-RESEARCH PAPER-

International Journal of Mining and Geo-Engineering IJMGE

IJMGE 60-3 (2026) 313-323 DOI: 10.22059/IJMGE.2026.400649.595291

Machine learning-based methods to predict the tunnel water inflow: insights from the water conveyance tunnel database

Sasan Ghorbani [a,*] and Ali Bameri [b]

[a] *Department of Mining Engineering, Hamedan University of Technology, Hamedan, Iran.*
[b] *Department of Mining Engineering, Isfahan University of Technology, Isfahan, Iran.*

Article History:
Received: 13 August 2025.
Revised: 02 February 2026.
Accepted: 08 September 2026.

ABSTRACT

Water inflow (WI) into the tunnel is one of the main geological hazards that can have a significant negative impact on the progress of the tunneling project. A precise prediction of the water inflow into the tunnel, as a significant challenge in rock tunneling, can guarantee project safety and progress. To address this, the current study applied five machine learning (ML) algorithms, including Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and Natural Gradient Boosting (NGBoost). Therefore, a dataset including hydraulic conductivity (K), geological strength index (GSI), and water head (H) from 73 sections across six water conveyance tunnels in Iran was collected. The ML algorithms were implemented and then three performance metrics, including the coefficient of determination (R^2), normalized root mean square error (NRMSE), and the variance account for (VAF) were employed to evaluate the efficiency of the ML models. As a result, the XGBoost model for testing data showed the greatest level of accuracy and reliability in predicting WI with R^2, NRMSE, and VAF of 84.9%, 12.9%, and 82%, respectively. Also, based on the results of score analysis and regression error characteristic curve (REC), XGBoost was suggested as the best method for predicting WI in the tunnel. Finally, the water head was found to be the most effective parameter in predicting WI using the Shapley Additive exPlanations (SHAP) and Partial Dependence Plot (PDP) methods.

Keywords: *Water inflow, XGBoost, PDP, Water conveyance tunnel, SHAP.*

1. Introduction

Water inflow (WI) into the tunnel is a major problem when tunneling in regions with complicated hydrogeological and geotechnical conditions. Water inflow into the tunnel is recognized as one of the most crucial geological hazards, has a substantial impact on workforce safety and project progress [1–3].

Creating an unsuitable working environment, tunnel collapse, and the complete burial of the TBM and drilling equipment are problems directly related to the amount of water ingress. These hazards reduce the speed of TBM excavation, increase costs, and increase construction time [4, 5]. Therefore, one of the main reasons for addressing this issue is its extraordinary importance, which can cause numerous problems during the implementation of tunneling projects, both from an economic and technical perspective [6].

Accurate prediction of water inflow in tunnel engineering is a critical component of hydro-mechanical risk assessment, as groundwater ingress directly influences effective stress conditions, face stability, lining performance, and long-term serviceability of underground structures. Excessive inflow can trigger hydraulic fracturing, internal erosion, piping, and progressive weakening of surrounding rock masses, particularly in highly fractured or karstified formations. Moreover, sudden high inflow events may lead to blowouts, equipment damage, construction delays, and substantial cost overruns. From a design perspective, reliable WI prediction is essential for optimizing pre-excavation grouting schemes, drainage capacity, waterproofing systems, and support pressure in mechanized tunneling (e.g. TBM operations), thereby ensuring both structural integrity and operational safety under varying hydrogeological conditions.

Given the water issue and the increasing construction of water conveyance tunnels worldwide, it is very important to recognize this hazard and be prepared to deal with it. Hence, by predicting the potential for water inrush along tunnels during project construction, one can be prepared to control the water inrush with operational methods such as grout injection [7–11]. Predicting and controlling WI before excavation is therefore essential to guaranteeing project viability, construction safety, and operational effectiveness [1, 12].

There are various methods for predicting WI in the tunnels. These methods include empirical methods, analytical methods, numerical methods, and recently machine learning (ML) algorithms. Machine-learning approaches provide a robust framework for WI prediction because groundwater inflow is governed by complex, nonlinear interactions among geological, structural, and hydraulic parameters, including permeability, fracture density, overburden depth, in-situ stress, and groundwater head. Conventional analytical or empirical models often rely on simplifying assumptions (e.g. homogeneity, isotropy, steady-state flow) that may not adequately represent real subsurface conditions. In contrast, ML algorithms can capture high-order nonlinearities and coupled effects directly from observational data, enabling improved generalization across heterogeneous tunnel sections. By integrating multi-parameter datasets and learning implicit patterns, ML-based WI prediction enhances probabilistic risk

* Corresponding author. *E-mail address:* s.ghorbani@hut.ac.ir (S. Ghorbani).

evaluation, supports adaptive construction management, and facilitates data-driven decision-making for safer and more economically efficient underground excavation projects.

ML algorithms have been applied in the last few years to predict WI in tunnels and have demonstrated great potential. For example, Li et al. [13] applied the nonlinear regression Gaussian process analysis to develop a new model for predicting WI in the tunnels. Their results showed that the developed model has better performance in predicting water inflow into the tunnel than other methods, such as support vector machine (SVM) and artificial neural network (ANN). Liu et al. [14] used the hybrid grey wolf optimization algorithm and support vector regression method (HGWO-SVR). They concluded that the HGWO-SVR model has high precision in the prediction of water inrush into the tunnel. Mahmoodzadeh et al. [15] applied an optimized model based on the gene expression programming (GEP) method to predict the WI in drill-and-blast tunnels. Zhou et al. [16] introduced a hybrid grey wolf optimization and random forest algorithm (GWO-RF) for the prediction of WI in tunnels. According to their findings, the suggested model outperforms other conventional ensemble models in terms of accuracy during both the training and testing stages. Jahanmiri et al. [17] utilized the Neuro-swarm, GEP tree-based, and whale optimization models for evaluating groundwater seepage in the tunnels. Their results showed that the whale optimization algorithm is the best model for water lost in the tunnel. Samadi et al. [1] developed several predictor networks, such as AdaGrad-gated recurrent unit, AdaGradlong short-term memory, AdaDelta-recurrent neural network, automatic linear forward stepwise information criterion, a novel stacking-ensemble model, and Adam optimization-back propagation neural network to predict WI in drill-and-blast tunnels. They reached the conclusion that the AdaG-GRU and stacking-ensemble models were the most reliable for forecasting WI in tunnels. Shen et al. [18] presented a mathematical model based on Physics-Informed Neural Networks (PINN) to determine the groundwater flow in the tunnel. Their results showed high accuracy in predicting underground water flow into the tunnel and agreement with analytical methods. Chen et al. [19] proposed a novel convolutional neural network (CNN)-based model to evaluate the water inflow into the tunnel. Ju et al. [20] used the eXtreme Gradient Boosting (XGBoost) model optimized with Bayesian optimization (BO) to predict the tunnel water inrush volumes. Their results showed that groundwater level is the most important input parameter in predicting tunnel water inrush volumes.

In the present study, five machine learning algorithms, including Random Forest (RF), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and Natural Gradient Boosting (NGBoost) have been harnessed to predict the WI in water conveyance tunnels. This study focuses on integrating three vital geomechanical and hydrogeological parameters, including hydraulic conductivity (K), geological strength index (GSI), and water head (H) to examine their combined impacts on WI in water conveyance tunnels. Next, the SHAP method has been used to reveal the water head as the most critical parameter, enhancing interpretability. In addition to traditional evaluation metrics, this study employs the Regression Error Characteristic (REC) curve and the Area Under the REC Curve (AOC) to assess model performance. The REC curve visualizes the cumulative distribution of absolute prediction errors, offering a comprehensive view of model accuracy across the dataset. The AOC, calculated as the area under the REC curve, quantifies the overall prediction error [21]. By integrating a real-world dataset from six tunnels in Iran, optimized models, and advanced evaluation metrics, this research offers a versatile, interpretable tool for managing hydrological risks in tunneling, extending prior work with broader applicability. In particular, the limitations of existing empirical, analytical, and hybrid machine-learning approaches for tunnel water inflow (WI) prediction have been critically discussed, emphasizing challenges related to nonlinear hydrogeological interactions, limited interpretability, and insufficient comparative benchmarking. This study explicitly highlights the study's contributions: (i) the systematic comparison of five advanced ensemble machine-learning models (RF, GB, XGBoost, CatBoost, and NGBoost), (ii) the integration of Bayesian hyperparameter optimization using Optuna to ensure robust and efficient model calibration, (iii) the incorporation of REC curves and AOC analysis as distribution-sensitive evaluation metrics, and (iv) the application of SHAP and PDP techniques to enhance interpretability using a real-world multi-tunnel dataset.

2. Machine learning algorithms

2.1. Random Forest (RF)

One well-known machine-learning technique is the RF method, which makes use of decision trees. To build an RF model, the following steps can be utilized [16]:

(A) Using input parameter embedding and bootstrapping, a collection of n trees is produced.

(B) Unpruned regression trees and random samples are used to choose the greatest number of split predictors.

(C) The number of decision trees to be employed (n_estimators) is combined through aggregation to anticipate the actual values.

2.2. Gradient Boosting (GB)

The main advantage of the GB is that it can combine weak learners to produce a more accurate model. It begins with an initialization, then uses the errors of the previous models to justify the next models, and it computes the gradient of the loss function based on ensemble predictions to indicate the size and direction of the update required to minimize the loss function [22].

2.3. Extreme Gradient Boosting (XGBoost)

XGBoost is among the algorithms developed based on the GB method, which improves efficiency and model prediction accuracy [23]. XGBoost checks it on the testing dataset concurrently with the training phases to improve performance. When the condition is satisfied, the modeling either ends or restarts the training process with a new iteration count.

2.4. Categorical Boosting (CatBoost)

The method can manage different types of data, from homogeneous to heterogeneous, used to predict a particular parameter. In order to assign numerical values to the categorical variables based on their goal statistics and how they appear on the datasets, it uses a technique called the "ordered goal encoding technique" [22].

2.5. Natural Gradient Boosting (NGBoost)

A supervised ML technique for general probabilistic prediction. The fundamental idea behind NGBoost is to estimate the parameters of a conditional probability distribution using a boosting technique. The three primary components of this machine learning system are the scoring rule, parametric probability distribution, and base learner [24].

2.6. Hyperparameter tuning

The 'Optuna' platform was utilized to optimize the hyperparameters of the models for better results. Table 1 lists each model's hyperparameter tuning values. The optimal parameters can only be discovered after completing the entire procedure mentioned above. Figure 1 shows the overall architecture of the ML models.

Hyperparameter tuning was conducted using Optuna due to its efficiency, adaptability, and suitability for optimizing multiple machine learning models with heterogeneous parameter spaces. Optuna employs a Tree-structured Parzen Estimator (TPE), a Bayesian optimization-based algorithm that dynamically guides the search process toward promising regions of the hyperparameter space. Unlike exhaustive approaches such as Grid Search, which evaluate all possible parameter combinations and become computationally prohibitive in high-dimensional settings, Optuna updates its sampling strategy based on prior trials, thereby substantially reducing computational cost while maintaining optimization accuracy.

Furthermore, Optuna offers advanced features such as adaptive sampling, pruning of unpromising trials, and support for conditional and non-continuous hyperparameters, making it particularly effective for tuning complex ensemble models. Given the dataset size (73 tunnel sections) and the need to optimize several models with diverse hyperparameters (e.g. n_estimators, learning_rate, and max_depth), Optuna provided a balanced solution between computational efficiency and predictive performance. Its seamless integration with common machine learning libraries further facilitated systematic optimization, ultimately contributing to the robust predictive results reported in this study.

Table 1. The best values of parameters in the used ML models.

Model	Parameters	Value	Model	Parameters	Value
RF	n_estimators	24	XGBoost	n_estimators	44
	max_depth	42		learning_rate	0.181
	min_samples_split	4		max_depth	45
	min_samples_leaf	1		min_child_weight	2
	min_child_weight	1	CatBoost	n_estimators	49
GB	n_estimators	35		learning_rate	0.218
	learning_rate	0.299		max_depth	9
	max_depth	35	NGBoost	n_estimators	10
	min_samples_split	3		learning_rate	5
	min_samples_leaf	1		max_depth	0.269

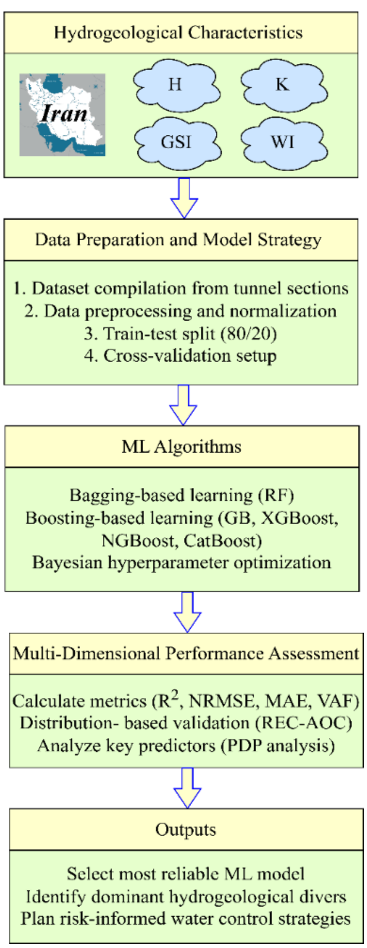

Figure 1. Conceptual framework for ML-based prediction of tunnel water inflow.

3. Data preprocessing

3.1. Data gathering

This study involved collecting and compiling data from six water conveyance tunnels in different geological regions of Iran: Gelas water conveyance tunnel (lot 2), Kerman water conveyance tunnel, AmirKabir conveyance tunnel, Zagros water conveyance tunnel, Lar-Kalan water conveyance tunnel, and Rozieh water conveyance tunnel.

The geological investigation of the Gelas tunnel showed that the major part of the tunnel route was hosted in a sequence of granite and granodiorite units [25]. The rock mass of the northern section of the Kerman water conveyance tunnel belongs to the Precambrian Era. This section is constituted by different sedimentary and volcanic rock sets [8]. Additionally, based on geological studies, the AmirKabir tunnel generally includes various sedimentary-volcanic complexes of the Karaj formation. Its lithology consists of a sequence of tuff, sandstone, fine-grained conglomerates, siltstone, shale, and agglomerate [26]. The main geological formations outcropped in the project area of the Zagros tunnel are Pabdeh, Gurpi and Ilam Formations. These formations are composed of a variety of carbonate and argillaceous rock units [27]. Based on geological studies, the Lar-Kalan tunnel includes sedimentary and sometimes chemical and biochemical, pyroclastic, and igneous rock units [7]. Finally, about 75% of the route of the Rozieh tunnel is made up of limestone, dolomite, silt, shale, and sandstone, and the rest of the tunnel route is covered with rocks such as tuff.

The used database includes 73 zones of the tunnels. Hydraulic conductivity (K), water head (H), and geological strength index (GSI) are the input parameters used to create ML models. Also, the output is the water inflow into the tunnel for each section. Based on the literature survey, these three input parameters are among the most effective parameters to predict the WI in tunnels [15, 28].

3.2. Data analysis

The collected dataset, comprising 73 zones from six water conveyance tunnels, was divided into training and testing subsets to ensure robust model development and evaluation. Specifically, 80% of the data was allocated for training, while the remaining 20% was reserved for testing. To maximize the utilization of the training data and enhance model generalization, a 5-fold cross-validation strategy was applied to the training subset. This approach ensured that all training data contributed to the model training process, reducing the risk of overfitting and improving the robustness of the machine learning models. The trained models were then evaluated on the separate 20% test set to assess their predictive performance.

The dataset used in this study consists of recorded measurements from tunnel sections where water inflow (WI) was monitored alongside key hydrogeological and geomechanical parameters. A comprehensive statistical characterization of the dataset, including the maximum, minimum, mean, and standard deviation of each variable, has been incorporated to improve transparency and reproducibility. The input variables-water head, hydraulic conductivity, and GSI-were selected based on well-established hydro-mechanical principles governing groundwater flow toward underground excavations. Water head represents the primary hydraulic driving force controlling seepage gradients; hydraulic conductivity quantifies the permeability and transmissivity of the surrounding rock mass; and GSI captures rock mass structure, fracture condition, and degree of block interlocking, which influence flow channelization and discontinuity connectivity. The selection of these three parameters was therefore physically justified rather than purely data-driven, ensuring that the modeling framework remains grounded in hydrogeological theory while capturing the dominant controls on tunnel water inflow.

Figure 2 displays the histograms, cumulative distributions, and basic descriptions of input and output parameters. The scatterplot matrix was also used to examine the relationships between hydraulic conductivity, water head, geological strength index with water inflow into the tunnel (Figure 3). This scatterplot matrix reveals no significant multicollinearity among the input features and water head. The absence of strong linear correlations ensures that these features are relatively independent, enhancing the interpretability of tree-based models like XGBoost, RF, GB, CatBoost, and NGBoost. In addition, goodness-of-fit tests in the EasyFit program were used to determine the best-fitted distribution functions on the three input parameters. The results of three goodness-of-fit tests, including Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared, were presented in Table 2.

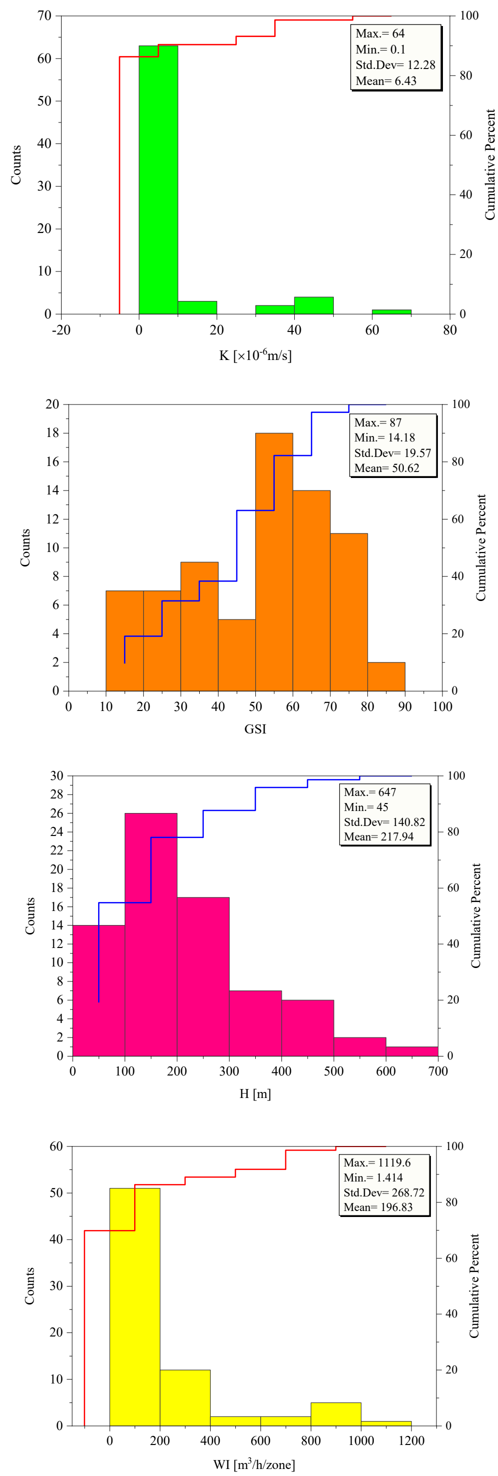

Figure 2. Cumulative frequency histograms for all parameters.

Table 2. Goodness-of-fit tests for input variables.

Input variables	Tests		
	Kolmogorov-Smirnov	Anderson-Darling	Chi-Squared
K	Dagum (4P)	Lognormal (3P)	Error Function
H	Gen. Pareto	Johnson SB	Log-Pearson 3
GSI	Dagum	Wakeby	Cauchy

To ensure robust predictive performance, hyperparameter optimization was performed using a Bayesian optimization strategy based on the Tree-structured Parzen Estimator (TPE) algorithm. The optimization objective was defined as minimizing the cross-validated Root Mean Square Error (RMSE) on the training dataset, thereby directly targeting prediction accuracy while controlling generalization error. A predefined number of optimization trials was conducted for each model, and early-stopping (pruning) rules were implemented to terminate underperforming trials based on intermediate validation scores. This adaptive search strategy enabled efficient exploration of high-dimensional hyperparameter spaces while avoiding exhaustive computation. The use of multiple ensemble-based models (bagging- and boosting-oriented algorithms) was not arbitrary; rather, it allowed systematic evaluation of different ensemble learning mechanisms in capturing nonlinear interactions among hydrogeological variables. This structured methodological design enhances both the reliability and interpretability of the proposed predictive framework.

4. Evaluation of the results

Before implementing intelligent models in geotechnical engineering applications, it's important to evaluate and validate their reliability [29]. The evaluation of the model's performance was done using traditional metrics. These metrics are the coefficient of determination (R2), the normalized root mean squared error (NRMSE), and the variance account for (VAF) values. These indices are elucidated using Eqs. (1)-(3) [22].

$$R^2 = 1 - \frac{\sum_{i=1}(WI_{observed} - WI_{predicted})^2}{\sum_{i=1}(WI_{observed} - \overline{WI})^2} \quad (1)$$

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(WI_{observed} - WI_{predicted})^2}}{\overline{WI}} \quad (2)$$

$$VAF = \left[1 - \frac{var(WI_{observed} - WI_{predicted})}{var(WI_{predicted})}\right] \times 100 \quad (3)$$

5. Results and discussions

5.1. Performance evaluation of ML models

Following the optimization of hyperparameters, the five machine learning models were trained using 80% of the dataset and validated on the remaining 20%, ensuring an unbiased assessment of predictive capability. The graphical comparison between observed and predicted water inflow values provides an initial qualitative understanding of model behavior. Figures 4 to 8 demonstrate the predicted WI values of the models alongside the observed values.

A more rigorous quantitative comparison using statistical performance indices further substantiates these observations. The XGBoost model achieved the highest R^2 alongside the lowest NRMSE and the highest VAF values among all tested models (Figure 9). The elevated R^2 reflects the model's ability to explain a substantial proportion of the variance in WI, while the reduced NRMSE indicates lower relative prediction error when normalized by the magnitude of observed values. Similarly, the higher VAF confirms strong agreement between predicted and measured inflow rates. Although the GB model also demonstrated competitive performance, its predictive accuracy remained slightly inferior to that of XGBoost across all evaluation metrics. This performance gap can be attributed to XGBoost's advanced

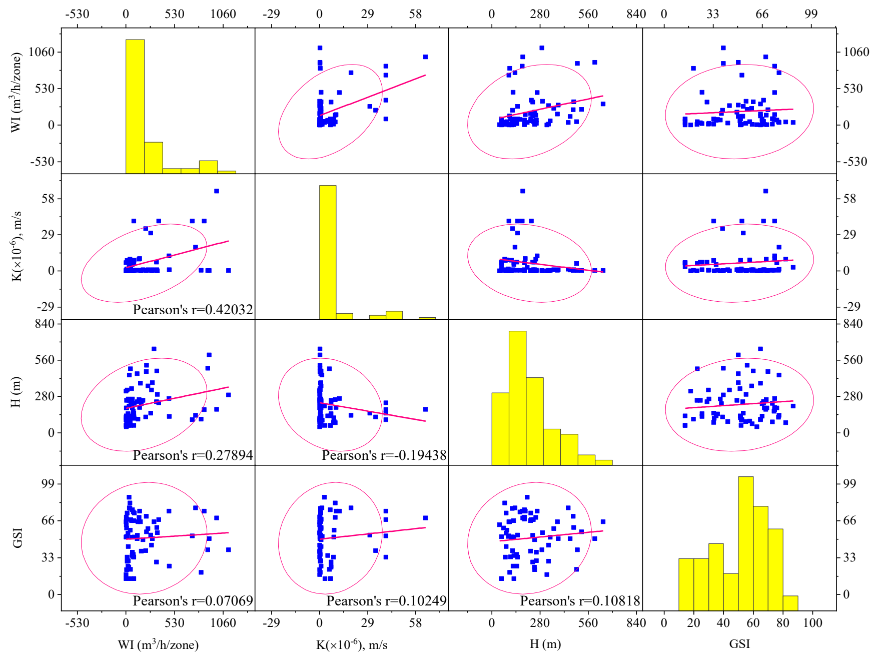

Figure 3. Scatterplot matrix between all parameters.

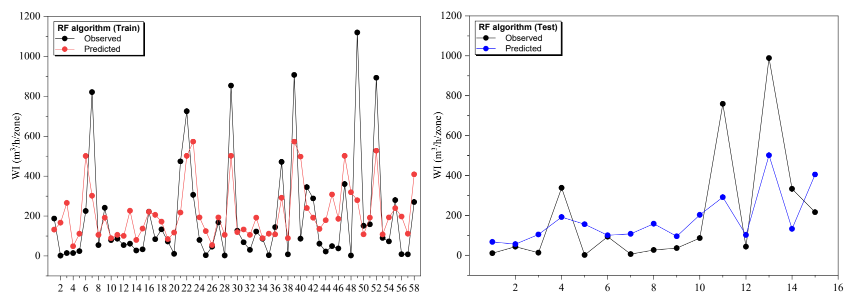

Figure 4. Comparison of the observed and predicted values of WI based on the RF model.

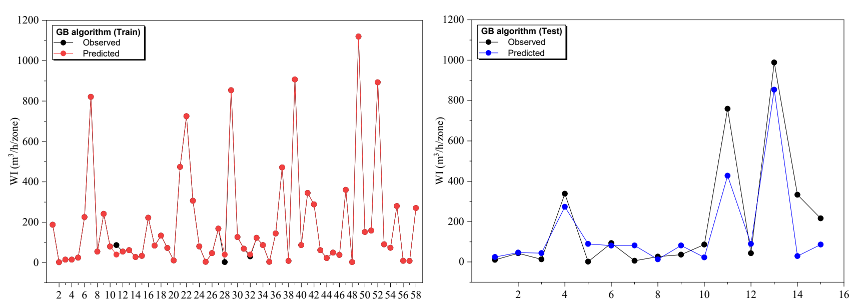

Figure 5. Comparison of the observed and predicted values of WI based on the GB model.

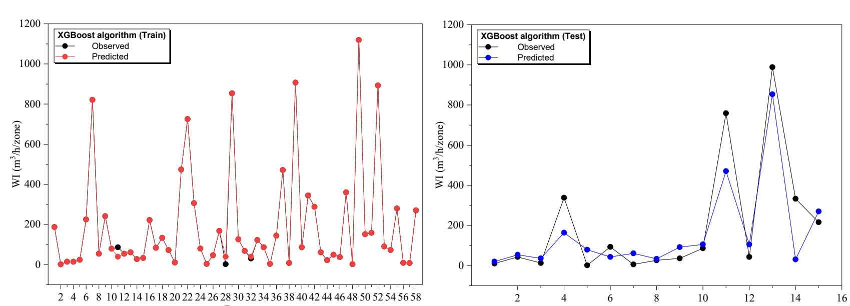

Figure 6. Comparison of the observed and predicted values of WI based on the XGBoost model.

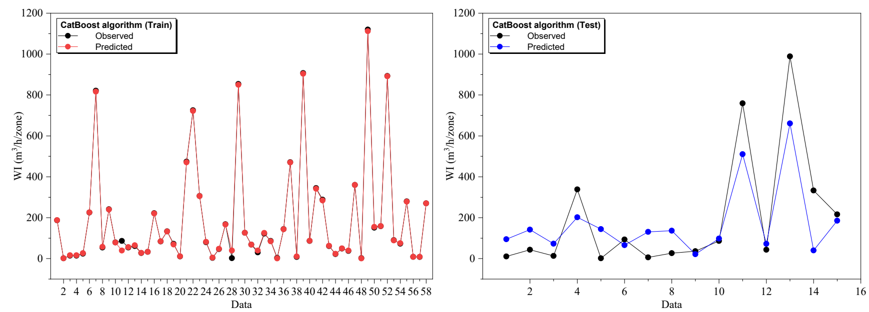

Figure 7. Comparison of the observed and predicted values of WI based on the CatBoost model.

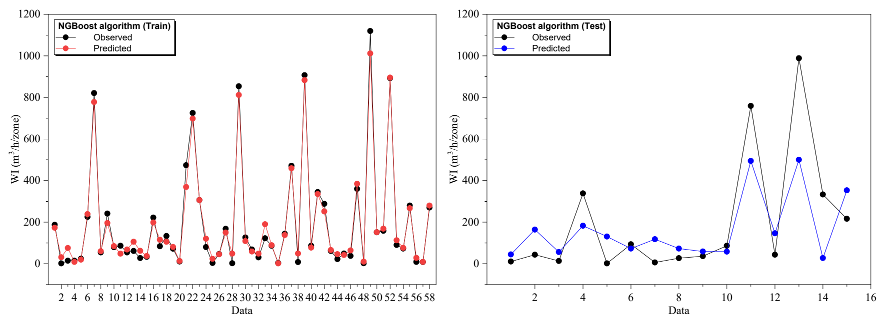

Figure 8. Comparison of the observed and predicted values of WI based on the NGBoost model.

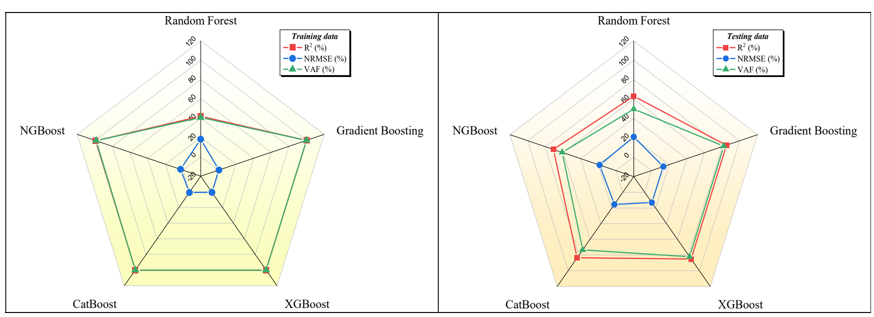

Figure 9. Performance metrics of the ML models.

regularization mechanisms, optimized tree pruning strategy, and second-order gradient information, which collectively enhance its ability to model complex nonlinear interactions among hydraulic conductivity, water head, and GSI.

The comparatively weaker performance of RF and NGBoost suggests limitations in their capacity to fully represent the coupled hydro-mechanical behaviour influencing WI. While RF benefits from variance reduction through ensemble averaging, its bagging-based structure may not correct residual errors as effectively as boosting-based approaches. NGBoost, despite its probabilistic framework, may exhibit sensitivity to limited dataset size and distributional assumptions, which can affect predictive stability in moderate-sample geotechnical datasets. In contrast, boosting algorithms iteratively refine residual errors, enabling them to adapt more effectively to subtle nonlinear dependencies and interaction effects inherent in heterogeneous rock masses.

From an engineering standpoint, the superior performance of XGBoost has important implications. Accurate WI prediction requires the capability to capture nonlinear amplification effects, particularly under high hydraulic head conditions where small variations in permeability or rock mass quality can trigger disproportionately large inflow responses. The enhanced predictive consistency observed in XGBoost suggests that it is better suited for modeling threshold-driven and interaction-dominated hydrogeological phenomena. Moreover, the narrow dispersion between training and testing performance metrics indicates strong generalization ability and minimal overfitting, reinforcing the reliability of the model for practical applications.

Overall, the comparative analysis demonstrates that boosting strategies, particularly XGBoost, provide a robust and technically sound framework for predicting tunnel water inflow. The combination of high explanatory power, reduced prediction error, and stable generalization underscores the suitability of XGBoost as a decision-support tool for hydrogeological risk assessment in underground construction projects.

5.2. Score analysis

To provide a more comprehensive and decision-oriented comparison among the developed machine learning models, a multi-criteria score analysis framework was implemented. While individual statistical indicators such as R^2, NRMSE, and VAF offer valuable insights into specific aspects of model performance, relying solely on a single metric may lead to biased or incomplete conclusions. Therefore, the present study employed a structured ranking procedure in which each model was evaluated simultaneously across four complementary performance indices. For each metric, models were ranked according to their predictive quality, and a score ranging from one (lowest performance) to five (highest performance) was assigned. In score analysis, models are allocated scores based on various evaluation metrics, and the model with the performance metrics closest to the desired values requires the highest score [30–32]. This approach allows the integration of goodness-of-fit, error magnitude, and variance explanation into a unified comparative scheme.

The results of the scoring process, summarized in Table 3, reveal a clear performance hierarchy among the evaluated algorithms. XGBoost obtained the highest cumulative score, indicating consistent superiority across all four statistical criteria. This finding demonstrates that its predictive capability is not limited to excelling in a single indicator but is reflected uniformly in high explanatory power (R^2), low normalized prediction error (NRMSE), minimized absolute deviation (MAE), and strong agreement between predicted and observed values (VAF). Gradient Boosting ranked second, showing competitive yet slightly less stable performance compared with XGBoost. In contrast, Random Forest and NGBoost received lower total scores, suggesting comparatively weaker generalization and higher residual dispersion.

From a methodological standpoint, the use of score analysis enhances the robustness of model selection by reducing ambiguity when performance metrics yield marginal differences. In geotechnical applications such as tunnel water inflow prediction, where safety, reliability, and risk mitigation are paramount, model evaluation must reflect overall stability rather than isolated statistical excellence. The dominance of XGBoost in the scoring framework corroborates the conclusions derived from graphical assessments and numerical performance indices, confirming its robustness, consistency, and practical suitability for hydrogeological prediction tasks. Consequently, the score analysis provides an objective and transparent basis for identifying XGBoost as the most reliable and technically appropriate model for estimating water inflow in complex tunneling environments.

Table 3. The results of the score analysis for the applied ML models.

Model	NRMSE	VAF	MAE	R^2	Total
GB	4	4	4	4	16
XGBoost	5	5	5	5	20
RF	1	1	1	1	4
CatBoost	3	3	3	3	12
NGBoost	2	2	2	2	8

5.3. Regression error characteristic curve

Regression models' prediction performance is graphically represented by the Regression Error Characteristic (REC) curve. The REC curve, in contrast to traditional metrics, concentrates on the absolute error's cumulative distribution function. The area under the REC curve (AOC) serves as an estimate of the prediction error, where a smaller AOC value indicates a more accurate model [21, 33]. The AOC derived from the Regression Error Characteristic (REC) curve provides a complementary perspective to conventional point-wise error metrics such as RMSE and MAE. While RMSE and MAE quantify the average magnitude of prediction errors, they do not explicitly describe how errors are distributed across different tolerance levels. In contrast, the REC curve represents the cumulative distribution of absolute errors over a continuum of admissible error thresholds. Consequently, the AOC summarizes the overall dispersion of prediction errors across all tolerance levels, offering a more comprehensive assessment of model reliability. A smaller AOC indicates that a larger proportion of predictions fall within tighter error bounds, reflecting higher predictive precision over a broad range of acceptable deviations.

From a practical standpoint, the AOC is particularly relevant in engineering and field-based applications where model usability depends not only on average error magnitude but also on the proportion of predictions satisfying predefined accuracy thresholds. For instance, in rock engineering estimation tasks, practitioners often operate within specific tolerance margins dictated by design or safety considerations. The AOC inherently captures model behavior across such tolerance intervals, thereby facilitating an application-oriented evaluation of predictive performance.

The calculated AOC values for all machine learning models are presented in Table 4, and the corresponding REC curves are illustrated in Figure 10. The graphical comparison reveals clear differences in the steepness and convergence patterns of the curves. Among the evaluated algorithms, XGBoost exhibits the lowest AOC value, indicating the highest concentration of predictions within tight error bounds. Its REC curve consistently dominates those of the competing models across nearly all tolerance thresholds, reflecting superior stability and error containment. GB demonstrates competitive but slightly inferior performance, whereas RF and NGBoost display flatter curves, suggesting greater dispersion of residuals.

Overall, the REC-AOC analysis reinforces the conclusions derived from traditional performance metrics and score analysis. The superior behavior of XGBoost in both numerical and distribution-based evaluations confirm its robustness, reliability, and suitability for predicting tunnel water inflow under complex hydrogeological conditions. By incorporating distribution-sensitive assessment tools such as the REC curve, the present study ensures a more comprehensive and application-oriented validation of the developed predictive models.

Table 4. The values of AOC for the used ML models.

ML model	AOC
GB	31.45
XGBoost	18.27
RF	30.68
CatBoost	36.91
BGBoost	36.21

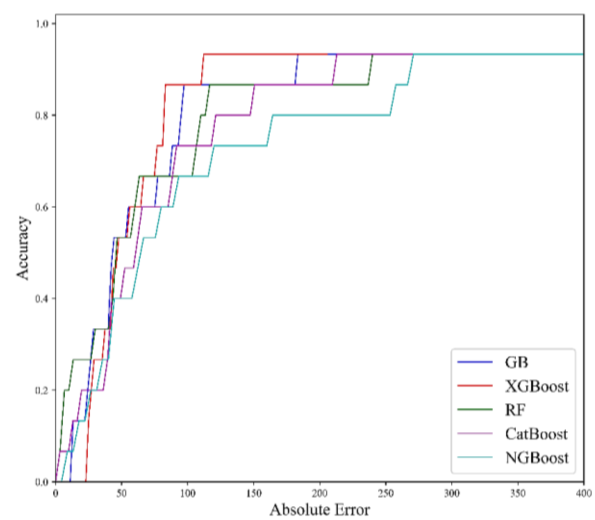

Figure 10. The REC curves of applied ML models.

6. Shapley Additive Interpretation (SHAP) analysis

Each feature's proportionate contribution to the anticipated results is calculated using the Shapley Additive Interpretation (SHAP) algorithm. The significance of the input variables in the ML model is measured by the higher Shapley values assigned to features that have a greater influence on the output [34].

Figure 11 displays the output of the SHAP value analysis in a force diagram for predicting the WI in ground behavior. The diagram illustrates the prediction process, starting from the SHAP baseline to the final output. Features are represented as forces driving the prediction. Water head has the greatest impact, with a long blue arrow indicating a significant reduction in the WI prediction. At the same time, the hydraulic conductivity and GSI have a positive impact (red arrow). However, water head, due to its longer arrow, is the most influential factor in reducing the WI.

Figure 12 illustrates how each input parameter affects the WI prediction outcomes. The horizontal axis represents the SHAP values, whereas the vertical axis denotes the properties that are pivotal for the prediction made by the model. The feature's exact value is shown by the color bar. According to Figure 12, the input parameters' contribution to the prediction of WI into the tunnels is as follows: water head (H)> hydraulic conductivity (K)> GSI. Notably, the water head has the greatest impact on water inflow into the tunnels. It should be mentioned that, given the superior performance of the XGBoost model, SHAP analysis was conducted specifically for this model to ensure interpretability of the most accurate predictions. As all models (RF, GB, XGBoost, CatBoost, NGBoost) are tree-based, supplementary SHAP analyses confirmed consistent feature importance rankings across models, with water head as the most influential parameter, followed by hydraulic conductivity and GSI. For brevity and focus, only the XGBoost SHAP results are presented.

To evaluate the robustness of the XGBoost regression model, the Gaussian noise was introduced at levels of 0.1, 0.3, and 0.5 to the test data features. The model's performance and feature importance, analyzed via SHAP values, remained stable across all noise levels. As can be seen in Figure 13, the model is robust to perturbations in the input data.

According to the results presented in this study and previous studies, it can be said that ML algorithms have been widely used to predict the WI by tunnel engineers and geologists. In other words, by considering the geological and hydrological conditions of other tunnels, precise predictions can be made using the hybrid ML methods. By considering the geological and hydrological characteristics of each tunnel, including water head, rock mass quality, tunnel radius, permeability, tunnel depth, etc., it is possible to make successful predictions of the amount of water inflow into the tunnel before encountering any risk of water inrush inside the tunnel. In this case, in addition to saving time and cost, it is possible to reduce the risk of tunnel construction in each tunnel zone.

The XGBoost model, with its high accuracy, is well-suited for real-time water inflow prediction in tunneling operations. Integrated with TBM systems, it can use real-time sensor data (hydraulic conductivity, GSI, water head) from borehole tests and piezometers to dynamically predict WI, enabling proactive measures like grouting to enhance safety and efficiency. SHAP analysis highlights water head as the most critical parameter. However, challenges such as data quality, computational demands, model retraining, operational integration, and regulatory compliance must be addressed. Solutions include robust sensor networks, edge computing, online learning, user-friendly interfaces, and human oversight to ensure reliable deployment in diverse geological settings.

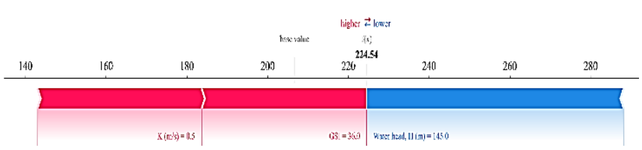

Figure 11. Shapley values explain individual predictors of WI.

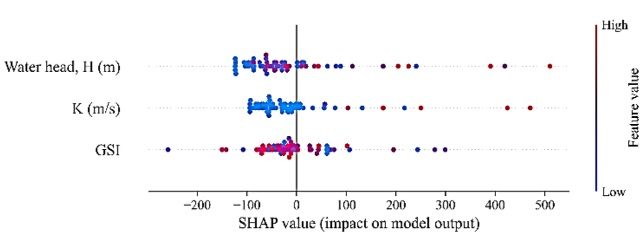

Figure 12. SHAP global explanation on XGBoost model.

7. Partial Dependence Plot (PDP) analysis

To enhance the interpretability of the developed machine learning framework, Partial Dependence Plot (PDP) analysis was conducted for the optimal XGBoost model. The PDP method is a powerful tool for analyzing the impact of individual input features on the output of machine learning models [35]. PDPs provide a global sensitivity assessment by quantifying the marginal effect of an individual predictor on the model response while averaging out the influence of other variables. Unlike correlation analysis which captures linear pairwise associations, PDPs reflect the actual functional relationship learned by the nonlinear ensemble model. Consequently, they allow for a deeper understanding of how each geotechnical and hydrogeological parameter contributes to tunnel water inflow (WI) predictions within the multidimensional feature space.

Based on the three input variables, PDPs were generated to visualize their marginal contributions to the predicted WI (Figure 14). The results clearly indicate that water head exerts the most significant influence on model output. This dominance is evidenced by the pronounced slope and wider vertical variation of its PDP curve compared to the other predictors. The steep gradient of the water head curve reflects strong model sensitivity, meaning that incremental increases in hydraulic head result in substantial changes in predicted inflow. From a hydrogeological standpoint, this behavior is physically consistent, as hydraulic head directly controls the driving force of groundwater flow toward the excavation boundary, in accordance with fundamental seepage principles.

In contrast, hydraulic conductivity exhibits a comparatively smoother and more gradual PDP trend, suggesting a moderate but less dominant contribution to WI variability within the studied dataset. Although permeability governs the ease of fluid movement through the rock mass, its effect appears to be partially moderated by the distribution of hydraulic gradients and geological conditions captured in the model. Similarly, GSI demonstrates a relatively limited marginal effect, with smaller fluctuations in the PDP curve. This implies that while rock mass quality influences fracture connectivity and flow pathways, its direct impact on inflow magnitude is secondary to the hydraulic forcing represented by water head.

Overall, the PDP analysis not only confirms the statistical importance of water head but also aligns with established hydro-mechanical principles governing groundwater inflow in tunnels. By integrating interpretable machine learning techniques with domain knowledge, the study strengthens confidence in the physical plausibility and engineering reliability of the XGBoost model. The dominance of water head in the PDP framework highlights the necessity of accurate hydrostatic pressure characterization during tunnel design and risk assessment, as even moderate variations in head conditions may significantly alter predicted inflow rates.

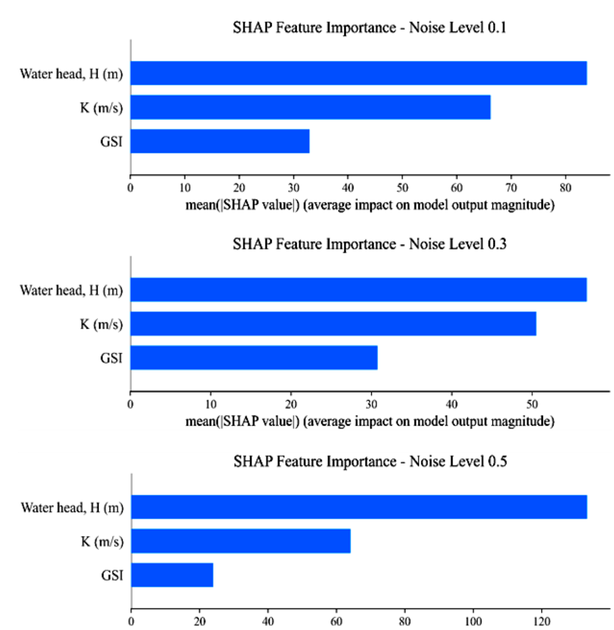

Figure 13. Feature importance for water inflow prediction with Gaussian noise.

8. Discussion

The present study demonstrates that advanced machine learning algorithms can effectively capture the complex, nonlinear interactions governing tunnel water inflow. The comparative results consistently indicate that boosting-based frameworks-particularly XGBoost-provide superior predictive accuracy relative to bagging-based and probabilistic ensemble approaches. This performance advantage can be attributed to the sequential error-correction mechanism inherent in gradient boosting algorithms, which iteratively reduces residual errors and improves generalization. In hydrogeological systems where inflow is controlled by nonlinear coupling between hydraulic gradients, permeability structure, and rock mass condition, such adaptive learning strategies are particularly advantageous. The ability of XGBoost to minimize both average error magnitude and error dispersion, as demonstrated through conventional metrics and REC-AOC analysis, confirms its robustness in handling heterogeneous geotechnical datasets.

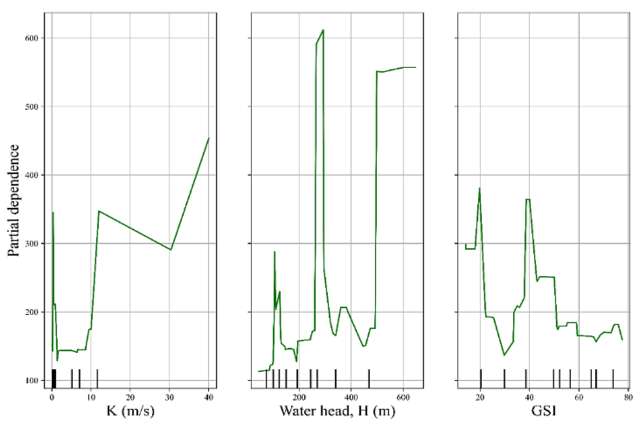

Figure 14. PDP analysis for the XGBoost model.

From a hydro-mechanical perspective, the dominance of water head identified in the PDP analysis is physically consistent with seepage theory. Hydraulic head directly governs the driving force for groundwater migration toward underground excavations, and even moderate variations can significantly alter inflow magnitude due to changes in hydraulic gradient. The comparatively smaller marginal influence of hydraulic conductivity and GSI does not imply limited importance; rather, it suggests that within the studied dataset, variability in inflow is more sensitive to hydraulic forcing than to material resistance parameters. Hydraulic conductivity controls transmissivity, but its influence becomes pronounced primarily under substantial head differentials. Similarly, GSI reflects structural integrity and discontinuity conditions, which indirectly affect fracture connectivity and permeability anisotropy. The model's ability to reflect these physically meaningful relationships enhances confidence in its engineering credibility and reduces concerns regarding purely data-driven overfitting.

The integration of REC-AOC analysis adds an important dimension to model validation. While traditional indices such as RMSE and MAE quantify the central tendency of prediction errors, they do not explicitly characterize the distribution of residuals across different tolerance levels. In tunnel engineering, where underestimation of inflow may lead to instability, excessive dewatering costs, or safety risks, understanding how many predictions fall within acceptable error thresholds is critical. The lower AOC value obtained for XGBoost indicates that a greater proportion of predictions remain confined within narrow deviation bounds, reinforcing its suitability for risk-sensitive design applications. This distribution-based validation framework strengthens the reliability assessment beyond single-value summary statistics.

Another important aspect of the findings is the structured comparison among ensemble learning philosophies. The results reveal that boosting-based methods outperform bagging-oriented RF and probabilistic NGBoost approaches in this specific hydrogeological context. This suggests that sequential learning with gradient-based optimization is more effective in modeling nonlinear seepage behavior than variance-reduction strategies alone. The observed performance hierarchy provides methodological insight for future geotechnical predictive modeling, indicating that algorithm selection should consider the physical complexity of the target phenomenon rather than relying solely on model popularity.

From a practical standpoint, the ML framework can be implemented during the preliminary design phase using site investigation data (e.g. borehole permeability tests, piezometric measurements, and rock mass classification).

Once trained, the optimal model (XGBoost) can rapidly estimate expected water inflow rates for different tunnel sections without requiring complex numerical groundwater simulations. This enables engineers to perform rapid scenario analyses under varying hydraulic head or permeability conditions and to identify high-risk zones prior to excavation. During construction, the model can be updated iteratively as new monitoring data become available, thereby supporting adaptive risk management and real-time decision-making. Such integration allows contractors to optimize drainage design, pre-grouting programs, and dewatering capacity planning, reducing both safety risks and unnecessary mitigation costs.

9. Conclusions

This study tried to predict water inflow (WI) in water conveyance tunnels using various machine learning algorithms by considering critical geomechanical and hydrological parameters, including hydraulic conductivity, GSI, and water head. For this purpose, five models, including Random Forest, Gradient Boosting, Extreme Gradient Boosting, Categorical Boosting, and Natural Gradient Boosting, were used. The study's findings could be interpreted as follows:

- The results revealed that the XGBoost model outperformed the other models in terms of accuracy and robustness. This model exhibited the highest performance in evaluation metrics such as R2, NRMSE, and VAF. In visual analyses of the prediction results, XGBoost showed better performance compared to RF and NGBoost models, with predicted points being closer to the 95% prediction band. Additionally, XGBoost has the lowest prediction error among the other models.
- Comparing the results of the ML models in terms of the REC and AOC analysis further demonstrated that XGBoost provided better predictions, as it achieved the lowest AOC value, indicating higher prediction accuracy. These findings highlight the model's capability in predicting WI in tunnels.
- Sensitivity analysis using the SHAP and PDP methods revealed that the water head had the most significant impact on WI predictions. Hydraulic conductivity and GSI followed, showing less influence on the WI predictions.
- U ltimately, this study underscores the potential of machine learning algorithms to significantly enhance the accuracy of geotechnical predictions, particularly in tunnel projects involving complex hydrological and geological conditions. The XGBoost model, identified as the most effective, can be used as a reliable tool for predicting and managing water inflow in tunnels. Specifically, this model can improve safety, efficiency, and reduce costs in tunnel construction projects, effectively mitigating risks associated with water ingress.

The used methods for WI prediction hold widespread applicability in similar geotechnical and hydrological projects, aiding in better predictions and decision-making related to managing hydrological hazards in underground construction projects. In fact, considering the high potential that the used ML algorithms have shown in solving various civil engineering, rock mechanics, geotechnical, etc. problems, this study enables to examine their ability in estimating the phenomenon of water flow in tunnels.

Competing interests

All authors declare that they have no conflict of interest.

Funding

No funding was received for this research.

Data availability

The datasets used and analyzed during the current study are available from the corresponding author on request.

References

[1]. Samadi, H., Mahmoodzadeh, A., Elhag, A. B., Alanazi, A., Alqahtani, A., & Alsubai, S. (2025). Application of hybrid-optimized and stacking-ensemble labeled neural networks to predict water inflow in drill-and-blast tunnels. *Tunnelling and Underground Space Technology*, *156*, 106273. doi: https://doi.org/10.1016/j.tust.2024.106273

[2]. Qi, B., Xu, P., & Wu, C. (2023). Analysis of the infiltration and water storage performance of recycled brick mix aggregates in sponge city construction. *Water*, *15*(2), 363. doi: https://doi.org/10.3390/w15020363

[3]. Fu, Q., Gu, M., Yuan, J., & Lin, Y. (2022). Experimental study on vibration velocity of piled raft supported embankment and foundation for ballastless high speed railway. *Buildings*, *12*(11), 1982. doi: https://doi.org/10.3390/buildings12111982

[4]. Park, K.-H., Owatsiriwong, A., & Lee, J.-G. (2008). Analytical solution for steady-state groundwater inflow into a drained circular tunnel in a semi-infinite aquifer: A revisit. *Tunnelling and Underground Space Technology*, *23*(2), 206-209. doi: https://doi.org/10.1016/j.tust.2007.02.004

[5]. Li, D., Li, X., Li, C. C., Huang, B., Gong, F., & Zhang, W. (2009). Case studies of groundwater flow into tunnels and an innovative water-gathering system for water drainage. *Tunnelling and Underground Space Technology*, *24*(3), 260-268. doi: https://doi.org/10.1016/j.tust.2008.08.006

[6]. West, G. (1983). Comparisons between real and predicted geology in tunnels: examples from recent cases. *Quarterly Journal of Engineering Geology and Hydrogeology*, *16*(2), 113-126. doi: https://doi.org/10.1144/GSL.QJEG.1983.016.02.04

[7]. Ghorbani, S., Bour, K., & Javdan, R. (2024). Evaluating the Efficiency of Pre-Grouting in Fault Zones of Tunnel and its Effect on Rock Mass Quality: A Case Study. *Geotechnical and Geological Engineering*, *42*(7), 6641-6654. doi: https://doi.org/10.1007/s10706-023-02557-8

[8]. Ghorbani, S., Bour, K., Javdan, R., & Bour, M. (2023). Design of Effective Grouting Pattern in Kerman Water Conveyance Tunnel Using DFN-DEM and Analytical Approaches. *International Journal of Geosynthetics and Ground Engineering*, *9*(2), 19. doi: https://doi.org/10.1007/s40891-023-00441-2

[9]. Ghorbani, S., Bour, K., & Javdan, R. (2024). Investigating the sealing efficiency of grouting in joints: insights from effects of the rheological properties of grout and joint characteristics. *Geomechanics and Geoengineering*, *19*(6), 1021-1037. doi: https://doi.org/10.1080/17486025.2024.2340117

[10]. Ghorbani, S., Bour, K., & Javdan, R. (2025). Applying the PROMETHEE II, WASPAS, and CoCoSo models for assessment of geotechnical hazards in TBM tunneling. *Scientific Reports*, *15*(1), 491. doi: https://doi.org/10.1038/s41598-024-84826-x

[11]. Zhang, D., Fang, Q., & Lou, H. (2014). Grouting techniques for the unfavorable geological conditions of Xiang'an subsea tunnel in China. *Journal of Rock Mechanics and Geotechnical Engineering*, *6*(5), 438-446. doi: https://doi.org/10.1016/j.jrmge.2014.07.005

[12]. Golian, M., Teshnizi, E. S., & Nakhaei, M. (2018). Prediction of water inflow to mechanized tunnels during tunnel-boring-machine advance using numerical simulation. *Hydrogeology Journal*, *26*(8). doi: https://doi.org/10.1007/s10040-018-1835-x

[13]. Li, S., He, P., Li, L., Shi, S., Zhang, Q., Zhang, J., & Hu, J. (2017). Gaussian process model of water inflow prediction in tunnel construction and its engineering applications. *Tunnelling and Underground Space Technology*, *69*, 155-161. doi: https://doi.org/10.1016/j.tust.2017.06.018

[14]. Liu, D., Xu, Q., Tang, Y., & Jian, Y. (2020). *Prediction of water inrush in long-lasting shutdown karst tunnels based on the HGWO-SVR model. IEEE Access* (Vol. 9). IEEE.

[15]. Mahmoodzadeh, A., Ghafourian, H., Mohammed, A. H., Rezaei, N., Ibrahim, H. H., & Rashidi, S. (2023). Predicting tunnel water inflow using a machine learning-based solution to improve tunnel construction safety. *Transportation Geotechnics*, *40*, 100978. doi: https://doi.org/10.1016/j.trgeo.2023.100978

[16]. Zhou, J., Zhang, Y., Li, C., Yong, W., Qiu, Y., Du, K., & Wang, S. (2023). Enhancing the performance of tunnel water inflow prediction using Random Forest optimized by Grey Wolf Optimizer. *Earth Science Informatics*, *16*(3), 2405-2420. doi: https://doi.org/10.1007/s12145-023-01042-3

[17]. Jahanmiri, S., Aalianvari, A., & Abbaszadeh, M. (2024). Developing GEP tree-based, Neuro-Swarm, and whale Optimization Models for evaluating Groundwater Seepage into Tunnels: A Case Study. *Journal of Mining and Environment*, *15*(4), 1409-1436. doi: https://doi.org/10.22044/jme.2024.13601.2513

[18]. Shen, Q., Yang, H., Zhou, Z., Chen, Z., & Zhang, Y. (2025). Simulation and parameter identification of water inrush in tunnel construction using physics-informed neural networks. *Bulletin of Engineering Geology and the Environment*, *84*(7), 370. doi: https://doi.org/10.1007/s10064-025-04381-1

[19]. Chen, J., Zhou, M., Zhang, D., Huang, H., & Zhang, F. (2021). Quantification of water inflow in rock tunnel faces via convolutional neural network approach. *Automation in Construction*, *123*, 103526. doi: https://doi.org/10.1016/j.autcon.2020.103526

[20]. Ju, S., Ou, G., Peng, T., Wang, Y., Song, Q., & Guan, P. (2025). Tunnel water inflow prediction using explainable machine learning and augmented partially missing dataset. *Frontiers in Earth Science*, *13*, 1590203. doi: https://doi.org/10.3389/feart.2025.1590203

[21]. Khatti, J., & Polat, B. Y. (2024). Assessment of short and long-term pozzolanic activity of natural pozzolans using machine learning approaches. *Structures, 68,* 107159. doi: https://doi.org/10.1016/j.istruc.2024.107159

[22]. Ghorbani, E., & Yagiz, S. (2024). Estimating the penetration rate of tunnel boring machines via gradient boosting algorithms. *Engineering Applications of Artificial Intelligence*, *136*, 108985. doi: https://doi.org/10.1016/j.engappai.2024.108985

[23]. Chen, T. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4).

[24]. Zhu, X., Chu, J., Wang, K., Wu, S., Yan, W., & Chiam, K. (2021). Prediction of rockhead using a hybrid N-XGBoost machine learning framework. *Journal of Rock Mechanics and Geotechnical Engineering*, *13*(6), 1231-1245. doi: https://doi.org/10.1016/j.jrmge.2021.06.012

[25]. Maleki, Z., Farhadian, H., & Nikvar-Hassani, A. (2021). Geological Hazard in Tunnelling: The Example of Gelas Water Conveyance Tunnel in Iran. *Quarterly Journal of Engineering Geology and Hydrogeology*, *54*(1), qjegh2019-114.

[26]. Farhadian, H., & Shahraki, F. B. (2024). Enhancing analytical methods for estimating water inflow to tunnels in the presence of discontinuity areas. *Environmental Earth Sciences*, *83*(10), 339. doi: https://doi.org/10.1007/s12665-024-11651-w

[27]. Hassanpour, J., Ghaedi Vanani, A. A., Rostami, J., & Cheshomi, A. (2016). Evaluation of common TBM performance prediction models based on field data from the second lot of Zagros water conveyance tunnel (ZWCT2). *Tunnelling and Underground Space Technology*, *52*, 147-156. doi: https://doi.org/10.1016/j.tust.2015.12.006

[28]. Farhadian, H., & Katibeh, H. (2017). New empirical model to evaluate groundwater flow into circular tunnel using multiple regression analysis. *International Journal of Mining Science and Technology*, *27*(3), 415-421. doi: https://doi.org/10.1016/j.ijmst.2017.03.005

[29]. Zhou, J., Qiu, Y., Khandelwal, M., Zhu, S., & Zhang, X. (2021). Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *International Journal of Rock Mechanics and Mining Sciences*, *145*, 104856. doi: https://doi.org/10.1016/j.ijrmms.2021.104856

[30]. Thamboo, J., Sathurshan, M., & Zahra, T. (2024). Reliable unit strength correlations to predict the compressive strength of grouted concrete masonry. *Materials and Structures*, *57*(7), 151. doi: https://doi.org/10.1617/s11527-024-02417-8

[31]. Alkayem, N. F., Shen, L., Mayya, A., Asteris, P. G., Fu, R., Di Luzio, G., Strauss, A., Cao, M. (2024). Prediction of concrete and FRC properties at high temperature using machine and deep learning: a review of recent advances and future perspectives. *Journal of Building Engineering*, *83*, 108369. doi: https://doi.org/10.1016/j.jobe.2023.108369

[32]. Kocak, B., Pınarcı, İ., Güvenç, U., & Kocak, Y. (2023). Prediction of compressive strengths of pumice-and diatomite-containing cement mortars with artificial intelligence-based applications. *Construction and Building Materials*, *385*, 131516. doi: https://doi.org/10.1016/j.conbuildmat.2023.131516

[33]. Fissha, Y., Khatti, J., Ikeda, H., Grover, K. S., Owada, N., Toriya, H., Adachi, T., Kawamura, Y. (2024). Predicting ground vibration during rock blasting using relevance vector machine improved with dual kernels and metaheuristic algorithms. *Scientific Reports*, *14*(1), 20026. doi: https://doi.org/10.1038/s41598-024-70939-w

[34]. Wu, X., Feng, Z., Liu, J., Chen, H., & Liu, Y. (2024). Predicting existing tunnel deformation from adjacent foundation pit construction using hybrid machine learning. *Automation in Construction*, *165*, 105516. doi: https://doi.org/10.1016/j.autcon.2024.105516

[35]. Ghorbani, S., & Bameri, A. (2025). Hybrid Machine Learning Models to Predict the Uniaxial Compressive Strength of Rocks Based on Non-Destructive Tests. *Transportation Infrastructure Geotechnology*, *12*(5), 148. doi: https://doi.org/10.1007/s40515-025-00603-x