values between the input and output variables. First, we select the input variable with the highest correlation with the output variable (in terms of absolute value of CC). According to Figure 1, d_p is showcased the highest correlated variable (CC = -0.661). Therefore, it was selected as the first input variable in predicting PFR. Similarly, d_{med}/R_h , C_{sed} , t_{sed}/d_{pipe} , and λ_f were selected as the input variables.

2.5. Model development and model evaluation

Selected parameters were engaged in developing various mathematical formulations for the prediction of PFR. Then,

these combinations were used to develop prediction models using MLPR and RC-MLPR algorithms. Table 3 summarizes the combinations that were used.

The performance of the developed prediction models was tested using various evaluation metrics. These include AI, NSE, MAE, and MSE, which are frequently used in ML prediction models [49, 50, 52]. The AI is given in Equation (9).

$$AI = 1 - \frac{\sum_{i=1}^N (PFR_i - \hat{PFR}_i)^2}{\sum_{i=1}^N |PFR_i - \overline{PFR}|^2 + \sum_{i=1}^N |\hat{PFR}_i - \overline{\hat{PFR}}|^2}, \quad (9)$$

$$-\infty \leq AI \leq 1$$

where PFR_i and $\hat{PFR}_i$ are the i_{th} observed and estimated values of the PFR, respectively. $\overline{PFR}$ and $\overline{\hat{PFR}}$ are the mean of observed and estimated values of the PFR, respectively.

Furthermore, the use of a 10-fold cross-validation strategy during model development provides robustness against potential

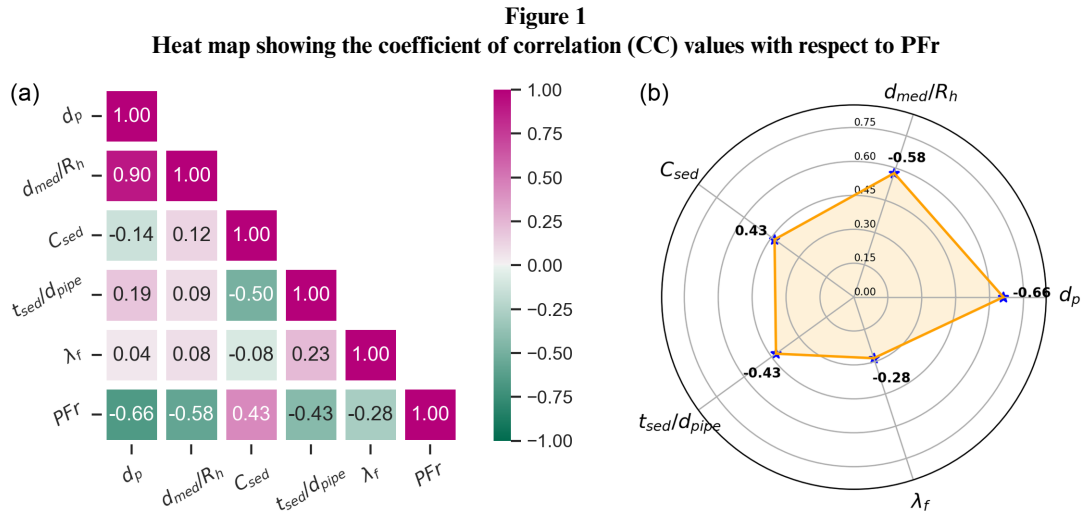


Table 3
Various combinations sorted by the CC for input variables

Models	Inputs	Combinations
MLPR	i	$f(d_p)$
	ii	$f(d_p, d_{med}/R_h)$
	iii	$f(d_p, d_{med}/R_h, C_{sed})$
	iv	$f(d_p, d_{med}/R_h, C_{sed}, t_{sed}/d_{pipe})$
	v	$f(d_p, d_{med}/R_h, C_{sed}, t_{sed}/d_{pipe}, \lambda_f)$
RC-MLPR	i	$f(d_p)$
	ii	$f(d_p, d_{med}/R_h)$
	iii	$f(d_p, d_{med}/R_h, C_{sed})$
	iv	$f(d_p, d_{med}/R_h, C_{sed}, t_{sed}/d_{pipe})$
	v	$f(d_p, d_{med}/R_h, C_{sed}, t_{sed}/d_{pipe}, \lambda_f)$

data bias by ensuring that the model is trained and tested on multiple subsets of the data. This reduces the likelihood that the model performance is driven by any specific subset of the dataset.

2.6. Sensitivity assessment

The sensitivity assessment was carried out to identify the input variable that has the highest impact on the accuracy of the prediction model. Here, two steps were involved. First, we found out which input combination has the superior estimation accuracy. Then, from the best input combination, each variable was removed one by one, and the prediction was done on the reduced set of the best input combination. Until all input variables had been removed, this process was repeated. As a result, given k inputs, we assessed the model's performance k times, removing exactly one input at a time. In this study, the minimum associated input parameter was removed first, followed by the second minimum associated parameter in the following iteration, and ultimately, the parameter with the maximum association was removed in the last iteration. We assessed the model's performance and found that the input parameter, which has the strongest influence on the estimation, by comparing metric values.

2.7. Proposed workflow

The flowchart in Figure 2 illustrates the overall methodology adopted in this study, beginning with dataset preparation and concluding with sensitivity analysis. Initially, a comprehensive dataset is compiled, consisting of key input variables such as pipe diameter, median particle size, hydraulic radius, sediment concentration, sediment transport parameters, pipe slope, and flow-related factors. The dataset is then systematically divided using a 10-fold cross-validation strategy, where the data are partitioned into training and testing subsets across ten iterations. This approach ensures robust model development and reduces the risk of overfitting by allowing each partition to serve as a test set once while the remaining data are used for training.

Following data splitting, different input variable combinations were formed (as shown in Table 3) and supplied to the selected models. The modeling framework includes the proposed ML models, RC-MLPR and MLPR, as well as several established empirical equations from the literature, namely, Safari and Danandeh Mehr [17], Nalluri et al. [36], and Montes et al. [48]. Model performance was evaluated using multiple statistical criteria as discussed in Section 2.5. Finally, a sensitivity analysis was conducted to assess the relative importance of input variables and to better understand their influence on model predictions, thereby providing insights into both model reliability and the underlying physical processes.

3. Results and Discussion

3.1. Most suitable variable arrangement

With the help of the CC value between the input variable and the output variable (see Table 3), we have a total of five different input variable combinations. It was found that the most accurate predictions are obtained when all variables are combined (input combination v). Among all the models (MLPR, RC-MLPR, and empirical equations), the hybrid RC-MLPR model produces the most accurate predictions of the PFr value (for input v).

Figure 2
Framework of the algorithms used in estimation of the PFr

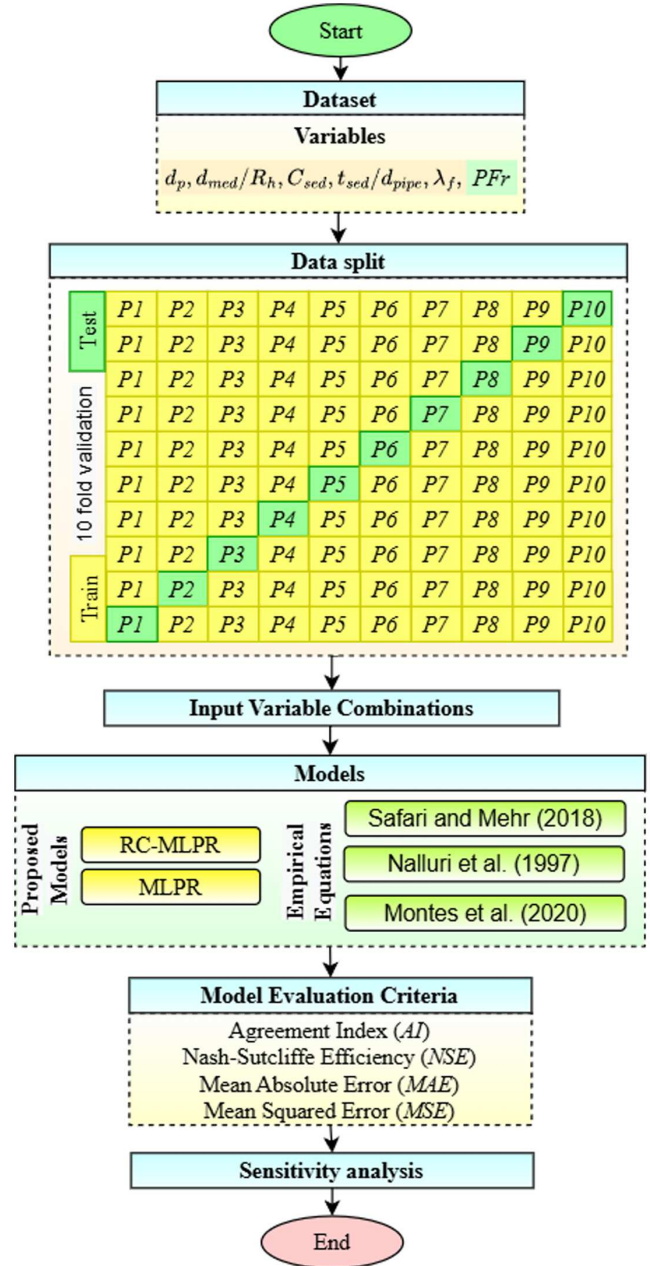


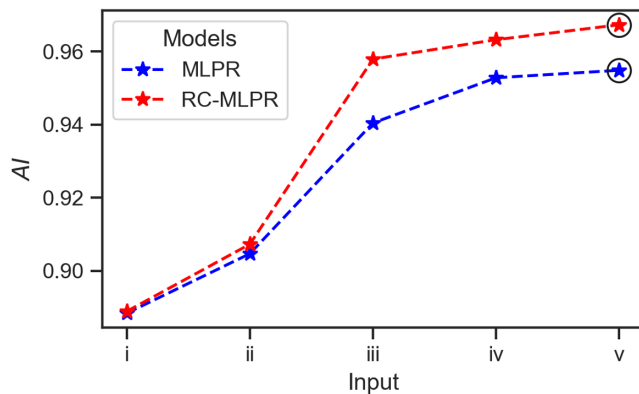
Figure 3 is a visual representation of the results of the AI, showing the performance of the model pictorially.

3.2. Model performance

Error assessment metrics of the proposed algorithms and empirically obtained expressions from the literature are represented in Table 4. These results show that according to AI value, RC-MLPR was able to most accurately predict PFr value (0.967), followed by MLPR (0.955), MGGP (0.922), genetic algorithm (GA) (0.873), bagging-linear regression (BA-LR) (0.866), Support Vector Machine (SVM) (0.792), RC-GPR (0.787), Montes et al. [48] (0.818), Nalluri et al. [36] (0.705), and Safari and Danandeh Mehr [17] (0.429), when the combination v was taken into account for estimating the PFr. According to NSE,

Figure 3

Line plot of AI value of all input combinations to predict PFr values



RC-MLPR was the most accurate predictive model (0.879), followed by MLPR (0.831), MGGP (0.703), GA (0.623), BA-LR (0.599), SVM (0.496), RC-GPR (0.477), Montes et al. [48] (−0.047), Nalluri et al. [36] (−0.021), and Safari and Danandeh Mehr [17] (−2.154).

To find the best predictive model, we also look at error metrics like MAE and MSE. According to MAE, RC-MLPR was able to most accurately predict PFr value (0.574), followed by MLPR (0.658), MGGP (0.925), GA (1.195), BA-LR (1.221), SVM (1.211), RC-GPR (1.277), Montes et al. [48] (1.647), Nalluri et al. [36] (2.005), and Safari and Danandeh Mehr [17] (3.938). According to MSE, RC-MLPR is the most accurate predictive model (0.847), followed by MLPR (1.185), MGGP (2.079), GA (2.644), BA-LR (2.808), SVM (3.531), RC-GPR (3.664), Montes et al. [48] (7.339), Nalluri et al. [36] (7.158), and Safari and Danandeh Mehr [17] (22.112).

Figure 4 ($R1 = 4a$ from Safari and Danandeh Mehr [17]; $R2 = 4b$ from Nalluri et al. [36], $R3 = 4c$ from Montes et al. [48], $4d$ from MLPR, and $4e$ from RC-MLPR) displays a line plot of predicted PFr values by the proposed algorithms and empirically obtained expressions from the literature with observed PFr values, where all actual and predicted data points are represented. The first two plots show predicted points by empirical equations, and the other two are predicted by the proposed models. From these figures, we can compare the results and observe that the proposed model predicts the PFr value more accurately with respect to empirically obtained expressions.

Figure 5 ($R1 = 5a$ from Safari and Danandeh Mehr [17], $R2 = 5b$ from Nalluri et al. [36], $R3 = 5c$ from Montes et al. [48], $5d$ from MLPR, and $5e$ from RC-MLPR) depicts the scatter plots of predicted PFr values by the proposed algorithms and empirically obtained expressions with the observed PFr values. A regression line, 95% confidence intervals (CI), and 95% prediction intervals (PI) are also shown. A CI can be represented on a scatter plot by plotting the interval as a shaded region around a point estimate, such as a mean or regression line. This region represents the range in which the true value is likely to fall, based on the level of confidence specified. For example, a 95% CI means that if the experiment were repeated many times, the true value would fall within the interval in 95% of cases. CI is a way to visualize the uncertainty of the estimate.

A PI is like a CI, but it is used to indicate the uncertainty of a single future observation rather than the mean of a population. On a scatter plot, a PI can be represented as a shaded region around a line of best fit or a predicted value. The width of the region will depend on the level of confidence specified and the distance of the individual data points from the line of best fit. It is a way to visualize the uncertainty of a prediction for a single point rather than the mean of a population.

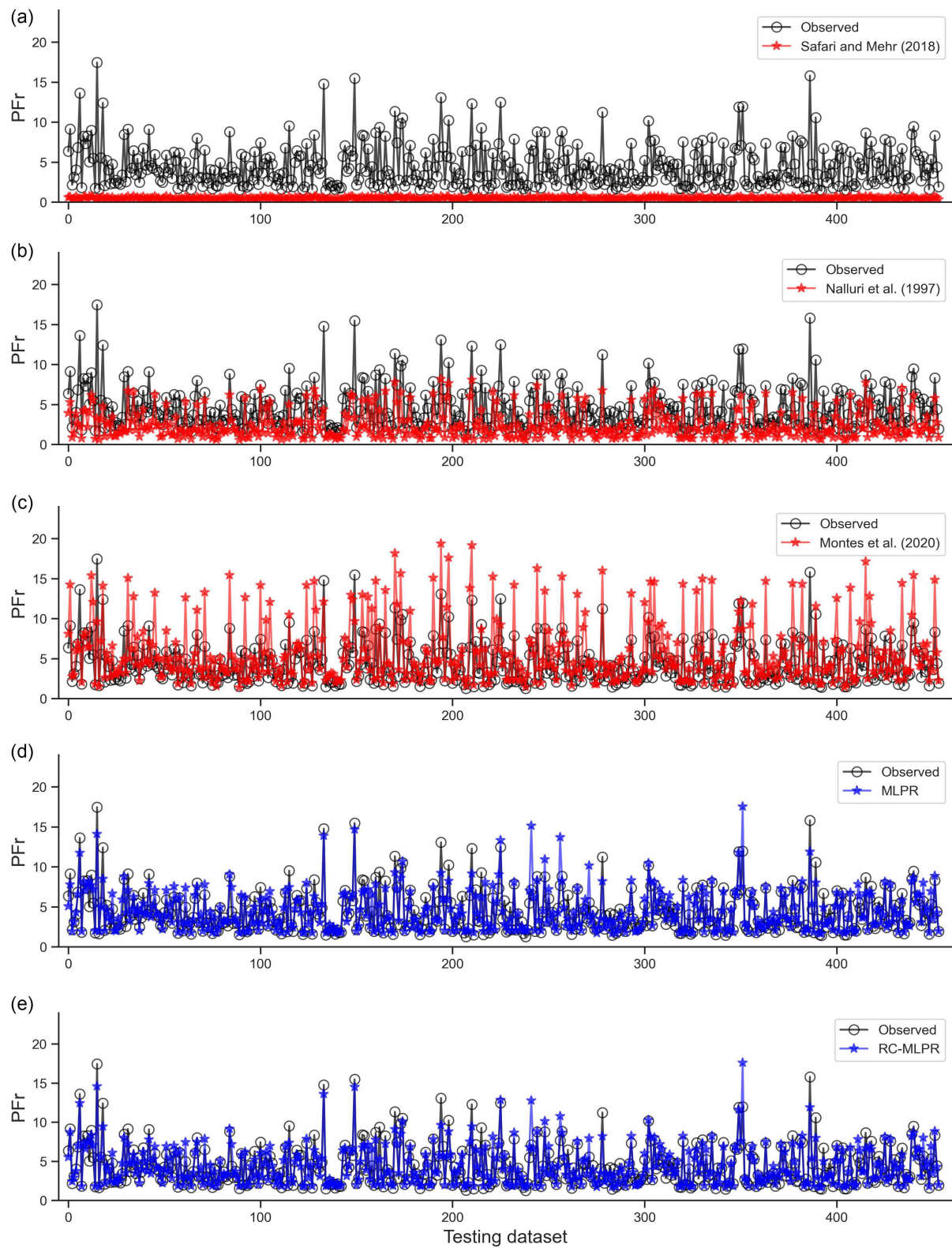
Comparing the spread of the data can be done by looking at the distance of the points from the line of best fit. If the points are tightly clustered around the line, the spread of the data is small. If the points are widely dispersed, the spread of the data is large. Outliers can be identified by looking for data points that fall far away from the main cluster of points. These outliers can have an impact on the predicted PFr value. Therefore, after an overall comparison by scatter plot, we can see that the RC-MLPR model is the best model to predict the PFr value.

Figure 6 shows the box and whisker plot. A box and whisker plot is a way to display the distribution of a dataset, showing the median, quartiles, and range of the data. The box represents the middle 50% of the data (between the first and third quartiles), while the whiskers extend to the minimum and maximum values of the data. The plot also displays any outliers, which are data points that fall outside of the range represented by the whiskers. By comparison of whiskers, we know that if the whiskers are longer, it means the data are more spread out. If the whiskers are shorter, it means the data are more tightly clustered. We can observe that the median and most quartile lines of the actual data and predicted data by the proposed models are in line with the actual values. So, we can conclude that RC-MLPR performs the best with respect to other models.

Table 4
Evaluation results of the proposed algorithms, state-of-the-art algorithms, and the empirically obtained expressions from the literature

	Model	AI	NSE	MAE	MSE
Empirical equation	Safari and Danandeh Mehr [17]	0.429	−2.154	3.938	22.112
	Nalluri et al. [36]	0.705	−0.021	2.005	7.158
	Montes et al. [48]	0.818	−0.047	1.647	7.339
State-of-the-art algorithms	RC-GPR	0.787	0.477	1.277	3.664
	SVM	0.792	0.496	1.211	3.531
	BA-LR	0.866	0.599	1.221	2.808
	GA	0.873	0.623	1.195	2.644
	MGGP	0.922	0.703	0.925	2.079
	MLPR	0.955	0.831	0.658	1.185
Proposed algorithms	RC-MLPR	0.967	0.879	0.574	0.847

Figure 4
Predicted vs observed PFr by the empirical equation of (a) $R1$, (b) $R2$, (c) $R3$, (d) MLPR, and (e) RC-MLPR



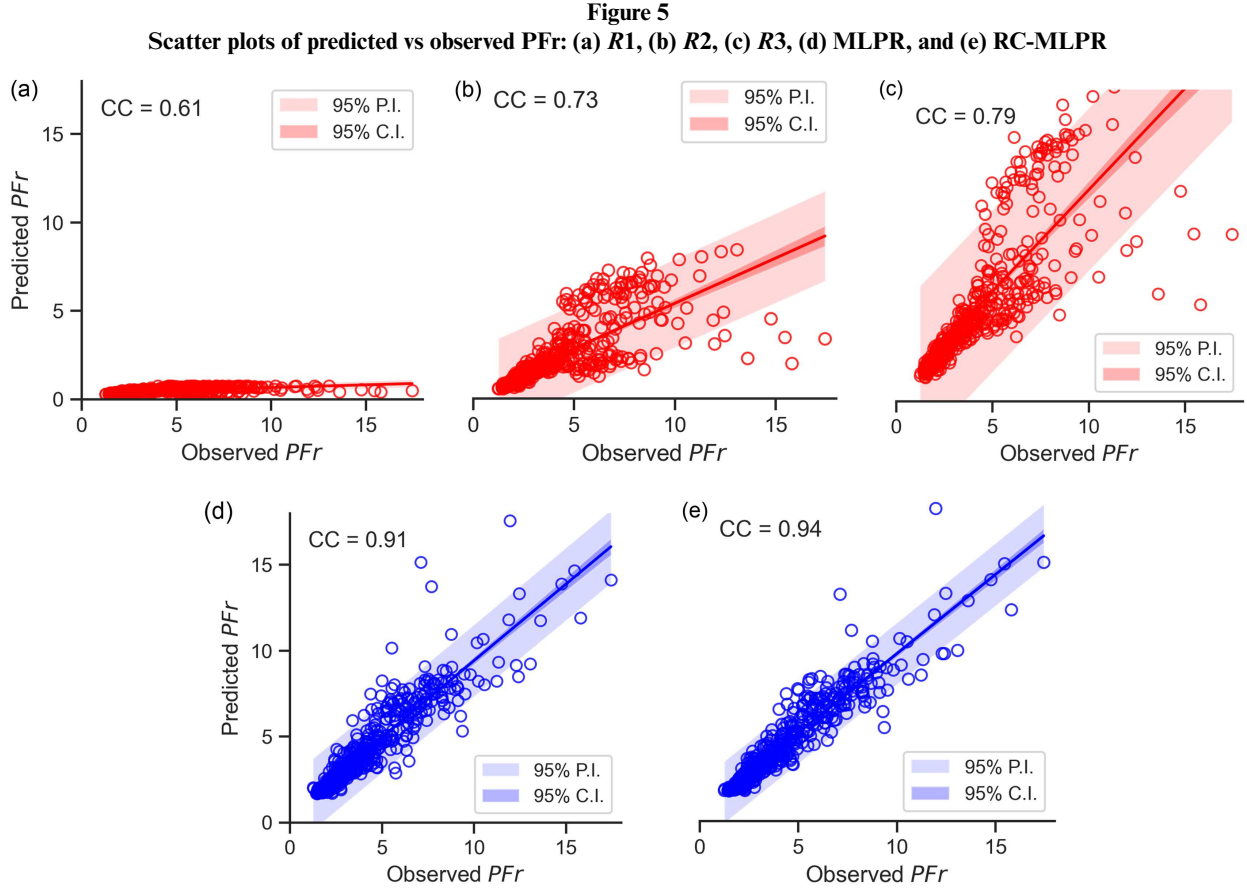
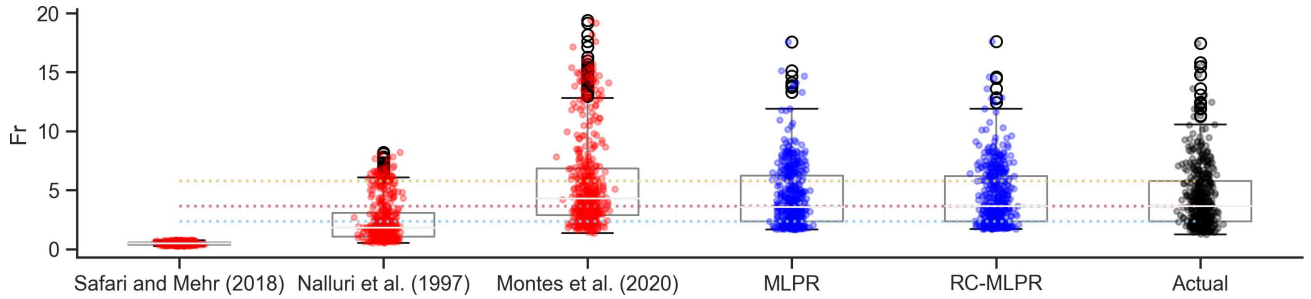


Figure 6
Observed and estimated values of PFr by the proposed algorithms and empirically obtained expressions from the literature employed in this investigation are shown in a box and whisker plot



The Taylor diagram in Figure 7 is used to compare the similarity of predicted values by proposed models (RC-MLPR and MLPR) and empirical equations with actual PFr values. The diagram plots the correlation coefficient and the standard deviation of the actual and predicted PFr values, allowing the user to compare the closeness of the different models to the actual PFr values. According to the diagram, the RC-MLPR model is the closest to the actual PFr values and is therefore concluded to be the best model for predicting PFr values in NSB conditions.

The improved performance of the RC-MLPR model can be explained by the principles of ensemble learning. The RC approach combines multiple MLPR models initialized with different weights, thereby introducing diversity among individual learners. This diversity allows the ensemble to capture complementary nonlinear patterns within the data. By aggregating predictions, the model effectively reduces variance and mitigates overfitting, resulting in more stable and accurate estimates of the PFr.

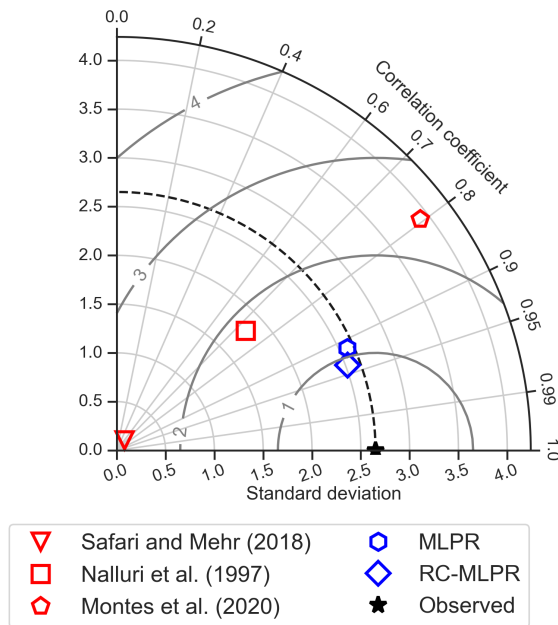
From a bias–variance trade-off perspective, the RC-MLPR model primarily reduces variance without significantly increasing bias, which explains its improved predictive performance compared to the standalone MLPR model. In addition, the ensemble structure enhances robustness to noise and inconsistencies present in the heterogeneous dataset. Since individual models may respond differently to noisy or uncertain data points, their aggregation helps smooth out extreme predictions and improves overall generalization performance.

These properties enable the model to better approximate complex nonlinear relationships among hydraulic and sediment-related variables compared to a single model.

3.3. Sensitivity assessment of the input variables/combinations

The input variables include five non-dimensional variables: d_p , $\frac{d_{med}}{R_h}$, C_{sed} , $\frac{t_{sed}}{d_{pipe}}$, λ_f . The sensitivity analysis was performed

Figure 7
Taylor diagram of proposed models and empirical equations to predict PFr value



by removing each of the input variables one at a time. For IC_1, we removed λ_f , and then we had only d_p , $\frac{d_{med}}{R_h}$, C_{sed} , $\frac{t_{sed}}{d_{pipe}}$, which are used in prediction of PFr value by RC-MLPR. Similarly, in IC_2, we removed $\frac{t_{sed}}{d_{pipe}}$ and evaluated the result on the rest input variables d_p , $\frac{d_{med}}{R_h}$, C_{sed} , λ_f . We repeated the same process until all input variables had been removed once. Table 5 showcases the results from these iterations. From the results, the elimination of, C_{sed} results in the most inadequate performance of the considered approach. Thus, we may assume that, C_{sed} is the most sensitive variable for predicting the PFr under NSB conditions.

3.4. Importance of deriving PFr in sewer systems

The sustainability of sewer systems is closely linked to the effective transport of sediments, which can be assessed using the PFr [53–55]. PFr demonstrates the ratio between flow velocity and the particle settling velocity, which indicates whether the

particles will remain in suspension or deposit along the sewer bed. Sediment deposition can reduce hydraulic capacity, increase maintenance requirements, and lead to long-term structural and operational challenges, making accurate assessment of PFr essential for sustainable sewer design. Recent research has highlighted that hybrid ML and ensemble-based approaches, such as bagging and RCs, significantly improve the prediction of PFr under both deposited and non-deposited bed conditions. By providing more reliable estimates, these methods allow engineers to optimize sewer dimensions, flow regimes, and cleaning schedules, thereby enhancing self-cleaning capacity, reducing blockages, and improving the long-term resilience and sustainability of urban wastewater networks. Furthermore, integrating data-driven predictions with traditional hydraulic models can support proactive maintenance strategies and better resource allocation, ultimately contributing to more efficient and environmentally sustainable sewer system management.

In addition, recent changes in climate have made sewer system management extremely challenging [56, 57]. Combined sewer overflows are frequent due to intensified rainfall events, which affect cities and metropolitan areas. Sedimentation in sewers reduces capacity and thus higher potential for combined sewer overflows [1, 58, 59]. Therefore, the presented research directly showcases the importance in a world that is under climate change.

3.5. Practical integration and decision-support applications

The proposed RC-MLPR model can be integrated into sewer design software and decision-support systems as a predictive module for estimating the PFr under NSB conditions. In design applications, the model can assist engineers in evaluating self-cleansing criteria, optimizing pipe dimensions, and selecting appropriate flow conditions to minimize sediment deposition. In operational settings, the model can be coupled with sensor-based monitoring systems to provide near real-time predictions, enabling proactive maintenance planning and early identification of blockage risks.

The model can be deployed either as an offline analytical tool or as part of an online decision-support system. For practical implementation, it is important to ensure reliable data acquisition, preprocessing, and periodic model recalibration using updated field data. Integration with existing hydraulic modeling platforms

Table 5
Sensitivity analysis of all input variables on prediction of PFr value by RC-MLPR model

IC #	Input variables	Removed	MAE	MSE
IC_1	d_p , $\frac{d_{med}}{R_h}$, C_{sed} , $\frac{t_{sed}}{d_{pipe}}$	λ_f	0.559	0.842
IC_2	d_p , $\frac{d_{med}}{R_h}$, C_{sed} , λ_f	$\frac{t_{sed}}{d_{pipe}}$	0.571	0.851
IC_3	d_p , $\frac{d_{med}}{R_h}$, $\frac{t_{sed}}{d_{pipe}}$, λ_f	C_{sed}	0.694	1.156
IC_4	d_p , C_{sed} , $\frac{t_{sed}}{d_{pipe}}$, λ_f	$\frac{d_{med}}{R_h}$	0.66	0.909
IC_5	$\frac{d_{med}}{R_h}$, C_{sed} , $\frac{t_{sed}}{d_{pipe}}$, λ_f	d_p	0.622	1.003
IC_6	d_p , $\frac{d_{med}}{R_h}$, C_{sed} , $\frac{t_{sed}}{d_{pipe}}$, λ_f	–	0.547	0.801

can be achieved through application programming interfaces or embedded computational modules.

4. Concluding Remarks

The existing auto-washout drainage network design methods, which rely on the optimum flow velocity and critical shear stress, have shortcomings since they ignore a number of vital hydraulic design elements. As a result, the performance of these traditional auto-washout methods is below par. In order to improve their performance, it is critical to take into account a larger number of characteristics while creating more accurate auto-washout algorithms. It is important to note that when creating a robust model, three elements should be taken into account: the variety of experimental data used, the inclusion of heterogeneous datasets, and the use of cutting-edge hybrid ML techniques that offer better prediction models.

The accurate prediction of PFr value plays a significant role in designing auto-washout sewer systems because it affects the flow behavior and transport of sediment in the sewer channel. By accurately predicting the PFr value, engineers can design the sewer system with the appropriate slope and dimensions to promote auto-washout and prevent sediment deposition. This can help to maintain the hydraulic capacity of the sewer channel, reduce the risk of blockages and overflows, and improve the overall efficiency and longevity of the system. In this study, two ML models, RC and MLPR, are used in the prediction of PFr values under NSB conditions.

Furthermore, this study finds that MLPR and RC-MLPR are both effective ML techniques for building prediction models for the prediction of PFr under NSB conditions. As a result, in actual use, they can serve as useful prediction models for sewer pipe design. Hybrid ML methods often have a flexible structure that often results in their better performance as compared to standalone methods. The AI, NSE, MAE, and MSE metrics are used to determine the effectiveness of proposed ML models. Overall, both ML models (RC-MLPR and MLPR) performed the best with Input v, when compared to all other input variable combinations. We discovered that RC-MLPR performed accurately, followed by MLPR, and empirically obtained expressions from the literature. The AI, the NSE, the MAE, and the MSE were used as performance indicators to assess the suggested algorithms. The RC-MLPR outperformed MLPR and empirical equations in terms of performance (AI = 0.967, NSE = 0.879, MAE = 0.574, and MSE = 0.847). We have also conducted a sensitivity assessment on the top-performing model (RC-MLPR) and found that the sediment concentration (C_{sed}) is the parameter, which is the most influential variable in the estimation of PFr. While the proposed RC-MLPR model demonstrates strong comparative performance, the results should be interpreted with consideration of underlying data variability. Future work will focus on incorporating uncertainty quantification and validating the model using standardized, high-quality datasets from real-world sewer systems.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available at <https://doi.org/10.5281/zenodo.20591098>.

Author Contribution Statement

Sanjit Kumar: Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Mayank Agarwal:** Conceptualization, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision. **Upaka Rathnayake:** Conceptualization, Validation, Writing – review & editing, Supervision, Project administration.

References

- [1] Wu, J., Ma, Y., & Song, S. (2024). Reducing particle accumulation in sewers for mitigation of combined sewer overflow impacts on urban rivers: A critical review of particles in sewer sediments. *Water Science & Technology*, 89(1), 89–115. <https://doi.org/10.2166/wst.2023.394>
- [2] Botturi, A., Ozbayram, E. G., Tondera, K., Gilbert, N. I., Rouault, P., Caradot, N., . . . , & Fatone, F. (2021). Combined sewer overflows: A critical review on best practice and innovative solutions to mitigate impacts on environment and human health. *Critical Reviews in Environmental Science and Technology*, 51(15), 1585–1618. <https://doi.org/10.1080/10643389.2020.1757957>
- [3] Chen, Y., Shi, X., Jin, X., & Jin, P. (2022). Characteristics of overflow pollution from combined sewer sediment: Formation, contribution and regulation. *Chemosphere*, 298, 134254. <https://doi.org/10.1016/j.chemosphere.2022.134254>
- [4] Perry, W. B., Ahmadian, R., Munday, M., Jones, O., Ormerod, S. J., & Durance, I. (2024). Addressing the challenges of combined sewer overflows. *Environmental Pollution*, 343, 123225. <https://doi.org/10.1016/j.envpol.2023.123225>
- [5] Zan, R., Blackburn, A., Plaimart, J., Acharya, K., Walsh, C., Stirling, R., . . . , & Werner, D. (2023). Environmental DNA clarifies impacts of combined sewer overflows on the bacteriology of an urban river and resulting risks to public health. *Science of the Total Environment*, 889, 164282. <https://doi.org/10.1016/j.scitotenv.2023.164282>
- [6] Yang, J., Zayed, T., Arimiyaw, D., & Xiao, R. (2025). A comprehensive review of influential factors and predictive techniques of time to failure for sewer pipes. *Tunnelling and Underground Space Technology*, 157, 106357. <https://doi.org/10.1016/j.tust.2024.106357>
- [7] Bertrand-Krajewski, J. L., Briat, P., & Scrivener, O. (1993). Sewer sediment production and transport modelling: A literature review. *Journal of Hydraulic Research*, 31(4), 435–460. <https://doi.org/10.1080/00221689309498869>
- [8] Murali, M. K., Hipsey, M. R., Ghadouani, A., & Yuan, Z. (2021). An investigation into the impacts of water demand management and decentralized water recycling on excess sewer sediment deposition. *Journal of Environmental Management*, 279, 111788. <https://doi.org/10.1016/j.jenvman.2020.111788>
- [9] Luo, Y., Bao, S., Yang, S., Zhang, Y., Ping, Y., Lin, C., & Yang, P. (2021). Characterization, spatial variation and management strategy of sewer sediments collected from combined sewer system: A case study in Longgang District, Shenzhen. *International Journal of Environmental Research and Public Health*, 18(14), 7687. <https://doi.org/10.3390/ijerph18147687>

- [10] Li, Z., Wang, B., Wang, F., Sun, B., & Li, L. (2024). Flow dynamics and turbulent coherent structures around sediment reduction plates of a sewer system. *Journal of Environmental Management*, 366, 121594. <https://doi.org/10.1016/j.jenvman.2024.121594>
- [11] Kumar, S., Deshpande, V., Agarwal, M., & Rathnayake, U. (2024). Forecasting particle Froude number in non-deposition scenarios within sewer pipes through hybrid machine learning approaches. *Results in Engineering*, 22, 102320. <https://doi.org/10.1016/j.rineng.2024.102320>
- [12] Sokáč, M., & Velísková, Y. (2021). Impact of sediment layer on longitudinal dispersion in sewer systems. *Water*, 13(22), 3168. <https://doi.org/10.3390/w13223168>
- [13] Liu, L., Kawaike, K., Wada, K., Koshiba, T., Chen, X., Han, C., & Matsui, H. (2025). Environmental drivers of sediment accumulation in urban sewer systems: A case study of Kyoto City, Japan. *Water Research*, 268, 122721. <https://doi.org/10.1016/j.watres.2024.122721>
- [14] Li, Z., Wang, X., Li, W., Liu, Y., Ma, J., & Cao, Z. (2025). Experimental study on sediment incipient motion characteristics in the Yangtze River source region. *Environmental Fluid Mechanics*, 25(5), 42. <https://doi.org/10.1007/s10652-025-10054-0>
- [15] Bong, C. H. J., Liew, S. C., Teo, F. Y., Lau, T. L., & Ab Ghani, A. (2021). Critical shear stress approach for self-cleansing design of a rectangular channel. *International Journal of Sediment Research*, 36(5), 678–685. <https://doi.org/10.1016/j.ijsrc.2021.01.002>
- [16] Safari, M.-J.-S., Aksoy, H., Unal, N. E., & Mohammadi, M. (2017). Non-deposition self-cleansing design criteria for drainage systems. *Journal of Hydro-environment Research*, 14, 76–84. <https://doi.org/10.1016/j.jher.2016.11.002>
- [17] Safari, M. J. S., & Danandeh Mehr, A. (2018). Multigene genetic programming for sediment transport modeling in sewers for conditions of non-deposition with a bed deposit. *International Journal of Sediment Research*, 33(3), 262–270. <https://doi.org/10.1016/j.ijsrc.2018.04.007>
- [18] Tian, R., Zheng, X., Wang, S., Tan, Q., Li, H., & Xie, J. (2025). Prediction of hydrodynamics in a liquid–solid fluidized bed using the densimetric Froude number-based drag model. *Chemical Engineering Science*, 305, 121201. <https://doi.org/10.1016/j.ces.2025.121201>
- [19] Yi, J.-J., Kim, M.-J., Kim, S.-H., Paik, B.-G., & Kim, K. C. (2022). Prediction of supercavitation shapes for a wide range of Froude numbers. *International Journal of Naval Architecture and Ocean Engineering*, 14, 100426. <https://doi.org/10.1016/j.ijnaoe.2021.11.009>
- [20] Song, C.-H., Nguyen, H. H. D., Kim, J.-W., Lee, J.-S., Kim, Y.-T., & Shin, H. (2025). Numerical estimation of debris flow impact pressure on barriers using Froude number-based empirical coefficient. *Engineering Geology*, 355, 108259. <https://doi.org/10.1016/j.enggeo.2025.108259>
- [21] Wan Mohtar, W. H. M., Afan, H., El-Shafie, A., Bong, C. H. J., & Ab. Ghani, A. (2018). Influence of bed deposit in the prediction of incipient sediment motion in sewers using artificial neural networks. *Urban Water Journal*, 15(4), 296–302. <https://doi.org/10.1080/1573062X.2018.1455880>
- [22] Roushangar, K., Ghasempour, R., & Biukaghazadeh, S. (2020). Evaluation of the parameters affecting the roughness coefficient of sewer pipes with rigid and loose boundary conditions via kernel based approaches. *International Journal of Sediment Research*, 35(2), 171–179. <https://doi.org/10.1016/j.ijsrc.2019.08.004>
- [23] Luo, H., Ji, H., Chen, Z., Liu, B., Xue, Z., & Li, Z. (2025). An analytical study for predicting incipient motion velocity of sediments under ice cover. *Scientific Reports*, 15(1), 1912. <https://doi.org/10.1038/s41598-025-85884-5>
- [24] Mahdi, L. A. A.-A., Abdul Wahhab, H. A., & Chaichan, M. T. (2024). Using Froude and Weber numbers to represent the changes in the flow pattern from stratified to stratified-wavy or plug for wire-on-tube condenser. *Results in Engineering*, 24, 103001. <https://doi.org/10.1016/j.rineng.2024.103001>
- [25] Zounemat-Kermani, M., Mahdavi-Meymand, A., & Hinkelmann, R. (2021). Nature-inspired algorithms in sanitary engineering: Modelling sediment transport in sewer pipes. *Soft Computing*, 25(8), 6373–6390. <https://doi.org/10.1007/s00500-021-05628-1>
- [26] Wahl, T. L. (2023). History and physical significance of the roughness Froude number. *Journal of Hydraulic Research*, 61(2), 173–182. <https://doi.org/10.1080/00221686.2022.2161958>
- [27] Moradi, M. A., & Mojra, A. (2024). Free-surface flow past a circular cylinder at high Froude numbers. *Ocean Engineering*, 295, 116804. <https://doi.org/10.1016/j.oceaneng.2024.116804>
- [28] Svantesson, J. L., Ersson, M., & Jönsson, P. G. (2021). Effect of Froude number on submerged gas blowing characteristics. *Materials*, 14(3), 627. <https://doi.org/10.3390/ma14030627>
- [29] Hřebřina, J., Conevski, S., Lia, L., & Pummer, E. (2026). Controls on isolated boulder incipient motion: Experimental insights into Froude number, relative submergence, and shape. *Journal of Hydraulic Engineering*, 152(5), 04026020. <https://doi.org/10.1061/JHEND8.HYENG-14692>
- [30] Ma, Y., Ma, Y., Lyu, J., & Liu, W. (2021). Wet gas pressure drop across multi-orifice plate in horizontal pipe in low gas-phase Froude number region. *Journal of the Taiwan Institute of Chemical Engineers*, 127, 92–101. <https://doi.org/10.1016/j.jtice.2021.08.019>
- [31] Gul, E., Safari, M. J. S., Haghighi, A. T., & Mehr, A. D. (2021). Sediment transport modeling in non-deposition with clean bed condition using different tree-based algorithms. *PLOS One*, 16(10), e0258125. <https://doi.org/10.1371/journal.pone.0258125>
- [32] El-Zaemey, A. K. S. (1991). *Sediment transport over deposited beds in sewers*. [Doctoral dissertation, Newcastle University]. Newcastle University Theses.
- [33] Perrusquía Gonzalez, G. S. (1991). *Bedload transport in storm sewers: Stream traction in pipe channels* (Publication No. 27755802), [Doctoral dissertation, Chalmers Tekniska Högskola]. ProQuest Dissertations & Theses Global
- [34] Ghani, A. A. (1993). *Sediment transport in sewers*. [Doctoral dissertation, Newcastle University]. Newcastle University Theses.
- [35] Montes, C., Kapelan, Z., & Saldarriaga, J. (2021). Predicting non-deposition sediment transport in sewer pipes using random forest. *Water Research*, 189, 116639. <https://doi.org/10.1016/j.watres.2020.116639>
- [36] Nalluri, C., El-Zaemey, A. K., & Chan, H. L. (1997). Sediment transport over fixed deposited beds in sewers—An appraisal of existing models. *Water Science and Technology*, 36(8-9), 123–128. [https://doi.org/10.1016/S0273-1223\(97\)00609-4](https://doi.org/10.1016/S0273-1223(97)00609-4)
- [37] Riazi, M., Khosravi, K., Shahedi, K., Ahmad, S., Jun, C., Bateni, S. M., & Kazakis, N. (2023). Enhancing flood

- susceptibility modeling using multi-temporal SAR images, CHIRPS data, and hybrid machine learning algorithms. *Science of the Total Environment*, 871, 162066. <https://doi.org/10.1016/j.scitotenv.2023.162066>
- [38] Mansour, E. A. (2021). Prognosis of thoracic cancer using the Bierman Random Committee machine learning. *International Journal of Online and Biomedical Engineering*, 17(12), 135–150. <https://doi.org/10.3991/ijoe.v17i12.27573>
- [39] Firouzi, M., Siddiqua, S., Kazemian, H., & Kiamahalleh, M. V. (2025). Green solvent-based extraction of cellulose from hemp bast fibers: From treatment efficacy to characterizations and optimization. *International Journal of Biological Macromolecules*, 288, 138689. <https://doi.org/10.1016/j.jbiomac.2024.138689>
- [40] Pan, T., Song, C., Gao, Z., Xia, T., & Wang, T. (2024). The prediction of flow stress in the hot compression of a Ni-Cr-Mo steel using machine learning algorithms. *Processes*, 12(3), 441. <https://doi.org/10.3390/pr12030441>
- [41] Alam, M., & Husain, S. (2025). Hyperparameter tuned machine learning predictions of specific capacitance of conducting polymers and their composites for high performance advanced supercapacitors. *Applied Physics A*, 131(1), 67. <https://doi.org/10.1007/s00339-024-08137-8>
- [42] Rahman, A., Ali, M. H., Malik, A. W., Mahmood, M. A., & Liou, F. (2025). Physics-based machine learning framework for predicting structure-property relationships in DED-fabricated low-alloy steels. *Metals*, 15(9), 965. <https://doi.org/10.3390/met15090965>
- [43] Al-Refaie, A., Al-atrash, M., & Lepkova, N. (2025). Prediction of the remaining useful life of a milling machine using machine learning. *MethodsX*, 14, 103195. <https://doi.org/10.1016/j.mex.2025.103195>
- [44] Wątróbski, J., Bączkiewicz, A., Król, R., & Rudawska, I. (2024). A neural network regression model supported by multi-criteria methods for ranking prediction in sustainable development assessment. In A. R. da Silva, M. M. da Silva, J. Estima, C. Barry, M. Lang, H. Linger, & C. Schneider (Eds.), *Advances in information systems development* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-57189-3_1
- [45] Tripathi, A., Tiwari, R. K., & Tiwari, S. P. (2022). A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation. *International Journal of Applied Earth Observation and Geoinformation*, 113, 102959. <https://doi.org/10.1016/j.jag.2022.102959>
- [46] Kruse, R., Mostaghim, S., Borgelt, C., Braune, C., & Steinbrecher, M. (2022). Multi-layer perceptrons. In R. Kruse, S. Mostaghim, C. Borgelt, C. Braune, & M. Steinbrecher (Eds.), *Computational intelligence: A methodological introduction* (pp. 53–124). Springer. https://doi.org/10.1007/978-3-030-42227-1_5
- [47] Gad, M. S., Soliman, M. S., & Helal, E. B. (2025). Hybrid XGBoost-RF-MLP model and PSO optimization for performance and emissions of CI engine using waste cooking biodiesel blends. *Scientific Reports*, 15(1), 43832. <https://doi.org/10.1038/s41598-025-29269-8>
- [48] Montes, C., Vanegas, S., Kapelan, Z., Berardi, L., & Saldarriaga, J. (2020). Non-deposition self-cleansing models for large sewer pipes. *Water Science and Technology*, 81(3), 606–621. <https://doi.org/10.2166/wst.2020.154>
- [49] Ali, Y. A., Awwad, E. M., Al-Razgan, M., & Maarouf, A. (2023). Hyperparameter search for machine learning algorithms for optimizing the computational complexity. *Processes*, 11(2), 349. <https://doi.org/10.3390/pr11020349>
- [50] Tiwari, A., Arun, G., & Vishwakarma, B. D. (2021). Parameter importance assessment improves efficacy of machine learning methods for predicting snow avalanche sites in Leh-Manali Highway, India. *Science of the Total Environment*, 794, 148738. <https://doi.org/10.1016/j.scitotenv.2021.148738>
- [51] Khawaja, L., Asif, U., Onyelowe, K., Al Asmari, A. F., Khan, D., Javed, M. F., & Alabduljabbar, H. (2024). Development of machine learning models for forecasting the strength of resilient modulus of subgrade soil: Genetic and artificial neural network approaches. *Scientific Reports*, 14(1), 18244. <https://doi.org/10.1038/s41598-024-69316-4>
- [52] Kumari, N., & Jana, P. K. (2024). A metaheuristic-based task offloading scheme with a trade-off between delay and resource utilization in IoT platform. *Cluster Computing*, 27(4), 4589–4603. <https://doi.org/10.1007/s10586-023-04193-6>
- [53] Mir, A. A., & Patel, M. (2024). A comprehensive review on sediment transport, flow dynamics, and hazards in steep channels. *Journal of Water Management Modeling*, 32, C517. <https://doi.org/10.14796/JWMM.C517>
- [54] Peng, X., Zhu, D. Z., & Zhang, W. (2024). Transport of non-flushable wipes in sewers and its application in sewer management. *Journal of Cleaner Production*, 434, 139876. <https://doi.org/10.1016/j.jclepro.2023.139876>
- [55] Braga, A. S., & Filion, Y. (2022). The interplay of suspended sediment concentration, particle size and fluid velocity on the rapid deposition of suspended iron oxide particles in PVC drinking water pipes. *Water Research X*, 15, 100143. <https://doi.org/10.1016/j.wroa.2022.100143>
- [56] Gogien, F., Dechesne, M., Martinerie, R., & Lipeme Kouyi, G. (2023). Assessing the impact of climate change on combined sewer overflows based on small time step future rainfall timeseries and long-term continuous sewer network modelling. *Water Research*, 230, 119504. <https://doi.org/10.1016/j.watres.2022.119504>
- [57] Wartalska, K., Szymczewski, S., Domalewska, W., Wdowikowski, M., Przestrzelska, K., Kotowski, A., & Kaźmierczak, B. (2025). The impact of climate change on the functioning of drainage systems in industrial areas—A case study. *Atmosphere*, 16(3), 347. <https://doi.org/10.3390/atmos16030347>
- [58] Muleta, T. N., & Knolmar, M. (2025). Ecological impacts of combined sewer overflows on receiving waters. *Discover Water*, 5(1), 24. <https://doi.org/10.1007/s43832-025-00212-2>
- [59] Pozzi, A. C. M., Petit, S., Marjolet, L., Youenou, B., Lagouy, M., Namour, P., ..., & Cournoyer, B. (2024). Ecological assessment of combined sewer overflow management practices through the analysis of benthic and hyporheic sediment bacterial assemblages from an intermittent stream. *Science of the Total Environment*, 907, 167854. <https://doi.org/10.1016/j.scitotenv.2023.167854>

How to Cite: Kumar, S., Agarwal, M., & Rathnayake, U. (2026). Hybrid Machine Learning for Predicting Particle Froude Number in Auto-washout Drainage Systems with Sedimented Beds. *Journal of Data Science and Intelligent Systems*. <https://doi.org/10.47852/bonviewJDSIS62028290>

Appendix

C_{sed}	: volumetric sediment concentration	d_{pipe}	: pipe diameter
d	: sediment size	d_p	: dimensionless sediment size
d_{med}	: the median sediment size	S_g	: sediment relative density
P_{Fr}	: particle Froude number	g	: gravitational acceleration
R_h	: the hydraulic radius	Y	: flow depth
t_{bed}	: deposited bed thickness	V_{mean}	: non-deposition mean flow velocity
ν	: kinematic viscosity	W_{bed}	: deposited bed width
λ_f	: canal/pipe friction factor	ρ_{sed}	: sediment density
ρ_f	: fluid density	t_{sed}	: deposited sedimentation thickness