Using machine learning to predict wall shear stress considering the presence of sediment in wellbores

Pelin Ilker[1*], *Mehmet* Sorgun[1]

[1]Izmır Katip Celebi University, Civil Engineering Department, Havaalanı Sosesi No:33/2 Cigli Izmir, Turkey

Abstract. Sediment transport is a fundamental topic that has been extensively investigated in hydraulic engineering and fluid mechanics. The shear stress acting on the particle surface is a primary variable governing sediment transport in closed conduits. In this study, the parameters influencing the wall shear stress are determined by a dimensional analysis applying the Buckingham Pi Theorem. According to dimensional analysis, the resulting wall shear stress is a function of the Reynolds number, Froude number, sediment concentration, hole inclination, and pipe rotation. The dimensionless wall shear stress coefficient is estimated using five experimentally measured inputs. The difficulty and cost of experimental studies are increasing the importance of computational approaches, including CFD and data-driven methods. Therefore, machine learning captures complex patterns without a fixed formula, provides fast alternatives for quick design checks, and quantifies uncertainty for risk-aware decisions. The features are standardized within a 5-fold cross-validation, and Ridge, Bayesian Ridge, Support Vector Regression (SVR-RBF), XGBoost, and Gaussian Process Regression (GPR) are compared. The results show that XGBoost outperforms the other models with highest accuracy ($R^2 \approx 0.994$, RMSE $\approx 2.3\times10^{-3}$), AAPE≈4.55%). SVR-RBF and GPR followed with ≈18.8% and ≈19.4%, respectively, while linear baselines (Ridge and Bayesian Ridge) exhibited higher errors (≥23%). These results indicate that boosted trees provide the most accurate percentage-error performance.

1 Introduction

Sediment transport in pipelines is a major concern in fluid mechanics directly affecting design parameters, energy usage and pumping power requirements. Since sediment transport is a very complex field, comprehensive studies on this subject are of great importance. In practice, computational methods are widely used to study the governing parameters of solid-liquid flows, considering solid concentration distribution, velocity profiles, and pressure drop, as well as complementary properties such as skin friction coefficient, shear stress distribution and turbulent viscosity. Analyses of asymmetries and residual variability between measured and predicted responses are conducted to compare experimental measurements and numerical predictions for determining the accuracy of the model. In recent years, there have been notable advances in investigation, measurement, and analysis, especially the growing reliability of computational modelling and machine learning approaches for predicting sediment-transport dynamics [1-3].

Over the years, various studies have been conducted on sediment transport and the factors affecting it. In addition to experimental studies, systematic analyses have been carried out using numerical modelling, optimization techniques, and data-driven methods such as machine learning and artificial neural networks. These approaches quantitatively demonstrate the relative effects of factors such as fluid rheology, fluid velocity, sediment concentration, and well geometry on transport efficiency, expanding the possibilities for prediction and optimization in field applications [4,5].

Rooki et al. [6] developed an ANN model that predicts terminal velocity for non-Newtonian power-law fluids. They used a database of 88 sets of Newtonian and non-Newtonian data from the literature and presented an ANN solution that can directly predict the terminal velocity with better accuracy than traditional equations. Li et al. [7] used Artificial Neural Networks (ANN) and Support Vector Regression (SVR) to predict the wall factor for 513 data points obtained from the experimental results of previous studies, considering the effects of particle shape and fluid rheology. To describe particle shape and size, they introduced a characteristic dimension and proposed a comprehensive prediction model applicable to arbitrarily shaped particles. The predicted values were then compared with the experimental results. Goldstein and Coco [8] predicted the settling velocity of noncohesive particles using a genetic-programming-based approach, showing that settling velocity is a nonlinear function of all provided independent variables and that the ML predictor performed satisfactorily. Al-Azani et al. [9] applied ANN and SVM to estimate cuttings concentration, using 116 experimental records with inputs spanning mud density/rheology, hole inclination, pipe eccentricity, rate of penetration, flow rate, rotational speed, and

* Corresponding author: pelin.ilker@ikcu.edu.tr

temperature. Their models outperformed an empirical baseline, indicating feasibility for real-time hole-cleaning assessment. Busch and Johansen [10] investigated the effect of pipe rotation and lateral motion on sediment transport by using computational fluid mechanics and a dynamic mesh technique. They revealed that the two processes of rotation and whirling greatly enhanced sediment transport.

Ozbayoglu et al. [11] developed an ANN-based model using four decades of experimental data from the Tulsa Drilling Research Projects and demonstrated higher accuracy than mechanistic models in most cases; they then proposed optimizing flow rate and pipe rotation with a genetic algorithm to minimize total energy consumption. Shirazi and Frigaard [12] proposed an integrated ANN and SVR model, that predicts both the critical velocity and the pressure drop across different flow regimes for solid-liquid slurry flow in horizontal pipes. The study used three different datasets, including 550 data points. By incorporating slurry flow regime classification into the prediction process, the integrated model demonstrated outstanding performance compared to models without classification. Badrouchi [13] implemented artificial neural network (ANN) models to predict terminal settling velocity while accounting for particle shape and wall effects. These ANN-based predictions were then used to estimate ECD (Equivalent Circulating Density), enabling a fast and accurate assessment of operational-parameter impacts. Purwandari et al. [14] evaluated Huber (robust) regression, Random Forest, and Gradient Boosting to predict cuttings concentration using 332 flow-loop measurements collected from three facilities. The results showed that Gradient Boosting achieved the best performance, with an error of approximately 2.15% and $R2\approx0.92$. Ning et al. [15] employed two data-driven models-an Artificial Neural Network (ANN) and a Least-Squares Support Vector Machine (LSSVM)-to predict drilling-fluid behaviour in nano-SiO_2 water-based muds. They predicted shear stress and filtration volume, achieving $R^2 > 0.99$ and MAE/MAPE < 7% with good agreement to experiments. The relationship between pipe rotation and sediment transport was examined by Zhao et al. [16] The tubular mechanical model and the sediment transport model were modified and combined to create the mechanical pipe-sediment interaction model. They revealed that the interaction model has a high prediction accuracy and performed better than the traditional mechanical and sediment transport models.

Yang et al. [17] showed that while the use of mesh-based CFD in pipe flows with sediment transport is superior in accuracy and efficiency, it requires additional developments in boundary and turbulence modeling when using meshfree particle-based smoothed particle hydrodynamics; integrating the discrete element method (DEM) improves the performance of both approaches. Tipu et al. [18] used a physics-informed neural network (PINN) to predict sediment transport rates, improving accuracy by integrating physical laws governing sediment transport dynamics into the model. When tested against empirical formulas using classical ML models such as RF and SVR, it outperformed with an average R^2 of ≈ 0.9573 and low error values. They concluded that combination of physics constraints and ML provides a more accurate and practical tool for sediment transport. Jing et al. [19] investigated the performance of Artificial Neural Networks (ANN), XGBoost, Random Forest, and Support Vector Regressor in predicting sediment concentration by using a large-scale dataset. The results showed that the machine learning techniques provided reliable predictions for sediment transport processes achieving R^2 values above 0.89 and RMSE within 5.41. Sharma et al. [20] presented a comparative study of machine learning algorithms for industrial-scale slurry transport systems by using 30,000 operational data points pre-processed from 1 million raw data. Among the compared models, Random Forest (RF) exhibited the superior performance with $R^2 = 0.964$ and MAE = 1.27. Their analysis confirmed that flow velocity and mixing density were the most dominant parameters for prediction. Saad et al. [21] used Backpropagation Neural Network (BPNN), Radial Basis Function Network (RBFN), and Support Vector Machine (SVM) to predict cuttings concentration in eccentric deviated wells, using field data from six Gulf of Suez wells. The RBFN model outperformed empirical and fuzzy-logic baselines and maintained high accuracy across neighbouring test wells.

The present study focuses on predicting the dimensionless wall shear stress coefficient C_τ from five measured inputs (Re, Fr, sediment concentration, hole inclination, and pipe rotation). We compare several machine-learning models under the same preprocessing and 5-fold cross-validation. R^2, RMSE, and AAPE are reported, with parity plots used to illustrate errors. A simple, effective model is presented to guide operating-condition selection before expensive tests or computationally expensive simulations.

2 Materials and Methods

2.1 Dimensional Analysis

The experimental data from Sorgun' s study [22], which concerns sediment transport in annular flow (Fig. 1.), is complemented with dimensional analysis. The dimensionless wall shear stress coefficient is defined by:

$$C_\tau = \frac{\tau_w}{1/2\rho v^2} \quad (1)$$

Major variables affecting wall shear stress coefficient are determined as:

$$C_\tau = func(v, \rho, \mu, D_h, \Omega, g, \theta, S_c) \quad (2)$$

where; v is the fluid velocity (L/T), ρ is the density of fluid (M/L^3), μ is the dynamic viscosity of fluid (M/LT), D_h is the hydraulic diameter (L) (D_h=D_o-D_i, D_o is the outer pipe diameter and D_i is the inner pipe diameter), Ω is the inner pipe rotation speed (1/T), g is the gravitational acceleration (L/T^2), θ is the hole inclination and S_c is the sediment concentration.

Dimensional analysis with Buckingham Pi Theorem was performed with the wall shear stress as the dependent variable and the other parameters given as independent variables. The relations created by dimensional analysis are;

$$\pi_1 = \frac{\rho v D}{\mu} \tag{3}$$

$$\pi_2 = \frac{v^2}{gD} \tag{4}$$

$$\pi_3 = C_c \tag{5}$$

$$\pi_4 = \theta \tag{6}$$

$$\pi_5 = \frac{\Omega D_h}{v} \tag{7}$$

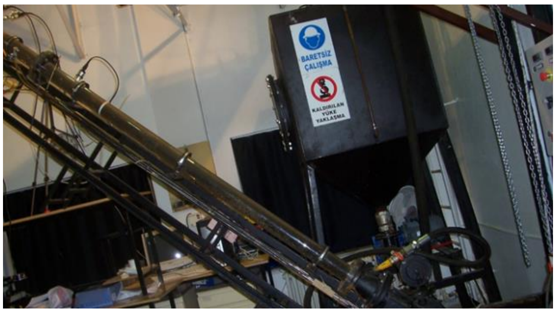

Fig. 1. Experimental Flow-Loop Setup for Sediment Transport (Sorgun,2010).

2.2 Data and Target Definition

In this study, an experimental dataset obtained under controlled conditions on well hydraulics with pipe rotation. Five input data contain dimensionless parameters found as a result of dimensional analysis:

- Re (Reynolds number),
- Fr (Froude number),
- S_C(sediment concentration),
- θ (wellbore inclination),
- $(\Omega D_h)/v$ (pipe rotation effect),

and one dimensionless target: C_τ (wall shear stress coefficient)
where τ_w is wall shear stress, ρ is density, and V is a characteristic velocity. In order to make comparisons with different fluids or different flow rates/velocities at the same scale, normalized shear stress was used as the target variable, aligning with the dimensional analysis presented in Section 2.1.

2.3 Preprocessing and Cross-Validation Scheme

All inputs are standardized to zero mean and unit variance within each training fold to prevent data leakage. Preprocessing and model fitting are carried out using scikit-learn Pipelines, ensuring that identical transformations are applied consistently during cross-validation and inference. Given the dataset size, 5-fold cross-validation is used as the primary evaluation scheme. For each model, R-squared (R^2), root mean squared error (RMSE), and average absolute percentage error (AAPE) are presented; out-of-fold (OOF) predictions are displayed on parity plots. The adopted error metrics:

• **R-squared (R^2):** Proportion of variance in the observations explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

where y_i are observations, $\hat{y}_i$ predictions, $\bar{y}$ the mean of y_i, and n the sample size.

• **Root Mean Square Error (RMSE):** Average magnitude of the prediction errors, expressed in the same units as the target.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

• **Average Absolute Percentage Error (AAPE):** Mean absolute error expressed as a percentage of the observed values.

$$\text{AAPE}(\%) = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{10}$$

2.4 Learning Algorithms

This section describes the five regression models used in this study (Ridge, Bayesian Ridge, SVR-RBF, XGBoost, GPR).

2.4.1 Ridge Regression

Ridge regression is a simple linear model with L2 regularization; it keeps coefficients stable when predictors are correlated and reduces over-variation in the estimates. Inputs are standardized so all features are on a comparable scale [23].

2.4.2 Bayesian Ridge Regression

Bayesian Ridge is a probabilistic version of linear regression. Instead of a single set of coefficients, it learns a distribution over them and then uses the average of that distribution for predictions. This approach usually behaves well when features are correlated and helps avoid unstable fits. Inputs are standardized as with Ridge [24].

2.4.3 Support Vector Regression

SVR fits a function that stays close to the data while ignoring very small errors (the ε-insensitive loss). With the RBF kernel, it can model smooth nonlinear relationships, here between {Re, Fr, conc, hole inclination, pipe rotation} and the target C_τ. Inputs are standardized so all features are on comparable scales. Only a few main settings are used in this study; C (regularization), ε (tube width), and γ (kernel width), and they are kept light and fixed across folds to avoid overfitting [25, 26].

2.4.4 Gaussian Process Regression (GPR) ith RBF kernel

GPR is a nonparametric, Bayesian method that fits a smooth function to the data using a kernel. With an RBF kernel plus a small noise term, it models general trends while accounting for measurement noise. It can also learn one length scale per feature, which hints at which inputs matter more. As with other kernel methods, we standardize inputs first [27].

2.4.5 Extreme Gradient Boosting (XGBoost)

XGBoost builds an ensemble of decision trees step by step, where each new tree focuses on the remaining errors of the previous ones. It captures nonlinear patterns and feature interactions very effectively and works well even when features have different scales. In our data, it delivered the strongest overall accuracy [28].

3 Results and Discussion

For this study, we use experimental annular-flow measurements from the METU-PETE (Middle East Technical University Petroleum Engineering Department) flow loop presented by Sorgun [22]. The data covers a wide range of operating conditions, including hole inclination, pipe rotation, and sediment concentration. For each test, we compute the Reynolds and Froude numbers and formed the dimensionless wall shear stress coefficient $C_\tau = \tau_w/(0.5\,\rho v^2)$. A Summary of the data used in this study is presented in Table 1.

Table 1. A summary of the data used in this study.

Re	Fr	Conc	Hole	Pipe	Target
14789.4	1.479265	0.667337	0.785	0	0.04779525
18075.93	2.209766	0.546003	0.785	0	0.035194684
21910.22	3.246672	0.450453	0.785	0	0.026132053
25744.51	4.482436	0.383364	0.785	0	0.020820468
29578.8	5.917059	0.333669	0.785	0	0.016489361
32865.33	7.305011	0.300302	0.785	0	0.011227104
14789.4	1.479265	1.334675	0.785	0	0.073604684
21910.22	3.246672	0.900906	0.785	0	0.035713805
25744.51	4.482436	0.766728	0.785	0	0.028076085
29578.8	5.917059	0.667337	0.785	0	0.023180696
32865.33	7.305011	0.600604	0.785	0	0.018969935
18075.93	2.209766	1.63801	0.785	0	0.061430721
21910.22	3.246672	1.351358	0.785	0	0.04442449
25744.51	4.482436	1.150092	0.785	0	0.033438933
29578.8	5.917059	1.001006	0.785	0	0.026048411
38342.89	9.942932	0.772205	0.785	0	0.012088297
43820.44	12.98669	0.675679	0.785	0	0.007186315

All analyses are carried out in Python using open-source libraries. We used scikit-learn for Pipelines, preprocessing, cross-validation, and the Ridge/Bayesian Ridge/SVR/GPR estimators; XGBoost (Python package xgboost) for gradient-boosted trees; NumPy/Pandas for data handling; and Matplotlib for parity plots and figures. Evaluation followed a 5-fold K-Fold scheme, and out-of-fold (OOF) predictions were collected for metrics and parity plots. The results are summarized in Table 2.

Table 2. Predictive Accuracy on C_τ: R^2, RMSE, and AAPE (OOF)a reference.

Model	R^2	RMSE	AAPE (%)
XGBoost	0.994408	0.002311	4.55
SVR-RBF	0.947981	0.007049	18.79
GPR	0.928199	0.008281	19.39
Ridge	0.893873	0.010068	23.26
Bayesian Ridge	0.898369	0.009853	24.37

Table 2 shows that XGBoost provides the best predictive performance ($R^2 \approx 0.994$, RMSE ≈ 0.00231, AAPE $\approx$ 4.55%). SVR-RBF and GPR demonstrate acceptable results capturing the overall trend but yield relatively higher percentage error. The linear baselines (Bayesian Ridge, Ridge) show the lowest accuracy.
Figure 2 shows that the XGBoost predictions versus actual values of C_τ. The predicted values are very close to the perfect line over the entire range of C_τ. The scatter is minimal, indicating small errors, conforming that XGBoost achieves the highest predictive accuracy among the tested models.

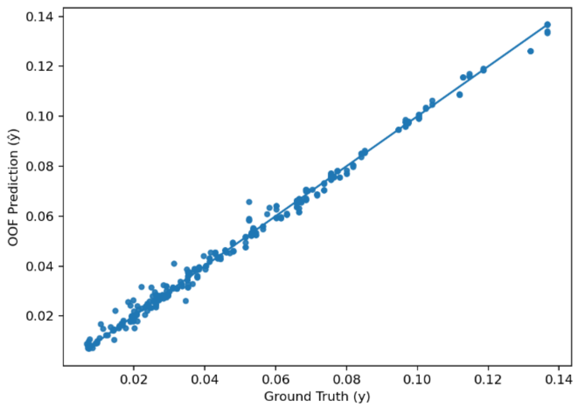

Fig. 2. Predicted vs. Observed C_τ, XGBoost (OOF Parity Plot)

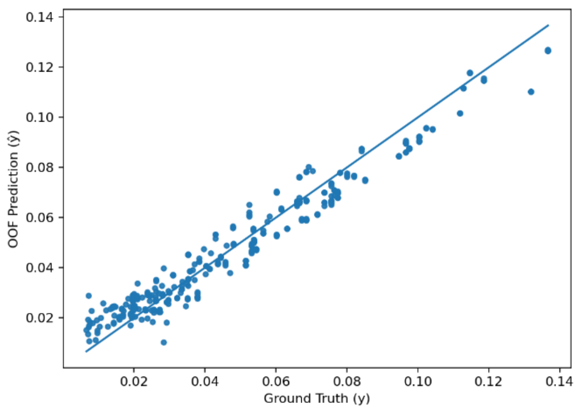

Fig. 3. Predicted vs. Observed C_τ, SVR-RBF (OOF Parity Plot).

However, when the SVR-RBF model is examined (Figure 3), the predictions follow the perfect line but dispersed more than XGBoost. This scatter increases at the low and high ends of C_τ, and a slight underprediction is observed at the upper range.
Similarly, as can be seen from Figure 4, GPR predictions also capture the overall trend but have a spread similar to SVR-RBF, resulting in a relatively higher percentage error.
Lastly, linear baseline models like Bayesian Ridge and Ridge (Figure 5 and Figure 6) show the lowest accuracy. The points in these plots are the most scattered and the furthest from the 1:1 line compared to the other models. This suggests that the relationship between the wall shear stress coefficient and the input variables is not linear, and that non-linear models yield better performance. This is also consistent with prior studies on cuttings transport, where the combined influence of parameters is reported to be complex and non-linear, and non-linear/data-driven models are commonly adopted to capture this behavior [11,12,14,16,18–20].

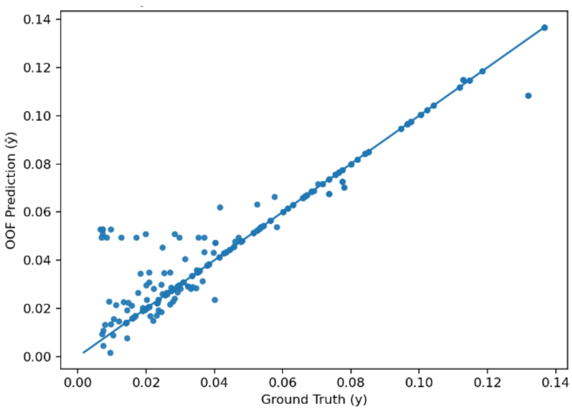

Fig. 4. Predicted vs. Observed C_τ, GPR (OOF Parity Plot).

Lastly, linear baseline models like Bayesian Ridge and Ridge (Figure 5 and Figure 6) show the lowest accuracy. The points in these plots are the most scattered and the furthest from the 1:1 line compared to the other models. This outcome supports the conclusion that the relationship between the wall shear stress coefficient and the input variables is not linear, and that non-linear models yield better performance.

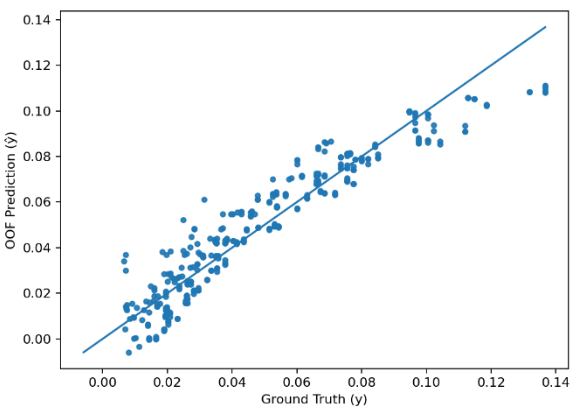

Fig. 5. Predicted vs. Observed C_τ, Bayesian Ridge (OOF Parity Plot).

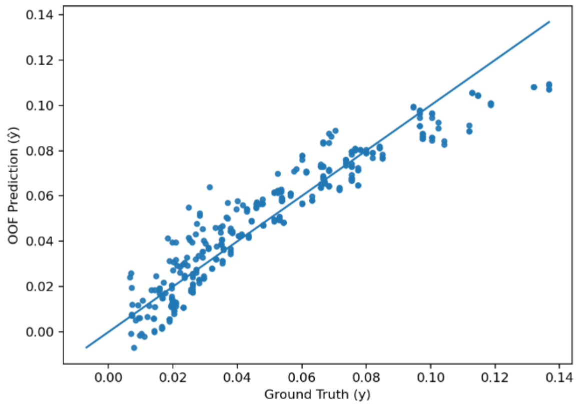

Fig. 6. Predicted vs. Observed C_τ, Ridge (OOF Parity Plot).

4 Conclusion

This study predicted the dimensionless wall shear stress coefficient $C_\tau = \tau_w/(0.5\,\rho v^2)$from five measured inputs which are, Reynolds number, Froude number, sediment concentration, hole inclination, and pipe rotation, by using a consistent Python workflow with 5-fold cross-validation and out-of-fold evaluation. Among the tested models, XGBoost achieved the highest predictive accuracy ($R^2 \approx 0.994$, RMSE ≈ 0.0023, AAPE ≈ 4.6%). SVR-RBF and GPR captured the overall trend but showed higher percentage errors, especially at low target values. Linear baselines (Ridge, Bayesian Ridge) underperformed, which is consistent with a nonlinear relationship between the inputs and C_τ and potential coupled effects among variables, as suggested in prior studies.
The findings practically support the use of ML models for quick assessment of operating conditions in inclined, rotating annuli, reducing the need for repeated high-cost tests.

Acknowledgements: The authors acknowledge the support of TUBITAK under the 2224-A International Scientific Meetings Participation Support Program for participation in this conference (Application No: 1919B022509571).

References

[1] M. Garcia (ed.), Sedimentation engineering: processes, measurements, modeling, and practice (American Society of Civil Engineers, Reston, VA, 2008). https://doi.org/10.1061/9780784408148

[2] Y. Hassanzadeh, Hydraulics of sediment transport. In Hydrodynamics: Theory and Model, 23–58 (2012).

[3] N.V.K. Reddy, J.K. Pothal, R. Barik, P.K. Senapati, Pipeline slurry transportation system: An overview. J. Pipeline Syst. Eng. Pract. **14**, 03123001 (2023). https://doi.org/10.1061/JPSEA2.PSENG-1391

[4] M.N. Belavadi, G.A. Chukwu, Experimental study of the parameters affecting cutting transportation in a vertical wellbore annulus. In SPE Western Regional Meeting (SPE-27880) (1994). https://doi.org/10.2118/27880-MS

[5] M.M. Huque, M.A. Rahman, S. Zendehboudi, S. Butt, S. Imtiaz, Experimental and numerical study of cuttings transport in inclined drilling operations. J. Pet. Sci. Eng. **208**, 109394 (2022). https://doi.org/10.1016/j.petrol.2021.109394

[6] R. Rooki, F.D. Ardejani, A. Moradzadeh, V.C. Kelessidis, M. Nourozi, Prediction of terminal velocity of solid spheres falling through Newtonian and non-Newtonian pseudoplastic power-law fluid using artificial neural network. Int. J. Miner. Process. **110**, 53–61 (2012). https://doi.org/10.1016/j.minpro.2012.03.012

[7] M. Li, G. Zhang, J. Xue, Y. Li, S. Tang, Prediction of the wall factor of arbitrary particle settling through various fluid media in a cylindrical tube using artificial intelligence. Sci. World J. **2014**, 438782 (2014). https://doi.org/10.1155/2014/438782

[8] E.B. Goldstein, G. Coco, A machine learning approach for the prediction of settling velocity. Water Resour. Res. **50**, 3595–3601 (2014). https://doi.org/10.1002/2013WR015116

[9] K. Al-Azani, S. Elkatatny, A. Ali, E. Ramadan, A. Abdulraheem, Cutting concentration prediction in horizontal and deviated wells using artificial intelligence techniques. J. Pet. Explor. Prod. Technol. **9**, 2769–2779 (2019). https://doi.org/10.1007/s13202-019-0672-3

[10] A. Busch, S.T. Johansen, Cuttings transport: On the effect of drill pipe rotation and lateral motion on the cuttings bed. J. Pet. Sci. Eng. **191**, 107136 (2020). https://doi.org/10.1016/j.petrol.2020.107136

[11] E. Ozbayoglu, M. Ozbayoglu, B.G. Ozdilli, Optimization of flow rate and pipe rotation speed considering effective cuttings transport using data-driven models. Energies **14**, 1484 (2021). https://doi.org/10.3390/en14051484

[12] A. Sarraf Shirazi, I. Frigaard, SlurryNet: Predicting critical velocities and frictional pressure drops in oilfield suspension flows. Energies **14**, 1263 (2021). https://doi.org/10.3390/en14051263

[13] F. Badrouchi, Hole cleaning and cuttings transportation modelling and optimization, Ph.D. thesis, University of North Dakota (2021).

[14] S.D. Purwandari, B. Lund, S. Hovda, Machine learning models for predicting cuttings concentration in annulus based on flowloop experimental data. In Int. Conf. Offshore Mech. Arctic Eng. (OMAE 2023) **86915**, V009T11A008 (2023). https://doi.org/10.1115/OMAE2023-104720

[15] Y.C. Ning, S. Ridha, S.U. Ilyas, S. Krishna, I. Dzulkarnain, M. Abdurrahman, Application of machine learning to determine the shear stress and filtration loss properties of nano-based drilling fluid. J. Pet. Explor. Prod. Technol. **13**, 1031–1052 (2023). https://doi.org/10.1007/s13202-022-01589-9

[16] J. Zhao, W. Huang, D. Gao, Interaction between pipe rotation and cuttings transport in extended-reach drilling: mechanism, model, and applications. SPE J. **29**, 2857–2876 (2024). https://doi.org/10.2118/219483-PA

[17] Z. Yang, X. Lian, C. Savari, M. Barigou, Evaluating the effectiveness of CFD-DEM and SPH-DEM for complex pipe flow simulations with and without particles. Chem. Eng. Sci. **288**, 119788 (2024). https://doi.org/10.1016/j.ces.2024.119788

[18] R.K. Tipu, R. Bhakhar, K.S. Pandya, V.R. Panchal, Physics-informed neural networks for predicting sediment transport in pressurized pipe flows. Environ. Earth Sci. **84**, 292 (2025). https://doi.org/10.1007/s12665-025-12295-0

[19] H. Jing, J. Wang, E. Ozbayoglu, S. Baldino, Real-time, data-driven cuttings concentration estimation across a wide range of drilling conditions. In SPE ATCE 2025 (paper D011S005R006) (2025). https://doi.org/10.2118/228095-MS

[20] S. Sharma, B.T. Agyeman, M. Palimi, J. Liu, X. Zhang, Machine learning-driven pressure drop prediction in multiphase slurry hydrotransport. Powder Technol. (article 121314) (2025). https://doi.org/10.1016/j.powtec.2025.121314

[21] M.Y. Saad, A.M. Salem, A.F. Ibrahim, O. Mahmoud, Learning-based prediction of cuttings concentration for enhancing hole cleaning efficiency in eccentric and deviated wells. Sci. Rep. **15**, 32646 (2025).

[22] M. Sorgun, Modeling of Newtonian fluids and cuttings transport analysis in high inclination wellbores with pipe rotation, Ph.D. thesis, Middle East Technical University, Turkey (2010).

[23] A.E. Hoerl, R.W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems. Technometrics **12**, 55–67 (1970).

[24] D.J. MacKay, Bayesian interpolation. Neural Comput. 4, 415–447 (1992). https://doi.org/10.1162/neco.1992.4.3.415

[25] V. Vapnik, The Nature of Statistical Learning Theory (Springer, New York, 2013).

[26] A.J. Smola, B. Schölkopf, A tutorial on support vector regression. Stat. Comput. **14**, 199–222 (2004).

[27] C.K.I. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning (MIT Press, Cambridge, MA, 2006).

[28] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system. In Proc. KDD ’**16**, 785–794 (2016). https://doi.org/10.1145/2939672.2939785