Geomechanics and Engineering, *Vol. 37, No. 1 (2024) 65-72*
https://doi.org/10.12989/gae.2024.37.1.065

A gene expression programming-based model to predict water inflow into tunnels

Arsalan Mahmoodzadeh*[1], Hawkar Hashim Ibrahim[2], Laith R. Flaih[3], Abed Alanazi[4], Abdullah Alqahtani[4], Shtwai Alsubai[4], Nabil Ben Kahla[5] and Adil Hussein Mohammed[6]

[1]*IRO, Civil Engineering Department, University of Halabja, Halabja, 46018, Iraq*
[2]*Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, 44002 Erbil, Kurdistan Region, Iraq*
[3]*Department of Computer Science, Cihan University-Erbil, Kurdistan Region, Iraq*
[4]*Department of Computer Science, College of Computer Engineering and Sciences in Al-Kharj, Prince Sattam bin Abdulaziz University, P.O. Box 151, Al-Kharj 11942, Saudi Arabia*
[5]*Department of Civil Engineering, College of Engineering, King Khalid University, PO Box 394, Abha 61411, KSA*
[6]*Department of Communication and Computer Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq*

(*Received October 16, 2022, Revised December 22, 2023, Accepted March 18, 2024*)

Abstract. Water ingress poses a common and intricate geological hazard with profound implications for tunnel construction's speed and safety. The project's success hinges significantly on the precision of estimating water inflow during excavation, a critical factor in early-stage decision-making during conception and design. This article introduces an optimized model employing the gene expression programming (GEP) approach to forecast tunnel water inflow. The GEP model was refined by developing an equation that best aligns with predictive outcomes. The equation's outputs were compared with measured data and assessed against practical scenarios to validate its potential applicability in calculating tunnel water input. The optimized GEP model excelled in forecasting tunnel water inflow, outperforming alternative machine learning algorithms like SVR, GPR, DT, and KNN. This positions the GEP model as a leading choice for accurate and superior predictions. A state-of-the-art machine learning-based graphical user interface (GUI) was innovatively crafted for predicting and visualizing tunnel water inflow. This cutting-edge tool leverages ML algorithms, marking a substantial advancement in tunneling prediction technologies, providing accuracy and accessibility in water inflow projections.

Keywords: gene expression programming; graphical user interface; machine learning; tunneling, water inflow

1. Introduction

Examination of the ramifications stemming from environmental phenomena such as earthquakes, floods, wind, tsunami waves, underground water, landslides, and the like holds paramount significance within the realm of research (Cui *et al.* 2023a,b, Cui *et al.* 2024, Dai *et al.* 2023, Liu *et al.* 2023). Among these factors, underground water stands out prominently, wielding the potential to infiltrate subterranean structures like tunnels, thereby precipitating substantial financial and human tolls. Tunnel designers and builders in karst regions face unique challenges due to the prevalence of groundwater inflows during the building process. It is not uncommon for construction workers and their equipment to be washed away when an unexpectedly large amount of water rushes in through the header. Worldwide, there have been numerous water inflow incidents, each causing significant human and economic damage. A notable instance was the Yesanguan tunnel disaster along the Yichang-Wanzhou railway in China on August 5, 2007, as detailed by Jin *et al.* (2016). To manage groundwater effectively during tunnel construction, the key step is to evaluate the risk of water inflow before beginning excavation (Farhadian and Nikvar-Hassani 2019, Cheng *et al.* 2019). In tunneling projects, the best approach to address water inflow issues is by forecasting their occurrence and implementing measures to either prevent or manage them effectively. In this proposal, the most popular machine learning (ML) methods are used to make predictions about the amount of water that will flow into tunneling projects.

Although precise water inflow prediction during tunneling operations is impossible, in the past few years, there have been various efforts by researchers to approach this issue through the use of empirical, numerical, and analytical techniques (Farhadian and Katibeh 2017, Apaydin *et al.* 2019, Xie *et al.* 2019, Su *et al.* 2017, Holmy and Nilsen 2014, Golian *et al.* 2018).

There's hope that analytical procedures will become more time and cost-effective in the near future. Analytical methods, based on simplified hydrogeological parameters often represented as circular or rectangular cross-sections, struggle to offer accurate predictions of water inflow in more complex scenarios involving intricate hydrogeological elements, such as rock cracks. Tunnel water inflow problems could also be studied using computer models in different complex geological settings (Hwang and Lu 2007). In recent decades, numerous conceptual modeling techniques have been developed, encompassing approaches

*Corresponding author, Ph.D.
E-mail: arsalan.mahmoodzadeh@uoh.edu.iq

ISSN: 2005-307X (Print), 2092-6219 (Online)

like discrete fracture network models, analogous porous media models, and a range of hybrid models (Berkowitz 2002, Li *et al.* 2017). While previous models have employed Rock Failure Process Analysis (2D) and COMSOL software to evaluate the possibility of water inflow, we will focus on this method (Yao *et al.* 2012). However, it is quite challenging for numerical modeling to attain the genuine state due to the geological data included in the model.

The application of analytical and computational techniques to gauge tunnel water inflows frequently falters, hampered by the need to delineate hydrogeological assumptions and the tendency to oversimplify heterogeneous environments. Additionally, these methods might not adequately provide a quantitative correlation between water inrush and its influencing factors, or for risk assessment. In fact, a chaotic system may be able to account for a wide variety of geological conditions. Therefore, the use of conventional exploration techniques is inadequate for the detection of hydrogeological features (Li *et al.* 2017).

A range of stochastic mathematical approaches has been implemented in the fields of tunnel construction and coal mining for precise analysis of water inflow risks. These techniques include the attribute mathematical model, as described by Wang *et al*. (2012) and Li *et al*. (2013), the analytic hierarchy process highlighted by Ho and Ma (2018), and the fuzzy extension theory discussed by Li *et al.* (2015). The assessments using these methods, which involve weight factors and index criteria, have been widely recognized and accepted (Li *et al.* 2017).

ML approaches have shown potential in addressing a variety of engineering challenges (Shi *et al.* 2023, Zhao *et al.* 2023, Yin *et al.* 2023). ML techniques have shown significant promise in the area of water inflow prediction in tunnels. For instance, Li *et al*. (2017) utilized the Gaussian Process Regression (GPR) method for water inflow prediction, comparing its results with those from Artificial Neural Networks (ANN) and Support Vector Regression (SVR). They found GPR to offer more accurate predictions than SVR and ANN. However, there's a consideration that their method might be limited to the specific dataset they used, raising questions about its applicability to different data sets or tunneling projects. Furthermore, the limited data they used for testing might not suffice for all test scenarios, underscoring the importance of data quantity and quality in determining the effectiveness of a predictive model. In a related study, Mahmoodzadeh *et al.* (2021) assessed the performance of various ML techniques, including Long Short-Term Memory (LSTM), K-Nearest Neighbors (KNN), GPR, SVR, and Decision Trees (DT), for predicting water inflow in tunnels. They suggested that the LSTM model was the most effective for this purpose.

Without a doubt, as more and more studies are conducted in a certain area, advances are gradually eliminating the limits and weaknesses of earlier studies (Mahmoodzadeh *et al.* 2021). Since the use of ML techniques for predicting water inflow in tunnels is still in its infancy, there is room for improvement in existing models. As an example, the work of Li *et al*. (2017) is noteworthy as the first study on this subject, but it also contains several flaws that need to be fixed. For instance, they require just 12 data points to create accurate prediction models, with 6 for training and 6 for testing. However, if the model was not properly trained, it would not accurately forecast future outcomes.

In this research, the method of gene expression programming was employed for predicting water inflow into tunnels. Several reasons led us to choose the GEP algorithm for simulation in various fields:

- GEP is known for its versatility, as it can handle both regression and classification problems. This flexibility makes it applicable to a wide range of simulation scenarios.
- GEP is particularly well-suited for symbolic regression problems, where the goal is to find an equation that describes the relationship between input variables and the output. This is valuable in simulations where understanding the underlying mathematical relationships is essential.
- GEP employs an evolutionary algorithm, making it effective for optimization tasks and finding solutions in complex search spaces. This is advantageous in simulations where the optimal solution may not be immediately apparent.
- GEP can adapt to the characteristics of the dataset. Its ability to evolve the model's structure and parameters allows it to capture complex patterns in the data.
- GEP is well-suited for capturing nonlinear relationships within data. This is beneficial in simulations where linear models may not adequately represent the simulated system.
- GEP can automatically perform feature selection, identifying the most relevant input variables. This can be crucial in simulations with many input features, helping to simplify and improve the model's efficiency.
- GEP provides symbolic expressions as solutions, making the results more interpretable compared to some other ML algorithms. This can be important when the simulation results need to be understood and communicated to a non-expert audience.

This research makes several significant contributions:

- It incorporates an extensive dataset of 750 data points.
- The study encompasses data from thirteen different tunnels across varied geological environments to cover a broad spectrum of variables.
- There's an exploration of how GEP performs in predicting tunnel water inflow.
- The GEP model is refined to produce a predictive equation for water inflow. This equation is then evaluated in comparison with real-world applications.
- A detailed sensitivity analysis is conducted to examine the impact of varying input factors on the output of the equation.

Introducing this model aims to mitigate the current uncertainties in tunnel construction and establishes a groundwork for integrating ML techniques into tunnel design processes.

2. GEP

Influenced by the principles of natural selection, GEP harnesses evolutionary computation to produce both mathematical models and computer programs. Ferreira

(2002) and Mansouri *et al*. (2016) elaborate on the dataset GEP uses for crafting its tree-like solutions. GEP's objective is to advance traditional methods of feature transmission by enhancing genetic algorithms (GA) and genetic programming (GP). Similar to GA, GP forecasts chromosomal phenotypes using a structure akin to a tree, varying in size and length (Ferreira 2002). GEP, however, deviates from its forerunners by discarding the need for chromosomes to serve dual purposes. This change results in more efficient and durable chromosome performance over a range of genetic operators, surpassing the capabilities of the GP algorithm (Ferreira 2002). A notable achievement of GEP is its ability to exceed both the Replicator and Phenotype Thresholds, signifying its evolutionary advancement in early stages.

The GEP methodology kicks off by spawning a population of linear chromosomes prior to embarking on the task of establishing connections among variables a, b, and y. These chromosomes might feature genes containing any of the variables. Once the chromosomes are crafted and filled with content, the subsequent stage entails assessing the fitness of each chromosome within the current generation. This evaluation entails employing an expression tree (ET) to depict the chromosomes. The gene's phenotype is delineated by an ET, functioning akin to a protein within a typical cell (Ferreira 2002, Ferreira 2006). In this process, a function that appropriately expresses the relationship between several variables is considered. This function can be manipulated using algebraic operators (such as +, -, /, *) and a variety of functions (like exponential and trigonometric functions). It's crucial to conduct a comprehensive analysis of the variable relationships involved.

In Ferreira's method (Ferreira 2002), the construction of a mathematical equation or program starts with the random assembly of terminals and functions. This is done using Ferreira's effective language, Karva, for gene expression, leading to the creation of unique entities. The fitness of these entities is measured by comparing the calculated y value with the actual y values for various combinations of a and b. The closer these values align, the more accurate the equation becomes, signaling improved health. During this process, the effectiveness of each chromosome in the initial generation is assessed, and this assessment influences the score for the next generation. Notably, the most promising individual from each generation is automatically carried over to the next, eliminating the need for a separate selection phase (Ferreira 2002, Ferreira 2006). This approach evaluates the fitness scenario by deducting and recording the discrepancy between the estimated and actual y values for each set of a and b, with reduced discrepancy indicating enhanced equation accuracy.

In the GEP algorithm development, the process begins with a series of critical steps. These include the establishment of the fitness function, the clear definition of various terms and functions, a detailed analysis of the chromosome structure encompassing aspects like the number of generations, their length, and gene count, the identification of the relationship's objective, and the construction of the algorithm based on the characteristics of identified operators. For the formation of the next generation, the process relies on the genotype, which is the linear arrangement of the chromosomes, ensuring complete transmission of both functional and non-functional chromosomes. During mutation in the next generation, regions of a gene that were previously inactive might become active, allowing for possible alterations (Ferreira 2002, Ferreira 2006). The progression to the next generation in this process is akin to the operation of a Roulette Wheel. In this metaphorical scenario, chromosomes are selected at random with each turn of the wheel. However, the randomness is tempered by the tendency to select higher-rated chromosomes more frequently, creating a form of selective randomization. This evolutionary process ensures the transference of genes from one generation to the next while maintaining a stable chromosome count. Subsequent to this, the chromosomes undergo a transformation through a technique that strategically positions genetic operators on the same chromosomes. The algorithm continues to evolve by constantly creating and evaluating new generations, aiming to gradually zero in on the most effective equation. To optimize efficiency and prevent loss of data and processing resources, it's advisable to impose a limit on the number of iterations, especially if the algorithm fails to show improvement within a predetermined period (Faradonbeh *et al.* 2016).

3. Modeling procedure

The comprehensive analysis and modeling undertaken in this study follows a systematic approach:

I. **Dataset Preparation**

Upon assembling all necessary samples, they are categorized into two groups: one designated for training purposes and the other for testing. The training set serves as the foundation for constructing the predictive model, while the testing set is instrumental in assessing the model's efficacy. During the training phase, the model scrutinizes potential correlations between the system's inputs and outputs.

II. **Model Construction Process**

Currently, there exists a singular viable option for configuring the model's hyperparameters.

III. **Model Analysis**

To assess the predictive model's efficacy, a range of statistical evaluation metrics are used. After identifying the most suitable hyperparameter values for the model, the process is repeated. A model is considered optimal if it demonstrates sustained resilience and maintains high performance over time.

IV. **Results Analysis and Discussion**

Subsequently, the model's ability to align with reality is tested by comparing predicted outcomes with observed results. This phase involves a thorough examination and discussion of the obtained results.

4. Database preparation

Recent research conducted by Mahmoodzadeh *et al*.

Table 1 An overview of the database utilized for this study

		H (m)	h (m)	RQD (%)	W (m3/hm)	Water inflow (m3/h)
Training data	count	600	600	600	600	600
	mean	88.8	50.5	48.2	5.7	83.5
	std	68.9	56.2	20.3	4.0	25.7
	min	15.0	0.0	4.0	1.0	47.7
	25%	54.0	22.0	34.0	3.2	67.4
	50%	66.0	33.0	45.0	4.7	75.6
	75%	89.0	54.0	63.0	7.0	90.5
	max	350.0	312.0	99.0	33.6	254.0
Test data	count	150	150	150	150	150
	mean	92.1	52.3	50.0	5.6	82.9
	std	74.3	57.8	20.3	4.0	24.1
	min	22.0	0.0	6.0	1.4	50.8
	25%	53.3	21.0	35.3	3.2	66.3
	50%	65.0	34.0	53.5	4.6	75.3
	75%	96.0	54.0	65.0	6.4	88.0
	max	340.0	283.0	95.0	24.7	162.9

(2021) posits that an excess of parameters may be superfluous when predicting water inflow in tunnels. Consequently, forecasting water inflow becomes progressively intricate as the number of parameters increases. Noteworthy factors, such as tunnel depth (H), groundwater level (h1), Rock Quality Designation (RQD), and water yield property (W), are acknowledged as pivotal parameters influencing water inflow into tunnels. These parameters are meticulously adjusted based on insights from other researchers in the field. Extensive studies affirm that these four factors wield significant influence over water inflow into tunnels. Given the inherent challenge of unknown or elusive subsurface geological conditions, data scarcity persists as a perennial issue in tunneling. Moreover, the selection of data and input parameters in this study is contingent upon the availability of relevant information. A dataset comprising 750 samples was employed, with 80% (600 samples) allocated for training and 20% (150 samples) for testing. All samples were documented by tunneling experts during the construction of 13 road tunnels in Iran, where the drilling and blasting technique was employed. Within the context of all the tunnels under consideration, grappling with water inflow emerged as a primary construction challenge. Given the paper's specific emphasis on forecasting water inflow into tunnels, the dataset collection strategy honed in on sections of the tunnel paths where water inflow occurred. Typically, these locations featured rock masses ranging from jointed to crushed Shale, Limestone, and Sandstone formations. Notably, in these tunnels, particularly within the crushed zones, water inflow precipitated tunnel face collapses, resulting in substantial financial setbacks for the project. Table 1 provides a summary of the training and testing samples.

5. Data normalization

Data normalization is a fundamental task in machine learning, especially when dealing with input data characterized by varying dimensions and units. The normalization process is vital for accelerating model convergence and reducing errors throughout the training phase. In this study, we embrace the Min-Max normalization technique to standardize all data within the zero-to-one range (see Eq. (1)).

$$x_{(0,1)} = \frac{x - Min}{Max - Min} \tag{1}$$

Here, x denotes the raw data, $x_{(0,1)}$ represents its normalized counterpart, and *Min* and *Max* signify the minimum and maximum values across the entire dataset, encompassing both training and testing data. *Min–Max* normalization stands as a widely accepted and effective technique in the research literature. However, it's worth noting that re-normalization might be necessary when additional samples are introduced, leading to an expanded range of values. To address this, it is recommended to increase the *Min–Max* range significantly beyond the original data's *Min–Max* values. This ensures that any future data falls within the predefined *Min–Max* range.

6. Assessment methods for model performance

To evaluate the model's effectiveness, various statistical metrics were applied. These included the mean squared error (MSE), the coefficient of determination (R^2), the variance account for (VAF) values, and the mean absolute error (MAE).

$$R^2 = 1 - \frac{sum\ squared\ regression\ (SSR)}{sum\ of\ squares\ total\ (SST)} \tag{2}$$

$$\mathrm{MAE} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}|y_i - y_i'| \tag{3}$$

$$\mathrm{MSE} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}(y_i - y_i')^2\ (\tag{4}$$

$$\mathrm{VAF} = 1 - \left[\frac{var(y_i - y_i')}{var(y_i)}\right] \times 100\% \tag{5}$$

7. Model implementation

The study utilized the Jupyter Notebook environment, part of the Anaconda Navigator 3.7, to apply the GEP method. Anaconda, known for its no-cost, open-source nature, facilitates package management and deployment in Python for scientific computing. An Intel (R) Core (TM) i7-10750H CPU with a 2.60 GHz speed and 32 GB of RAM was employed for the computational tasks.To achieve the best outcomes, multiple rounds of GEP modeling, as outlined in Table 2, were carried out. Throughout the process, key parameters such as "population size (n_pop)" and "generation count (n_gen)" underwent systematic modifications. Finally, the top-performing models were

Table 2 GEP model configuration parameters

	Parameter	Description
General	Head	10
	No. of genes in each chromosome	3
	RNC array	7
	Linking	addition
	Fitness	mean squared error
	Function set	-, +, *, /, Sqrt (x), Pow (x,y), x^2, x^3, 1/x, $x^{1/3}$, Exp, sin, cos, tan, MSE, R^2
Genetic operators	IS transposition	0.2
	Inversion	1
	Mutation	1
	One–point recombination	0.4
	Gene recombination	0.3
	Two–point recombination	0.2
	RIS transposition	0.2
	Gene transposition	0.3
Numerical constants	Upper bound	+10
	Lower bound	-10
	Constants per each gene	3
	Data type	Floating-Point

chosen based on a comprehensive analysis of their outcomes and an evaluation of their accuracy using various statistical metrics.

8. Results and discussion

8.1 Analysis of results

In the course of this investigation, the GEP model underwent 44 runs to guarantee the attainment of optimal accuracy. Each iteration involved the adjustment of parameters, specifically n_pop and n_gen, with values ranging from minimal settings of 10 and 5, respectively, to the maximum explored values of 200 and 180. Ultimately, the greatest resilient model, equipped to predict water inflow within tunnels effectively, was identified. The resulting equation for the proposed model is as follows

$$\begin{aligned}\text{Water inflow} = {} & 0.40659625W^2 \\ & -1.4729596\text{RQD} + 11.595781\text{W} \\ & +0.253535676\sin(W) - 0.394398645\text{H} \\ & +0.21946735877W(W+\sin(W)+6) \\ & +2.275976\text{h1} + 0.12765886\tan(5) \\ & +\frac{0.43765583h1^2}{H} + \frac{8.87688852W^2}{H} \\ & +\frac{235.845464W}{H} - 1.66575\cos(H*RQD) \\ & +0.1206754\cos(6h1) \\ & +445.999836 \quad m^3/h\end{aligned} \tag{6}$$

Eq. (6) serves as a universal formula for predicting water inflow in tunnels created through the drilling and blasting technique. Within this equation, the incorporation of all four input parameters from the dataset is evident.

Table 3 Results of statistical criteria for training and testing phases in proposed models

Dataset	R^2	MAE [m^3/h]	MSE [$(m^3/h)^2$]	VAF [%]
Test	0.9779	4.8950	33.7213	97.6340
Training	0.9758	4.5562	28.4586	97.4266

Table 4 Statistical evaluation results for the 5-fold cross-validation step

R^2	MAE [m^3/h]	MSE [$(m^3/h)^2$]	VAF [%]
0.9536	6.3214	41.5721	94.2519

Table 5 Comparison of the GEP's performance with other ML models' performances to estimate water inflow

Method	R^2	MAE [m^3/h]	MSE [$(m^3/h)^2$]	VAF [%]
GEP	0.9779	4.8950	33.7213	97.6340
SVR	0.9321	5.1532	36.4321	96.3412
GPR	0.9283	6.2136	68.9043	96.2314
DT	0.8941	9.3407	89.9535	92.1424
KNN	0.8703	12.5388	100.3569	89.0639

The effectiveness of Eq. (6) was evaluated by applying it to the input variables from both the training and testing data sets. To gauge its accuracy, several statistical measures were employed, as outlined in Table 3. The outcomes unequivocally illustrate the model's precise prediction of water inflow.

Figs. 1 and 2 illustrate the results of the proposed model for both the test and training datasets, respectively. The variances between the observed and predicted values are marginal, with tunnel water inflow estimation errors of such magnitude deemed inconsequential. Consequently, the model can be confidently endorsed as a dependable method for predicting water inflow within tunnels constructed through the drilling and blasting technique.

To bolster confidence in the precision of the results and mitigate concerns like overfitting, we implement the 5-fold cross-validation method. This approach ensures a rigorous examination of the model's performance by partitioning the dataset into five subsets, iteratively using four for training and the remaining one for validation. This enhances the robustness of the outcomes and guards against potential biases, fostering a more reliable and generalizable assessment of the model's effectiveness. The 5-fold cross-validation results for the GEP model are provided in Table 4. These results show the high accuracy of the GEP model and its correct performance.

Table 5 presents a comprehensive comparison of the GEP model's performance with other leading ML models, including SVR, GPR, KNN, and DT, all evaluated on the test data points. The findings unequivocally highlight the supremacy of the optimized GEP model in this study, showcasing unparalleled performance and accuracy when juxtaposed with its algorithmic counterparts. The meticulous optimization undertaken has undoubtedly positioned the GEP model as the frontrunner, attaining the pinnacle of precision among the diverse array of ML approaches tested.

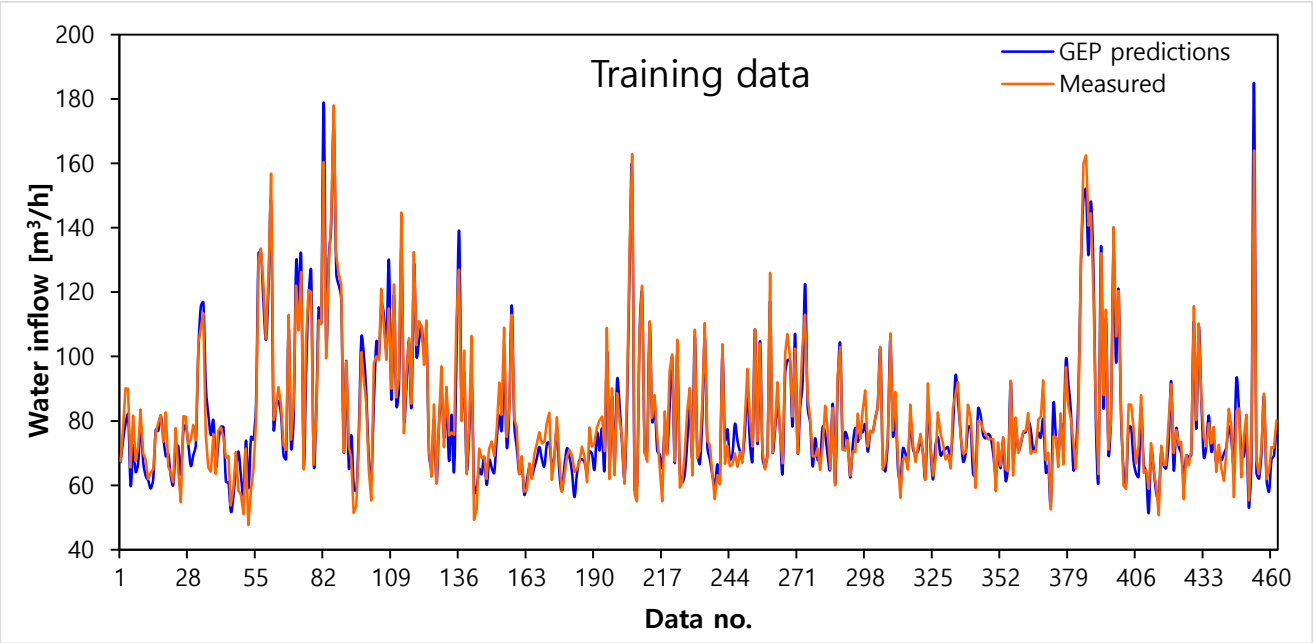

Fig. 1 Evaluating the accuracy of the proposed model's cost estimates by comparing them with the actual costs encountered during the training phase

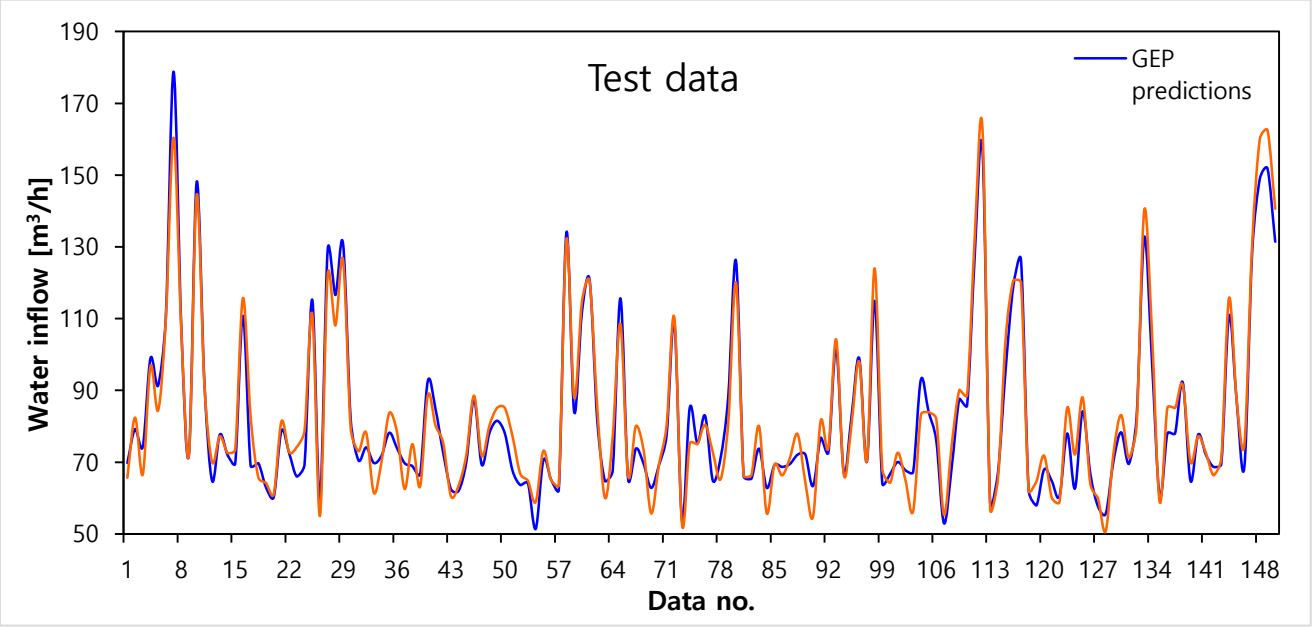

Fig. 2 Evaluating the costs calculated by the recommended model against the real costs encountered in the testing phase

8.2 Graphical user interphase (GUI)

An ML-based GUI is a pivotal tool that bridges the gap between sophisticated ML models and end-users, enabling a more widespread and effective application of ML techniques. In this groundbreaking study, a cutting-edge ML–based GUI has also been innovatively crafted to forecast and visualize water inflow projections for tunnels (see Fig. 3). This sophisticated GUI harnesses the power of ML algorithms to provide an intuitive and insightful tool for predicting water inflow, marking a significant leap forward in tunneling prediction technologies.

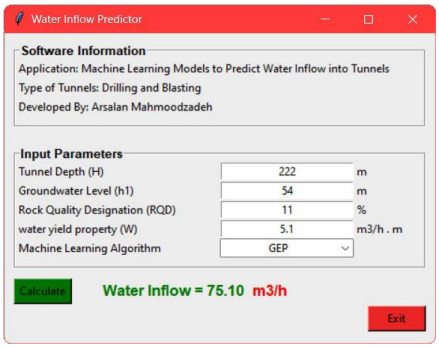

Fig. 3 A ML–based GUI to predict water inflow into tunnels

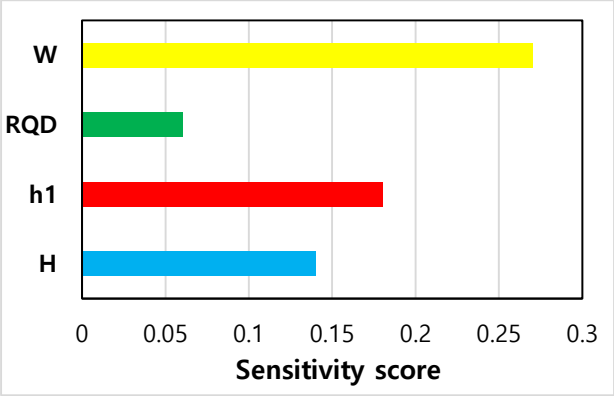

Fig. 4 The MI method was used to determine the impact of each input parameter's sensitivity score on the outcome of Eq. (6)

8.3 Sensitivity analysis

Facing a novel dataset can be overwhelming, especially when handed numerous features without a clear description. The initial challenge is figuring out where to begin. An effective method involves using a feature utility meter to assess the extent to which a feature aligns with a targeted objective. This ensures that efforts are focused on the most critical aspects, providing a clear direction without wasting time.

Our chosen statistical measure for this purpose is mutual information (MI). Unlike correlation, which reveals linear relationships only, MI uncovers connections in any form. Particularly beneficial in early-stage feature development when the model isn't predetermined, MI is versatile, capable of identifying various links, easy to employ, and quick. It operates based on the notion of uncertainty, measuring the extent to which one piece of information diminishes uncertainty when applied to another. Essentially, it addresses the question: If we possessed knowledge of a specific characteristic, to what degree could we enhance our confidence in attaining our objective?

The radar plot in Fig. 4 illustrates the computed sensitivity scores for each input parameter using the MI method. Notably, parameters W and h1 emerge with the highest ratings, indicating their substantial impact on the proposed model. Despite the substantial impact of RQD and H on water inflow in tunneling projects, Eq. (6) demonstrates a lower sensitivity to alterations in these parameters. The sensitivity of the prediction model's parameters is intrinsically linked to the nature of the data and its ranges, underscoring the influence of data characteristics on the sensitivity of the derived equation.

9. Conclusions

The conclusions drawn from this study based on the obtained results are as follows:

- The GEP model demonstrates exceptional proficiency in discerning relationships among various inputs and the volume of water inflow.
- Sensitivity analysis highlights the critical role of two parameters, n_pop and n_gen, in the GEP model's performance, necessitating careful tuning for a specific dataset. Through 44 independent runs of the GEP algorithm, diverse combinations of n_pop and n_gen were evaluated. Ultimately, the most robust model for predicting water inflow was identified.
- The GEP technique endeavors to discover the optimal alignment between observed data and each model, yields a corresponding equation. In this study, a comprehensive equation for the proposed model was formulated, incorporating four input variables to precisely predict water inflow into tunnels.
- Thorough evaluation of the suggested equation's capability to predict water inflow, utilizing various statistical indices, was conducted during training (R^2 = 0.9758; MAE = 4.5562 m^3/h; MSE = 28.4586 m^3/h; and VAF = 97.4266%) and testing (R^2 = 0.9779; MAE = 4.8950 m^3/h; MSE = 33.7213 $(m^3/h)^2$; and VAF = 97.6340%). The results underscore the potential effectiveness of the equation in predicting water inflow.
- The optimized GEP model's efficacy in forecasting water inflow into tunnels was meticulously assessed, including a robust comparison with alternative ML algorithms such as SVR, GPR, DT, and KNN. In this comprehensive evaluation, the GEP model distinctly outperformed its ML counterparts, showcasing superior predictive capabilities and accuracy in anticipating water inflow. This compelling demonstration positions the optimized GEP model as a frontrunner in the realm of tunnel water inflow prediction, attaining noteworthy superiority over its algorithmic peers.
- A state-of-the-art ML-driven GUI was ingeniously designed to predict and visually represent water inflow forecasts for tunnels. This cutting-edge GUI leverages ML algorithms to offer an intuitive and insightful tool, propelling tunneling prediction technologies to new heights. Its innovative approach marks a substantial advancement, providing a powerful platform for accurate and accessible water inflow projections in tunneling scenarios.
- During the sensitivity analysis conducted using the MI approach, parameters W and h1 surfaced as possessing the most substantial influence on the suggested equation.
- The paramount importance of this study lies in furnishing geotechnical engineers with the tools necessary to precisely predict water inflow into tunnels.

Acknowledgments

This study is supported via funding from Prince Satam bin Abdulaziz University project number (PSAU/2024/R/1445).

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through large group Research Project under grant number RGP2/130/44.

References

Apaydin, A., Korkmaz, N. and Ciftci, D. (2019), “Water inflow into tunnels: Assessment of the Gerede water transmission

tunnel (Turkey) with complex hydrogeology", *Q. J. Eng. Geol. Hydrogeol.*, **52**(3), 346-359. https://doi.org/10.1144/qjegh2017-125.

Berkowitz, B. (2002), "Characterizing flow and transport in fractured geological media: A review", *Adv. Water Resour.*, **25**(8-12), 861-884. https://doi.org/10.1016/S0309-1708(02)00042-8.

Cheng, P., Zhao, L., Li, Q., Li, L. and Zhang, S. (2019), "Water inflow prediction and grouting design for tunnel considering nonlinear hydraulic conductivity", *KSCE J. Civil Eng.*, **23**(9), 4132-4140. https://doi.org/10.1007/s12205-019-0306-9.

Cui, W., Caracoglia, L., Zhao, L. and Ge, Y. (2023a), "Examination of occurrence probability of vortex-induced vibration of long-span bridge decks by Fokker–Planck–Kolmogorov equation", *Struct. Saf.*, **105**, 102369. https://doi.org/10.1016/j.strusafe.2023.102369.

Cui, W., Zhao, L. and Ge, Y. (2023b), "Wind-induced buffeting cibration of long-span bridge considering geometric and aerodynamic nonlinearity based on reduced-order modeling", *J. Struct. Eng.*, **149**(11). https://doi.org/10.1061/JSENDH.STENG-11543.

Cui, W., Zhao, L., Ge, Y. and Xu, K. (2024), "A generalized van der Pol nonlinear model of vortex-induced vibrations of bridge decks with multistability", *Nonlinear Dynam.*, **112**(1), 259-272. https://doi.org/10.1007/s11071-023-09047-9.

Dai, Z., Li, X. and Lan, B. (2023a), "Three-dimensional modeling of Tsunami waves triggered by submarine landslides based on the smoothed particle hydrodynamics method", *J. Mar. Sci. Eng.*, **11**(10), 2015. https://doi.org/10.3390/jmse11102015.

Faradonbeh, R.S., Armaghani, D.J., Monjezi, M. and Mohamad, E.T. (2016), "Genetic programming and gene expression programming for flyrock assessment due to mine blasting", *Int. J. Rock Mech. Min. Sci.*, **88**, 254-264. https://doi.org/10.1016/j.ijrmms.2016.07.028.

Farhadian, H. and Katibeh, H. (2017), "New empirical model to evaluate groundwater flow into circular tunnel using multiple regression analysis", *Int. J. Min. Sci. Technol.*, **27**(3), 415-421. https://doi.org/10.1016/j.ijmst.2017.03.005.

Farhadian, H. andNikvar-Hassani, A. (2019), "Water flow into tunnels in discontinuous rock: a short critical review of the analytical solution of the art", *Bull. Eng. Geol. Environ.*, **78**(5), 3833-3849. https://doi.org/10.1007/s10064-018-1348-9.

Ferreira, C. (2002), "Gene Expression Programming in Problem Solving", In Soft Computing and Industry, 635-653. Springer London. https://doi.org/10.1007/978-1-4471-0123-9_54.

Ferreira, C. (2006), "Gene Expression Programming", 21, Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-32849-1.

Golian, M., Teshnizi, E.S. and Nakhaei, M. (2018), "Prediction of water inflow to mechanized tunnels during tunnel-boring-machine advance using numerical simulation", *Hydrogeol. J.*, **26**(8), 2827-2851. https://doi.org/10.1007/s10040-018-1835-x.

Ho, W. and Ma, X. (2018), "The state-of-the-art integrations and applications of the analytic hierarchy process", *Eur. J. Operat. Res.*, **267**(2), 399-414. https://doi.org/10.1016/j.ejor.2017.09.007.

Holmøy, K.H. and Nilsen, B. (2014), "Significance of geological parameters for predicting water inflow in hard rock tunnels", *Rock Mech. Rock Eng.*, **47**(3), 853-868. https://doi.org/10.1007/s00603-013-0384-9.

Hwang, J.H. and Lu, C.C. (2007), "A semi-analytical method for analyzing the tunnel water inflow", *Tunn. Undergr. Sp. Tech.*, **22**(1), 39-46. https://doi.org/10.1016/j.tust.2006.03.003.

Jin, X., Li, Y., Luo, Y. and Liu, H. (2016), "Prediction of city tunnel water inflow and its influence on overlain lakes in karst valley", *Environ. Earth Sci.*, **75**(16), 1162. https://doi.org/10.1007/s12665-016-5949-y.

Li, L., Lei, T., Li, S., Zhang, Q., Xu, Z., Shi, S. and Zhou, Z. (2015), "Risk assessment of water inrush in karst tunnels and software development", *Arabian J. Geosci.*, **8**(4), 1843-1854. https://doi.org/10.1007/s12517-014-1365-3.

Li, S., He, P., Li, L., Shi, S., Zhang, Q., Zhang, J. and Hu, J. (2017), "Gaussian process model of water inflow prediction in tunnel construction and its engineering applications", *Tunn. Undergr. Sp. Tech.*, **69**, 155-161. https://doi.org/10.1016/j.tust.2017.06.018.

Li, S., Zhou, Z., Li, L., Xu, Z., Zhang, Q. and Shi, S. (2013), "Risk assessment of water inrush in karst tunnels based on attribute synthetic evaluation system", *Tunn. Undergr. Sp. Tech.*, **38**, 50-58. https://doi.org/10.1016/j.tust.2013.05.001.

Liu, W., Liang, J. and Xu, T. (2023), "Tunnelling-induced ground deformation subjected to the behavior of tail grouting materials", *Tunn. Undergr. Sp. Tech.*, **140**, 105253. https://doi.org/10.1016/j.tust.2023.105253.

Mahmoodzadeh, A., Mohammadi, M., Noori, K.M.G., Khishe, M., Ibrahim, H.H., Ali, H.F.H. and Abdulhamid, S.N. (2021), "Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques", *Automat. Constr.*, **127**, 103719. https://doi.org/10.1016/j.autcon.2021.103719.

Mansouri, I., Hu, J. and Kisi, O. (2016), "Novel predictive model of the debonding strength for masonry members retrofitted with FRP", *Appl. Sci.*, **6**(11), 337. https://doi.org/10.3390/app6110337.

Su, K., Zhou, Y., Wu, H., Shi, C. and Zhou, L. (2017), "An analytical method for groundwater inflow into a drained circular tunnel", *Groundwater*, **55**(5), 712-721. https://doi.org/10.1111/gwat.12513.

Shi, M., Hu, W., Li, M., Zhang, J., Song, X. and Sun, W. (2023), "Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine", *Mech. Syst. Signal Pr.*, **188**, 110022. https://doi.org/10.1016/j.ymssp.2022.110022.

Wang, Y., Yang, W., Li, M. and Liu, X. (2012), "Risk assessment of floor water inrush in coal mines based on secondary fuzzy comprehensive evaluation", *Int. J. Rock Mech. Min. Sci.*, **52**, 50-55. https://doi.org/10.1016/j.ijrmms.2012.03.006.

Xie, H., Jiang, C., He, J. and Han, H. (2019), "Analytical solution for the steady-state Karst water inflow into a tunnel", *Geofluids*, **2019**, 1-9. https://doi.org/10.1155/2019/1756856.

Yao, B., Bai, H. and Zhang, B. (2012), "Numerical simulation on the risk of roof water inrush in Wuyang Coal Mine", I*nt. J. Min. Sci. Technol.*, **22**(2), 273-277. https://doi.org/10.1016/j.ijmst.2012.03.006

Yin, H., Wu, Q., Yin, S., Dong, S., Dai, Z. and Soltanian, M.R. (2023). "Predicting mine water inrush accidents based on water level anomalies of borehole groups using long short-term memory and isolation forest", *J. Hydrology*, **616**, 128813. https://doi.org/10.1016/j.jhydrol.2022.128813.

Zhao, N., Li, D.Q., Gu, S.X. and Du, W. (2024), "Analytical fragility relation for buried cast iron pipelines with lead-caulked joints based on machine learning algorithms", *Earthq. Spectra*, **40**(1), 566-583. https://doi.org/10.1177/87552930231209195.

CC