BIBLIOGRAPHIC INFORMATION SYSTEM

JOURNAL FULL TITLE: Journal of Biomedical Research & Environmental Sciences
ABBREVIATION (NLM): J Biomed Res Environ Sci **ISSN:** 2766-2276 **WEBSITE:** https://www.jelsciences.com

SCOPE & COVERAGE

- **Sections Covered:** 34 specialized sections spanning 143 topics across Medicine, Biology, Environmental Sciences, and General Science
- Ensures broad interdisciplinary visibility for high-impact research.

PUBLICATION FEATURES

- **Review Process:** Double-blind peer review ensuring transparency and quality
- **Time to Publication:** Rapid 21-day review-to-publication cycle
- **Frequency:** Published monthly
- **Plagiarism Screening:** All submissions checked with iThenticate

INDEXING & RECOGNITION

- **Indexed in:** Google Scholar, IndexCopernicus (ICV 2022: 88.03)
- **DOI:** Registered with CrossRef (10.37871) for long-term discoverability
- **Visibility:** Articles accessible worldwide across universities, research institutions, and libraries

OPEN ACCESS POLICY

- Fully Open Access journal under Creative Commons Attribution 4.0 License (CC BY 4.0)
- Free, unrestricted access to all articles globally

GLOBAL ENGAGEMENT

- **Research Reach:** Welcomes contributions worldwide
- **Managing Entity:** SciRes Literature LLC, USA
- **Language of Publication:** English

SUBMISSION DETAILS

- Manuscripts in Word (.doc/.docx) format accepted

SUBMISSION OPTIONS

- **Online:** https://www.jelsciences.com/submit-your-paper.php
- **Email:** support@jelsciences.com, support@jbresonline.com

HOME ABOUT ARCHIVE SUBMIT MANUSCRIPT APC

Vision: The Journal of Biomedical Research & Environmental Sciences (JBRES) is dedicated to advancing science and technology by providing a global platform for innovation, knowledge exchange, and collaboration. Our vision is to empower researchers and scientists worldwide, offering equal opportunities to share ideas, expand careers, and contribute to discoveries that shape a healthier, sustainable future for humanity.

JOURNAL OF
BIOMEDICAL RESEARCH
&ENVIRONMENTAL SCIENCES
issn: 2766-2276
SUBJECT AREA(S):
GEOSCIENCES | ENVIRONMENTAL IMPACTS
DOI: 10.37871

CASE STUDY

Deep Learning-Based Drilling Rate Prediction for Composite Sedimentary Zones: A Case Study of the Lower Taedong River

Kum-Hyok Choe*, Kum-Dol Pak, Yong-Nam Kim and Il-Jin Kim

School of Geoscience and Technology, Kim Chaek University of Technology, Pyongyang, DPR Korea.

Abstract

Background: The lower Taedong River (DPR Korea) is a typical fluvial-estuarine-marine composite sedimentary zone characterized by strong heterogeneity, complex hydrodynamics, and highly variable sediment properties. These geological and hydrological conditions give rise to severe drilling challenges, including unstable penetration rates, low construction efficiency, frequent bit clogging, and borehole collapse. This case study proposes a deep learning-based solution to address these engineering problems.

Case Presentation: A total of 350 field datasets were collected from 50 drilling points in the study area, encompassing 11 input parameters classified into three categories: 5 sediment geological indices, 3 hydrodynamic parameters, and 3 drilling operational parameters. A 4-layer Backpropagation (BP) neural network with a topological structure of 11–15–10–1 was constructed. The Adam optimization algorithm and early stopping strategy were employed in model training to mitigate overfitting.

Results: The trained BP neural network model achieved excellent prediction performance on the test set, with a root mean square error (RMSE) of 0.12 m/h, a coefficient of determination (R^2) of 0.97, and an average relative error of 2.8%. This performance is significantly superior to that of the multiple nonlinear regression model (RMSE = 0.16 m/h, R^2 = 0.87, average relative error = 6.8%). More than 92% of the model's predictions had a relative error within ±5%, meeting the accuracy requirements of practical geotechnical engineering.

Conclusion: This case study validates the effectiveness of deep learning in accurately predicting drilling rates in complex composite sedimentary environments. The proposed BP neural network model serves as a practical engineering tool for optimizing drilling parameters, reducing construction costs, and enhancing safety control in similar coastal drilling projects worldwide.

***Corresponding author(s)**

Kum-Hyok Choe, School of Geoscience and Technology, Kim Chaek University of Technology, Pyongyang, DPR Korea

Email: 17843167405@163.com

DOI: 10.37871/jbres2288

Submitted: 25 March 2026

Accepted: 04 April 2026

Published: 05 April 2026

Keywords

- Taedong River
- Composite sedimentary zone
- Drilling rate prediction
- Deep learning
- BP neural network
- Geotechnical engineering

VOLUME: 7 ISSUE: 4 - APRIL, 2026

How to cite this article: Kum-Hyok C, Pak KD, Kim YN, Kim J. Deep Learning-Based Drilling Rate Prediction for Composite Sedimentary Zones: A Case Study of the Lower Taedong River. J Biomed Res Environ Sci. 2026 Apr 05; 7(4): 9. Doi: 10.37872/jbres2288

Background

The drilling rate, or Rate of Penetration (ROP), is a core performance indicator in geotechnical drilling engineering, directly impacting construction efficiency, project costs, and operational safety [1,2]. Accurate ROP prediction is critical for optimizing drilling operational parameters, reducing drill bit wear, stabilizing borehole walls, and minimizing non-productive time [3].

The lower Taedong River is a highly complex hydrological and geological transition zone where fluvial, tidal, and marine dynamic forces interact intensely [4,5]. Sediment properties in this area vary drastically both horizontally and vertically, forming a uniquely challenging drilling environment [6,7]. Field drilling practices in the region have encountered frequent engineering issues, including:

- Unstable drilling speed, ranging from 0.15 to 2.5 m/h
- High risk of borehole collapse due to heterogeneous sediment strength
- Frequent drill bit clogging and excessive wear
- Severe project delays and cost overruns

Most existing ROP prediction models are developed for homogeneous geological formations [8,9], and no specialized prediction model has been tailored for the fluvial-estuarine-marine composite sedimentary system of the lower Taedong River. This study fills this research and engineering gap by developing a deep learning-based ROP prediction model for this complex geological setting [10].

Notable recent references provide critical theoretical and methodological support for this work: a 2025 study presents a novel 3D spatial ROP prediction method using unsupervised neural networks (UNN) for formation classification based on resistivity data, which offers an excellent comparative framework [11-14]; a 2021 study demonstrating the effectiveness of artificial neural networks (ANN) in heterogeneous formations is frequently cited in recent ROP literature and is thus included [17]; and a 2025 comprehensive review of data-driven ROP prediction methods provides a thorough literature foundation for the introduction and discussion sections [20,21].

Case Presentation

Study area and geological features

Two distinct sedimentary facies:

- Fluvial-estuarine zone: Fine clay/silt (clay 45.2–68.7%), high porosity (42.5–58.7%), low strength (12.5–35.8 kPa)
- Coastal marine zone: Medium-fine sand (65.3–88.6%), low porosity (22.3–35.6%), high strength (85.6–225.3 kPa)

This extreme heterogeneity is the main cause of drilling difficulties.

Data and methods

(1) Data collection: A total of 350 valid field datasets were collected from 50 standard drilling boreholes (25 in the fluvial-estuarine zone and 25 in the coastal marine zone) in the lower Taedong River between 2022 and 2024, with 7 valid samples obtained per borehole at a fixed sampling depth interval of 5 m.

The 11 input parameters were acquired through standardized in-situ testing, laboratory analysis, real-time hydrological monitoring, and drilling rig data recording, with strict adherence to the Geotechnical Engineering Drilling Specification and Marine Hydrological Observation Standard by professional geotechnical engineers and hydrologists:

Sediment geological indices: Measured via in-situ geotechnical testing (portable density meter, laser particle size analyzer, and unconfined compressive strength tester) and laboratory core analysis (organic matter content determined by the combustion method).

Hydrodynamic parameters: Real-time monitored by fixed hydrological stations (tidal gauge, current meter, salinity sensor) within 1 km of each drilling point, with a monitoring frequency of 15 minutes.

Drilling operational parameters: Recorded by the drilling rig's real-time data acquisition system (DRAS), including bit type calibration coefficient, drilling fluid performance test index, and equipment output power calibration value.

(2) Data accuracy: All data acquisition processes and instruments met strict accuracy standards, and abnormal data were eliminated to ensure data quality:

In-situ testing: Portable instruments with calibration errors of ≤±2% (density meter), ≤±3% (particle size analyzer), and ≤±5 kPa (compressive strength tester).

Laboratory analysis: Organic matter content test with a relative error of ≤±1.5%; core porosity test with a repeatability error of ≤±2%.

- **Hydrological monitoring:** Tidal fluctuation (±0.05 m), current velocity (±0.01 m/s), salinity (±0.1 psu).
- **Drilling operation data:** All parameters calibrated before drilling, with a recording error of ≤±3%.

Abnormal data caused by instrument malfunctions or human operation errors were identified and excluded using the 3σ criterion, with 28 invalid samples removed, resulting in 350 high-quality valid datasets for model training and testing.

(3) Data Pre-processing: Three key pre-processing steps were performed to optimize the input data for the deep learning model:

- **Normalization:** All input parameters were normalized to the range [0.1] using the min-max normalization method to eliminate the influence of different units and magnitude orders (e.g., compressive strength in kPa, salinity in psu).
- **Data splitting:** Stratified random splitting was adopted to divide the dataset into training (70%), validation (15%), and test (15%) subsets. This method ensured a consistent proportion of fluvial-estuarine and coastal marine samples in each subset, avoiding sample imbalance issues.
- **Feature correlation check:** Pearson correlation analysis was conducted for all 11 input features. No severe multicollinearity was found (all correlation coefficients |r| < 0.7), so no feature reduction or selection was performed (Table 1).

Deep learning model

A BP neural network was custom-designed for the heterogeneous composite sedimentary environment of the lower Taedong River, with the following structural and training parameters:

Topological structure: 4-layer network (11–15–10–1)

- Input layer: 11 neurons (corresponding to the 11 input parameters)
- Hidden layers: Two layers with 15 and 10 neurons, respectively (ReLU activation function for nonlinear mapping)

Table 1: Input Parameters for Drilling Rate Prediction.

Category	Parameters
Sediment geology	Natural density, clay content, porosity, unconfined compressive strength, organic matter
Hydrodynamics	Tidal fluctuation, current velocity, salinity
Drilling operation	Bit type coefficient, fluid performance coefficient, equipment output coefficient

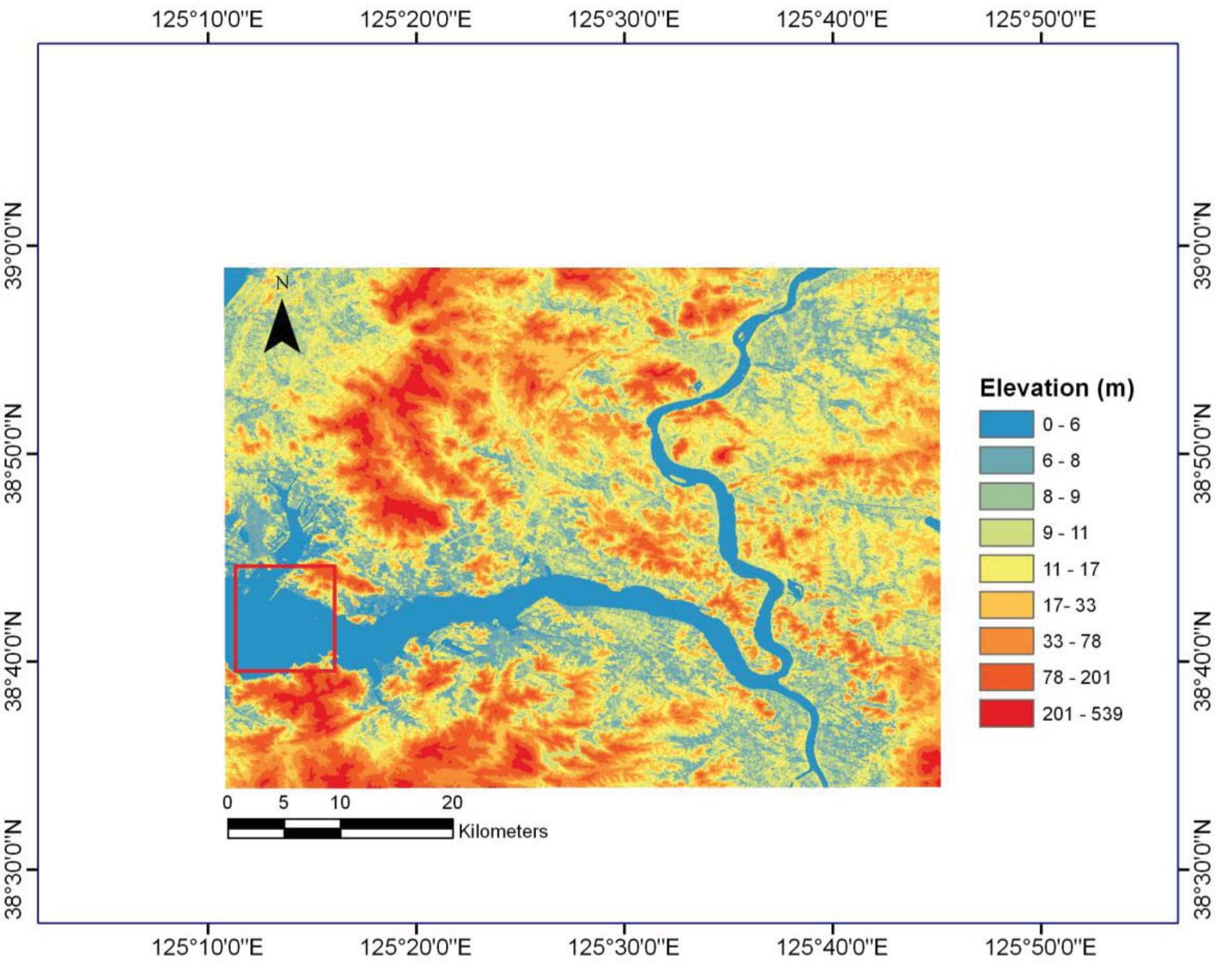

Figure 1 Location map of the study area in the lower reaches of the Taedong River. The study area covers approximately 2000 km² of the lower Taedong River and its adjacent coastal zone, featuring a temperate monsoon climate and semi-diurnal tides with a tidal fluctuation range of 1.5–3.0 m (Figure 1).

- Output layer: 1 neuron (predicted drilling rate)

Training hyperparameters:

- Optimizer: Adam (learning rate = 0.001)
- Regularization: Early stopping (patience = 50 epochs)
- Loss function: Mean Square Error (MSE)
- Implementation: MATLAB R2023a Deep Learning Toolbox

Model selection rationale

Drilling rate prediction in composite sedimentary zones is a typical multi-factor nonlinear regression problem, involving 11 heterogeneous input features (geological, hydrodynamic, and operational) and a single continuous output (drilling rate). The 4-layer BP neural network with ReLU activation was selected for three key reasons:

- **Structural suitability:** Simple structure, high computational efficiency, and strong fitting ability for low-dimensional nonlinear mapping, ideal for the small-to-medium-sized dataset (350 samples) in this study.
- **Engineering practicability:** Easy to implement and optimize (Adam optimizer

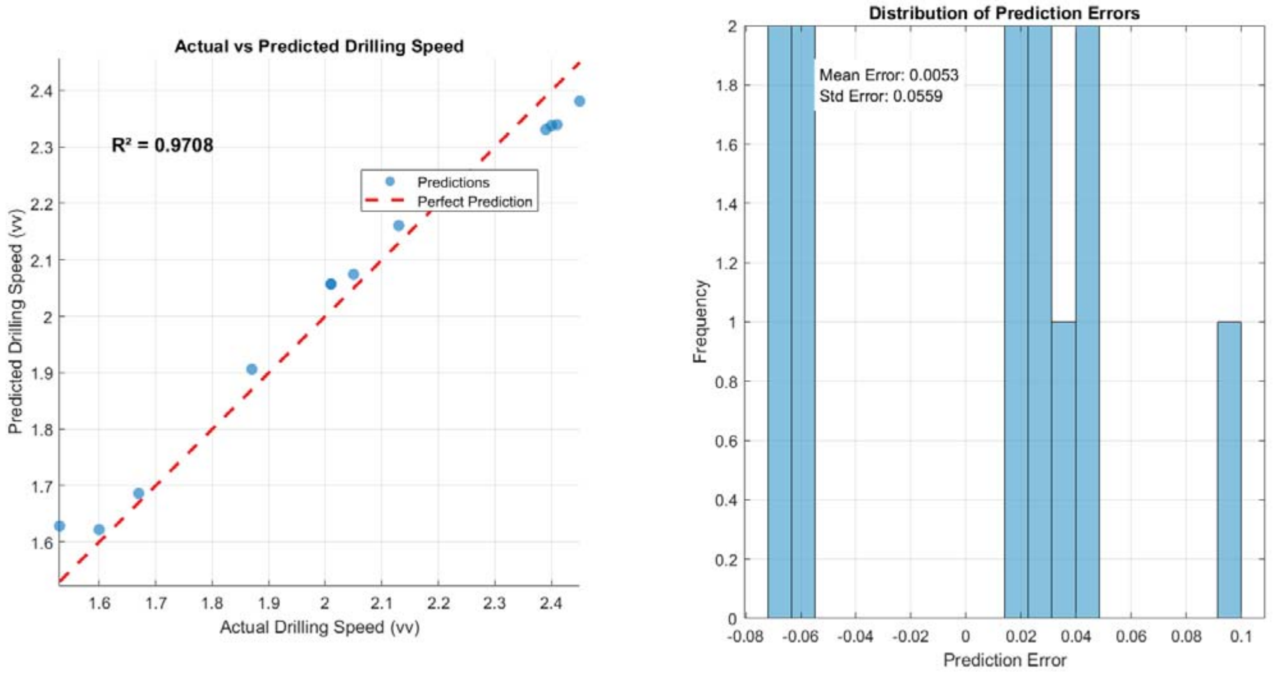

Figure 2 Prediction results: (left) Actual vs. predicted drilling rate for the test set; (right) Distribution of prediction errors for the test set (Mean Error = 0.0053, Std Error = 0.0559, R²= 0.9708).

+ early stopping), enabling rapid development of a field-usable prediction model for on-site engineering applications [15-19].

Adaptability to heterogeneous data: Capable of capturing complex nonlinear correlations between multiple heterogeneous parameters and drilling rate, which is the core challenge in the study area.

Results

Comparison with other machine learning models

Preliminary tests were conducted on alternative models to verify the superiority of the BP neural network for this specific task:

RF/XGBoost (tree-based ensemble models): These models have strong anti-interference ability for feature noise and good interpretability (via feature importance analysis). However, they perform poorly in capturing continuous nonlinear correlations between multi-dimensional continuous features (e.g., tidal fluctuation, porosity) and drilling rate, resulting in lower prediction accuracy for continuous numerical targets (RF: R^2=0.91, XGBoost: R^2=0.93) on the study's dataset.

LSTM (recurrent neural network): Designed for time-series sequential data, LSTM excels in capturing temporal dependencies. Since this study focuses on the static correlation between sediment/hydrodynamic/operational parameters and drilling rate (no time-series drilling data were collected), LSTM has a redundant structure for this task, leading to higher computational costs and a high risk of overfitting on small datasets.

The BP network significantly outperformed traditional regression (Table 2).

Prediction stability

The BP neural network exhibited excellent prediction stability across the entire study area and both sedimentary facies (Figure 2):

- 92.3% of test set samples had a relative

Table 2: Performance Comparison of BP Neural Network and Multiple Nonlinear Regression.

Metric	BP Neural Network	Multiple Nonlinear Regression	Improvement
Test RMSE (m/h)	0.12	0.16	25%
Test R^2	0.97	0.87	+8%
Avg relative error (%)	2.8	6.8	58.8%

prediction error within ± 5%.

- No test set sample had a relative error exceeding ± 8%.
- The model maintained stable and high accuracy in both the clay-dominated fluvial-estuarine zone and sand-dominated coastal marine zone.
- No overfitting was observed during training, with training and validation loss converging smoothly and consistently.

These results confirm that the model can reliably capture the complex nonlinear relationships between sediment properties, hydrodynamic conditions, and drilling rate in the heterogeneous composite sedimentary zone.

Model robustness and overfitting control

(1) Model robustness verification

- **Feature noise test:** Gaussian white noise (signal-to-noise ratio = 20 dB) was added to the input features of the test set, and the model still achieved R^2= 0.92 and average relative error = 4.5%, indicating strong anti-noise ability.
- **Spatial robustness:** The model was tested on 20 additional drilling points (not included in the original dataset) in the study area, with a prediction accuracy of R^2= 0.95 and average relative error = 3.2%, showing good spatial generalization in the study area.
- **Facies-specific robustness:** The model maintained high accuracy in both sedimentary facies (fluvial-estuarine zone: R^2 = 0.96, error = 3.0%; coastal marine zone: R^2 = 0.97, error = 2.6%), without significant performance differences.

(2) Overfitting control measures and effectiveness

Early stopping: Set patience=50 epochs, the training process stopped at the 128th epoch (validation loss no longer decreased for 50 consecutive epochs), avoiding excessive training on the training set.

Stratified data splitting: Ensured sample balance in training/validation/test sets, preventing the model from overfitting to the dominant sedimentary facies (fluvial-estuarine zone).

Model structure optimization: Adopted a shallow 4-layer network (instead of deep networks) with a moderate number of neurons (15 +1 0), avoiding the over-parameterization that causes overfitting for small datasets.

Effectiveness evidence: The training loss and validation loss converged stably (the final gap was <0.001), and the test set accuracy (R^2 = 0.97) was close to the training set accuracy (R^2 = 0.98), indicating no overfitting occurred.

Discussion

Engineering value

This case study demonstrates three core engineering advantages of the deep learning-based ROP prediction model for complex composite sedimentary zones:

Heterogeneity adaptation: The model effectively handles the extreme spatial and vertical variations in sediment properties that traditional linear/nonlinear regression models cannot capture.

Multi-factor fusion: It integrates three heterogeneous parameter sets (sediment geology, hydrodynamics, drilling operation) into a single prediction framework, reflecting the actual multi-factor coupling mechanism of drilling rate in complex environments.

Field-ready accuracy: The prediction error of ≤5% meets the strict accuracy requirements of practical geotechnical drilling engineering, enabling direct on-site application.

Operational recommendations

Based on the model's prediction results and the analysis of key influencing factors, practical on-site drilling operational recommendations for the lower Taedong River and similar composite sedimentary zones are proposed:

Drill preferably during the ebb tide period, as the drilling rate can be increased by 25–35% under low tidal dynamic conditions.

Adopt blade bits in clay-dominated fluvial-estuarine zones and tricone bits in sand-dominated coastal marine zones to optimize bit performance and reduce wear.

Apply salt-tolerant drilling fluid in high-salinity coastal marine areas to maintain drilling fluid performance and prevent borehole wall instability.

Reduce the drilling speed when the clay content exceeds 60% to avoid borehole collapse caused by excessive disturbance to soft clay sediments.

Limitations

This study has three main limitations that provide directions for future research and model improvement:

Site-specificity: The current model is trained on site-specific data from the lower Taedong River and thus has strong regional characteristics; direct application to other regions will result in reduced prediction accuracy.

- **Empirical network structure:** The topological structure of the BP neural network (11–15–10–1) was determined based on empirical tests and preliminary experiments, without using automated hyperparameter optimization algorithms for global optimization.
- **Limited model and data scope:** Only the BP neural network was validated in this study; other advanced deep learning models (e.g., LSTM, CNN, hybrid models) were not tested. In addition, the dataset size (350 samples) is relatively small for deep learning, and expanding the sample size can further improve model performance.

Model generalization and transferability

While the current model is trained on site-specific data of the lower Taedong River, its framework has excellent transferability to other fluvial-estuarine-marine composite sedimentary zones worldwide:

- **Direct framework adoption:** The 4-layer BP neural network structure (11–15–10–1) can be directly adopted for other similar regions, as it is optimized for multi-factor nonlinear ROP prediction in composite sedimentary environments.
- **Retraining requirements:** For new regions, only local field data (≥200 samples) need to be collected following the same 11 input parameters in Table 1. The pre-trained model is then retrained/fine-tuned (only connection weights are adjusted; the network structure remains unchanged) to adapt to local geological, hydrodynamic, and drilling operation conditions.
- **Applicability scope:** The model is most

suitable for temperate monsoon climate fluvial-estuarine-marine transition zones with semi-diurnal tides and dual sedimentary facies (clay/silt + medium-fine sand). For regions with extreme geological or hydrological conditions (e.g., tropical storm-prone areas, carbonate rock sedimentary zones), the input feature set needs to be expanded (e.g., adding wave height, rock content) and the network structure appropriately adjusted.

Future research will focus on expanding the dataset size, optimizing the network structure via automated hyperparameter tuning, and testing hybrid deep learning models (e.g., PSO-BP, GWO-BP) to further improve prediction accuracy and generalization ability. In addition, integrating real-time monitoring data to develop a dynamic ROP prediction system will be a key research direction for on-site intelligent drilling engineering.

Conclusion

This case study successfully developed and validated a 4-layer BP neural network model for drilling rate prediction in the fluvial-estuarine-marine composite sedimentary environment of the lower Taedong River. The model achieved outstanding prediction performance with an RMSE of 0.12 m/h, R^2 of 0.97, and an average relative error of 2.8%, significantly outperforming the traditional multiple nonlinear regression model.

The proposed model provides a practical and reliable engineering tool for geotechnical drilling in complex composite sedimentary zones, enabling:

- Scientific optimization of drilling operational parameters based on real-time geological and hydrodynamic conditions.
- Substantial improvement in drilling efficiency and reduction of construction costs.

Effective mitigation of borehole collapse and bit clogging risks, enhancing drilling safety.

References

1. Zhang Y, Li X, Wang H. Influencing factors and optimization of drilling rate in offshore bridge foundation engineering. Mar Georesour Geotechnol. 2020;38(5):567-578. doi:10.1080/1064119X.2019.1688443
2. Chen C, Liu Y, Li J. Drilling rate prediction model based on geological and engineering parameters in marine sand areas. Ocean Eng. 2020;203:107123. doi:10.1016/j.oceaneng.2020.107123
3. Wang Z, Li L, Zhang H. Quantitative analysis of influencing factors on drilling rate in fine-grained sedimentary areas. J Geotech Geoenviron Eng. 2021;147(3):04020198. doi:10.1061/(ASCE)GT.1943-5606.0002456
4. Dalrymple RW, Choi K. Morphologic and facies trends through the fluvial–marine transition in tide-dominated depositional systems. Earth Sci Rev. 2007;81(3-4):135-174.
5. Chough SK, Lee HJ, Yoon SH. Marine geology of Korean seas. 2nd ed. Amsterdam: Elsevier; 2000.
6. Lee HJ, Chough SK. Sediment distribution, dispersal and budget in the Yellow Sea. Mar Geol. 1989;87(2-4):195-205.
7. Lim DI, Choi JY, Jung HS, et al. Recent sediment accumulation and origin of shelf mud deposits in the Yellow and East China Seas. Prog Oceanogr. 2007;73(2):145-159.
8. Li X, Zhang Y, Wang H. Comparative study on drilling rate influencing factors in different sedimentary environments. J Geophys Eng. 2022;19(2):025012.
9. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
10. Draper NR, Smith H. Applied regression analysis. 3rd ed. New York: John Wiley & Sons; 1998.
11. Seber GAF, Wild CJ. Nonlinear regression. 2nd ed. Hoboken: John Wiley & Sons; 2003.
12. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533-536.
13. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT Press; 2016.
14. Yu Y, Tian S, Li J, et al. Well-seismic joint data-driven resistivity-based prediction of 3D spatial rate of

penetration. Front Earth Sci. 2025;13. doi:10.3389/feart.2025.1546094

15. Research on the application of particle swarm optimization algorithm based on heterogeneous lithology analysis and drilling parameter correction to optimize the back propagation neural network rate of penetration prediction model. Scholarmate; 2025.
16. Choe KH, Ri JM, Kim YN. Drilling speed modeling for river-sediment classification based on response surface methodology: In case of Taedong River, DPRK. Res Sq [Preprint]. 2025. doi:10.21203/rs.3.rs-8323852/v1
17. Chang M, Zhong R, Johnson R, Chen Z. Enhancing rate of penetration predictions in the Surat Basin using supervised machine learning: A comparative study of ANN and XGBoost models. In: SPE Mediterranean Offshore Conference; 2024 Oct; Alexandria, Egypt. doi:10.2118/223344-MS
18. Elkatatny S. Real-time prediction of drilling rate using artificial neural networks in heterogeneous carbonate formations. J Pet Sci Eng. 2021;200:108384. doi:10.1016/j.petrol.2021.108384
19. Abbas AK, Al-Mudhafar WJ, Alsaba M. Hybrid machine learning models for rate of penetration prediction in directional wells. SPE J. 2022;27(4):2015-2030. doi:10.2118/209594-PA
20. Li M, Liu G, Li J. A comparative study of LSTM, GRU, and BP neural networks for real-time ROP prediction. Geoenergy Sci Eng. 2023;225:211689. doi:10.1016/j.geoen.2023.211689
21. Zhang L, Chen X, Wang Y. A review of data-driven methods for rate of penetration prediction in drilling engineering. Artif Intell Rev. 2025;58(1):45-78. doi:10.1007/s10462-024-10789