Journal of Mining and Environment (JME) Vol. 15, No. 4, 2024, 1409-1436

Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Developing GEP tree-based, Neuro-swarm, and whale optimization models for evaluating Groundwater Seepage into Tunnels: A Case Study

Shirin Jahanmiri, Ali Aalianvari*, and Maliheh Abbaszadeh

Department of Mining Engineering, University of Kashan, Kashan, Iran

Article Info

Received 13 September 2023

Received in Revised form 7 April 2024

Accepted 12 April 2024

Published online 12 April 2024

DOI: 10.22044/jme.2024.13601.2513

Keywords

Tunnel Seepage

Groundwater

Optimization

Meta-heuristic algorithms

Abstract

Groundwater inflow is a critical subject within the domains of hydrology, hydraulic engineering, hydrogeology, rock engineering, and related disciplines. Tunnels excavated below the groundwater table, in particular, face the inherent risk of groundwater seepage during both the excavation process and subsequent operational phases. Groundwater inflows, often perceived as rare geological hazards, can induce instability in the surrounding rock formations, leading to severe consequences such as injuries, fatalities, and substantial financial expenditures. The primary objective of this research is to explore the application of machine learning techniques to identify the most accurate method of forecasting tunnel water seepage. The prediction of water loss into the tunnel during the forecasting phase employed a tree equation based on gene expression programming (GEP). These results were compared with those obtained from a hybrid model comprising particle swarm optimization (PSO) and artificial neural networks (ANN). The Whale Optimization Algorithm (WOA) was selected and developed during the optimization phase. Upon contrasting the aforementioned methods, the Whale Optimization Algorithm demonstrated superior performance, precisely forecasting the volume of water lost into the tunnel with a correlation coefficient of 0.99. This underscores the effectiveness of advanced optimization techniques in enhancing the accuracy of groundwater inflow predictions and mitigating potential risks associated with tunneling activities.

1. Introduction

Hydrology, geotechnical engineering, structural geology, rock engineering, and other related disciplines all pay close attention to the entry of groundwater into tunnels [1]. Tunnels frequently experience groundwater input both during and after construction, especially those built below the water table [2]. These unpredictably occurring geological hazards have the potential to destabilize the nearby rocks and result in serious harm, including injuries, fatalities, and high financial consequences [3][4]. Groundwater conditions are crucial factors to be taken into consideration during both the construction and operation of tunnels. [5].

Accurate groundwater flow prediction and evaluation are crucial given the possible dangers and difficulties involved with groundwater input [6]. Many scholars have tried utilizing a number of techniques to precisely anticipate groundwater flow into tunnels despite the persistent difficulties [7]. However, a thorough review of these strategies has not yet been done, leaving potential for more study in this important field [8]. The movement of groundwater into rock tunnels has been the subject of several research [9]. There are really several methods to accomplish this. Analytical (including semi-analytical), experimental, and numerical methods are a few examples of this technique [10].

However, there are a number of potential reasons why it may be challenging to estimate groundwater flow into tunnels accurately [11].

Corresponding author: ali_aalianvari@kashanu.ac.ir (A. Aalianvari)

This is due to the complexity and variety of rock masses, which make it challenging to pinpoint their essential characteristics. Assumptions are often used to downplay important elements and rock environment features when selecting the optimal strategy or tactics for a given scenario, which is not always easy. Choosing the optimal strategy or tactics for a given scenario is not always simple [12]. Seepage studies have investigated how groundwater flows into subterranean constructions. Because of a lack of information, it has been impossible to compare and analyze all available techniques for keeping track of groundwater inputs in tunnels [13]. This study provides a comprehensive overview of recent research advancements aimed at addressing the challenges associated with the inflexibility of mechanized drilling, particularly with Tunnel Boring Machines (TBMs), in response to sudden changes in geological conditions. Consequently, it offers a concise summary of the latest research developments in this domain. The primary objective of the study is to explore various machine-learning algorithms capable of estimating the volume of water infiltrating tunnels during drilling and blasting activities.

In the domains of hydrology, geotechnical engineering, structural geology, rock engineering, and related fields, the infiltration of subterranean water into tunnels remains a significant concern, attracting considerable attention [14]. Groundwater flow poses a recurring challenge during tunnel construction and subsequent phases, particularly in the case of tunnels situated below the water table [15]. These unforeseeable geological risks give rise to floods that compromise the integrity of subterranean rock structures, resulting in severe consequences such as fatalities, injuries, and substantial economic losses [9]. Consequently, precise forecasting and assessment of groundwater flow into tunnels become imperative. Despite the persistent challenges associated with such predictions, numerous scholars have endeavored to address this issue using diverse methodologies [16]. However, a comprehensive examination of these varied strategies is notably lacking. The literature review underscores a multitude of investigations conducted over recent decades concerning the phenomenon of groundwater flow through rock tunnels [17]. Several approaches have been employed for this purpose, including analytical (including semi-analytical), experimental, and numerical methods. Nevertheless, accurate estimation of groundwater flow into tunnels proves to be a formidable task due to a multitude of potential factors [18]. The intricacies and diverse nature of rock masses present a formidable challenge in precisely determining their fundamental characteristics [19]. Consequently, assumptions are frequently employed to mitigate the complexity of essential elements and actual environmental features within the rock strata. Selecting the optimal strategy or tactics for a given scenario is a non-trivial task [20]. The focus of this research is to investigate the ingress of subterranean water into tunnels excavated within a rock environment. However, due to a lack of comprehensive information, a thorough comparison and analysis of all available techniques for monitoring groundwater inputs in tunnels have been elusive [21].

Various experts have delved into the dynamics of subsurface water flow into underground constructions, employing a spectrum of techniques to forecast water seepage into tunnels. For instance, Jiang et al. [22] applied the Conformal Mapping Technique to compute the seepage in deep circular tunnels with grout. Maleki et al. [23] utilized the stochastic discontinuous technique to investigate the uncertainty associated with groundwater influx into subterranean excavations, accounting for geological, hydraulic, and tunnel-related factors. Ying et al. [7] employed a numerical model in the COMSOL program to validate the semi-analytical solution for predicting ground seepage into tunnels, meticulously considering factors such as tunnel burial depth, lining thickness, and lining permeability.

Some studies conducted in the field of predicting groundwater seepage into tunnels in recent years are given in the table 1.

Table 1. Some studies conducted in the field of predicting groundwater seepage into tunnels in recent years

Remarks	Specificity and Applicability	methods	Reference
Variable spatial distribution of groundwater flows in the media is considered. Complex modeling. Limited as hydraulic aperture of fractures not fully reached.	Real porous media. Large fractures sparsely distributed. Fractures permeability greater than that of rock mass. Simulate groundwater movements in the fractures.	Discrete Fracture Network (DFN)	[24]–[26]
Very limited due to the non-consideration of the real properties of the media.	Equivalent Porous media, seepage flow through fractured rock. Darcy flow	Equivalent Continuous Model (ECM)	[24]
Variable geotechnical and hydrogeological conditions	Modeling groundwater inflows into tunnels. Continuous media	Finite Element Method (FEM)	[27]
The dimensionality of the studied problem is reduced. Domain problem is changed to boundary problem.	Analysis and Description of groundwater flow. Isotropic and anisotropic Porous Media. Darcy flow.	Boundary Element Method (BEM)	[5]
Good representation of fractures in 3D. Direct treatment of the non-linearity behavior of materials.	Simulation of stress-flow coupling. Hydro mechanical properties of discontinuous rocks can be derived by equivalence.	Distinct Element Method (DEM)	[28]
Hydro mechanical properties of tunnels surrounding rocks are required. FLAC can be used alone, or coupled to mechanical modeling for interactions of fluid-media.	Simulation of groundwater inflows or bursting in subsurface tunnels or mine in homogeneous media. Darcy's flow regime is adopted.	FLAC 2D / FLAC 3D	[13][29], [30]
Hydraulic conductivity, Hydraulic head, and others relevant hydrogeological data are needed.	Prediction of groundwater inflows into shallow and deep Tunnels. Porous media. Laminar flow	MODFLOW	[14], [29], [30]
Tunnel diameter, Reynolds Number, Permeability, Sinuosity as requirements for the CFB	Simulation of groundwater inflows into Conveyance Tunnels in heterogeneous media. Laminar and Turbulent flow.	Conduit Flow Process (CFP) and adapted MODFLOW	[18]
Geological and hydrogeological features of the areas are required for the analysis. RFPA is based on Finite Element Method (FEM).	Prediction of groundwater outburst in underground mine. Heterogeneous media and fractured zones. Darcy flow adopted	Rock Failure Process Analysis code (RFPA), 2D	[31]
Hydraulic Conductivity and Volumetric Water Content are required.	Simulation of groundwater inflow into tunnels in saturated and unsaturated zones. Confined or unconfined aquifers. Flow regime can be Steady or Transient.	SEEP/W	[27]
Hydraulic head, tunnel radius and joint spacing are required for optimum accuracy.	Computation of groundwater inflows rate in discontinuous media. Laminar flow	Universal Distinct Element Code (UDEC), 2D	[32]
Hydro mechanical properties of surrounding rocks are required.	Computation of groundwater inflow into tunnels and mines in both saturated and unsaturated discontinuous media. Darcy flow	COMSOL Multiphysics	[2], [33]

As stipulated earlier, diverse approaches have been employed for the anticipation of water seepage occurrences within tunnel structures. In light of the escalating utilization of machine learning methodologies and their applicability in forecasting various parameters, there exists a discernible need for the exploration and implementation of novel prediction and optimization techniques specifically tailored for forecasting water seepage in tunnels. A number of studies conducted in the field of predicting water seepage in tunnels are given in the table 2.

Table 2. Some studies conducted in the field of predicting water seepage in tunnels using machine learning in the last few years.

Machine Learning methods	Capabilities and Applicability	Remarks	Ref
Gaussian Process Regression (GPR)	Groundwater inflows quantification into tunnels built in heterogeneous media, based on basic evaluation index and the associated criteria. Maximum Performance of inflows $R2$= 0.9956	No need to consider the relationship between hydrogeological features and water discharge rate. Large amounts of statistical data are required to obtain accurate results.	[34]
Support Vector Machine (SVM)	Prediction of groundwater inflows into tunnels built in karst and faults zones. Maximum Performance of inflows: $R2$ = 0.9767	Relevant hydrogeological properties of the concerned media and the depth of tunnels are required.	[34]
Convolutional Neural Network (CNN)	Prediction of groundwater inflow information in rock tunnels face.	Classification of RMR-based groundwater inflow image datasets based, and associated segmentations.	[35]
BP Neural Network	Prediction of groundwater inrush risk in karst tunnels using relevant factors	Hydrogeological factors and engineering factors could be combined for the prediction.	[36]
Artificial Neural Network (ANN)	Prediction of groundwater inflows into tunnels. Maximum Performance of inflows: $R2$= 0.8331	Relevant hydrogeological properties of the media and tunnels depth are necessary.	[34]
Bayesian Network (BN) & GIS	Water inrush prediction in coal mine located in faults areas. The accuracy of the prediction is about 83.4%.	BN used a graphical network of probabilistic rationale. GIS is coupled to BN for water inrush quantification, and for encroachment analysis. Relevant features of openings are required.	[37]
Long short-term memory (LSTM)	Groundwater prediction in tunnels excavated by DB. Performance: $R2$= 0.9866	Data: Tunnel depth, groundwater level, Rock Quality Designation, and Water yield property.	[38]
Deep Neural Networks (DNN)	Groundwater prediction in tunnels excavated by Drill-and-Blast. Performance: $R2$= 0.9815	Data: Tunnel depth, groundwater level, Rock Quality Designation, and Water yield property.	[38]
K-nearest neighbors (KNN)	Groundwater prediction in tunnels excavated by Drill-and-Blast. Performance: $R2$= 0.7665	Data: Tunnel depth, groundwater level, Rock Quality Designation, and Water yield property.	[38]
Decision Trees (DT)	Groundwater prediction in tunnels executed by DB. Performance: $R2$= 0.721	Data: Tunnel depth, groundwater level, Rock Quality Designation, and Water yield property.	[38]
Integrated model (VMD, ORELM, MOGWO)	Groundwater inflows prediction into deep mines. Prediction Performance: $R2$= 0.9685	Procuration of water inflow series by VMD, Prediction of components by ORELM, Optimization by MOGWO.	[39]
Hybrid model (HGWO-SVR)	Prediction of water in rush into Karts Tunnels. Transport Tunnels Model Performance: $R2$= 0.99953	Appropriated Rainfall data are required. HGWO algorithm optimizes SVR parameters.	[26]

In prior studies, a singular method and its hybrid counterparts were commonly employed, while a comprehensive comparison among various machine learning methods was often lacking. In this current investigation, a distinctive approach has been adopted, incorporating three hybrid neural network methods: the particle swarm algorithm, the gene expression algorithm method, and the optimization of the selected method using the Whale Algorithm. Notably, the Whale Algorithm stands out as one of the latest optimization techniques in recent years. This study effectively demonstrates the utility and efficiency of this algorithm in predicting water seepage into the tunnel. The following is a brief overview of the methods employed in this study. Particle Swarm Algorithm (PSO)[40] leverages the principles of swarm intelligence, where a population of particles collaboratively explores the solution space to optimize the neural network parameters for predicting water seepage. Gene Expression Algorithm Method (GEP)[41] involves the evolution of computer programs using genetic algorithms. It has been applied to formulate a tree equation that captures the underlying principles of water seepage into the tunnel. Introduced as one of the latest optimization methods, the Whale Algorithm is utilized in this study to fine-tune and optimize the selected neural network method. This algorithm has shown promise in enhancing the overall predictive accuracy for water seepage [42]. The combined application of these methods represents a novel and comprehensive approach, showcasing the potential of advanced hybrid neural network techniques and the efficacy of the Whale Algorithm in the specific context of predicting water seepage into tunnels.

In the present study, a repertoire of machine learning techniques has been harnessed, including the amalgamated neural network algorithm, the particle swarm algorithm, the gene expression algorithm, and the optimization facilitated by the innovative whale algorithm. The distinctiveness of this study lies in its concurrent deployment of multiple machine learning techniques, coupled with their optimization and a thorough inter-method comparison. This multifaceted approach enhances the robustness and comprehensiveness of the predictive modeling employed in the investigation.

2. Study area

Amir Kabir tunnel has been conceived and is currently under implementation with the primary objective of conveying drinking water from Amir Kabir dam to Tehran. The tunnel spans an approximate length of 30 km, divided into two distinct sections. In the initial phase of implementing the first segment, extending from ET to K', two primary tunnel route options were considered. Ultimately, the arc route was chosen for various reasons, prominently influenced by geological considerations, particularly the need to circumvent the Porkan-Verdij fault. The tunnel route plan for the ET-K' section is visually represented in Figure 1, encompassing a length of 15,980 meters and executed through the utilization of a full-section Tunnel boring machine. This study is dedicated to the scrutiny of the groundwater inflow in the initial segment of Amir Kabir tunnel, employing the SGR method in the context of tunnel construction and assessing the associated risk of underground water seepage. Both qualitative and quantitative aspects of groundwater seepage are thoroughly examined and discussed.

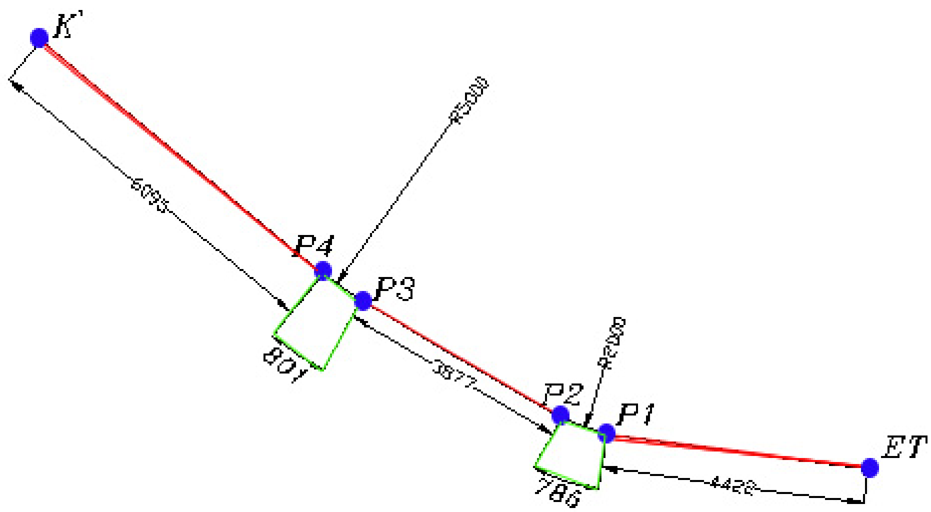


Figure 1. Tunnel route plan (Section ST-K') - Total length 15980 meters

It is possible to forecast that the studied region would encounter tectonic issues caused by inverted fault activity because of its geostructural location in the central Alborz zone. The area that has been studied geologically stretches from the heights east of Karaj to the highlands west of the Olympic Village and north of the railroad town (west of Tehran). The exact location of the tunnel is shown in Figure 2.

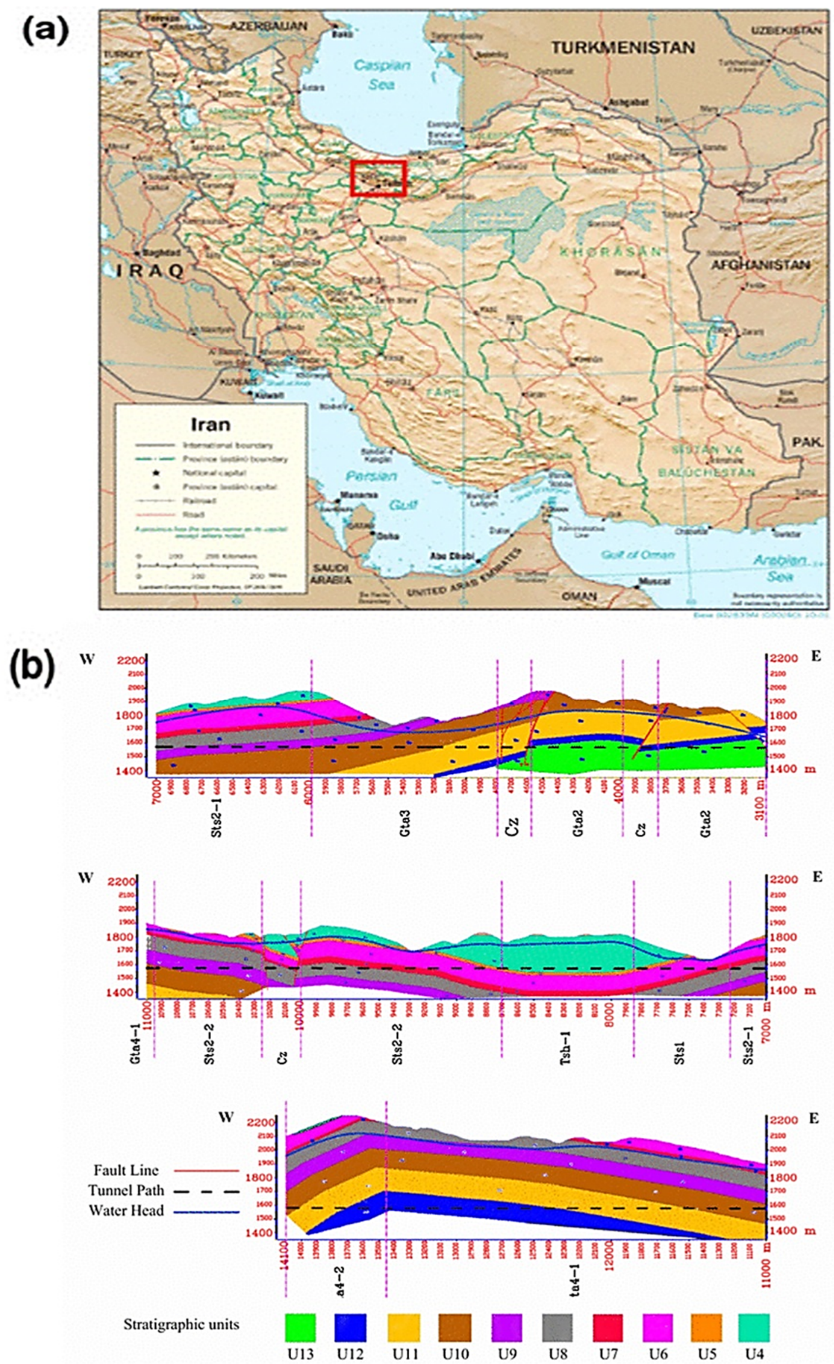


Figure 2. Location (a) and geological section (b) of Amir Kabir tunnel path [43]

Based on the investigations conducted in the geological visits of the region and the information obtained from the geotechnical operations (exploratory guesses and laboratory tests), regardless of the sediments and quaternary deposits, there are a total of 8 lithological types in the route. Amir Kabir tunnel can be identified and separated from each other from 1.3 to 1.14 kilometers [43] . These lithological types are introduced in Table 3. The boundary of these lithological types only in some cases coincides with the boundary of the stratigraphic units and in most cases the geotechnical characteristics of the units have been the factor of separation of lithological types.

Table 3. Lithological types identified in the tunnel route

Row	Lithological species	Groundwater conditions	Stability Category / Description
1	Gta2	Moist to wet	Fairly strong, thick layered, fairly crushed, stable
2	Gta3	Moist to wet	Fairly strong, thick layered, fairly crushed, stable
3	Gta4	Moist to wet	Strong, thick layering, stable
4	Sts1	Flow locally	Weak to somewhat strong, thin to slightly thick layers, crushed
5	Sts2	Moist to wet	Strong, thick layering, stable
6	Tsh	Moist	Crushed, thin to medium layers, weak to moderately strong
7	Cz	Flow	Very weak, faulted and crushed rocks, unstable

3. Materials and methods

The workflow for predicting water seepage into the tunnel is delineated in the accompanying figure. The initial phase involves the collection of data pertinent to Amir Kabir tunnel, followed by a comprehensive analysis of the gathered information. In the prediction stage, the principles governing seepage into the tunnel are computed using a tree equation formulated through gene expression programming (GEP). The ensuing results are then juxtaposed with the outcomes derived from a hybrid artificial neural network (ANN) augmented with particle swarm optimization (PSO). Moving on to the optimization phase, the Whale Optimization Algorithm (WOA) is both chosen and developed to enhance the efficiency of the modeling process.

3.1. Data set

The collected data related to Amir Kabir tunnel included 448 data, 4 parameters were considered as the input parameter and the input flow to the tunnel as the output parameter of the tunnel. In Table 4, the variables used to predict water loss into and out of Amir Kabir tunnel are presented.

Table 4. Estimating variables for water loss into and out of Amir Kabir tunnel

Type	Parameters	Unit	Min	max
Model input parameters	Equivalent Permeability	Keq(Lu)	0.500	23.450
	Head of water above tunnel	(m)	55.000	535.000
	R tunnel	(m)	2.350	2.350
	Overburden	(m)	65.000	660.000
Model output parameter	Q	(lit/s)	0.008	0.023

Figure 3 shows correlation analysis and data matrix.

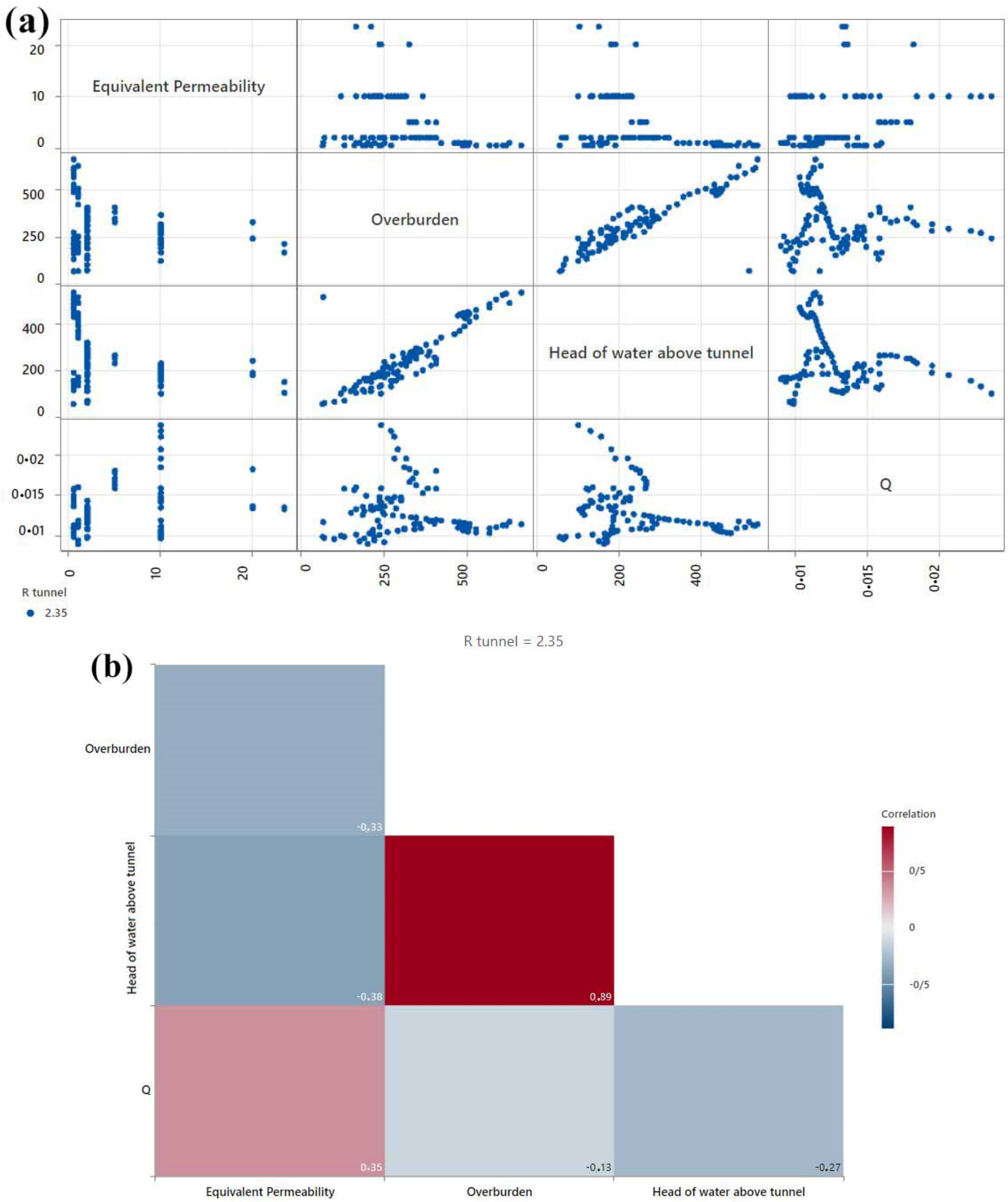


Figure 3. (a) Correlation analysis and (b) data matrix

3.2. Methods

3.2.1 Artificial Neural Network (ANN)

Donald Hebb initially established the idea of neural networks in the 1950s by presenting a straightforward learning process [44]. He developed this technique by researching how learning affects human brain neurons [45]. Each ANN neuron receives information from the preceding neuron through its dendrites [46], which are then processed before being sent by the axons to the subsequent section, or neuron. All neurons are connected in a layered architecture, where the mapping between inputs and outputs is conducted using the following formula[45], [47]:

$$h_i = \max(0, W_i . h_{i-1} + b_i) for 1 \le i \le L, and . h_0 = x$$
$$y = \max(0, V . h_L) \quad (1)$$

Where L is the number of layers, matrices W_1; . . . W_L; V and vector b_1; . . . b_L are model parameters learned from the dataset.

Through synapses, chemical signaling occurs between cells. The behavior of a computer neuron used in neural networks, given a sigmoid activation function, is akin to that of a real neuron with inputs and outputs [48] [49]. Each of an ANN's layers, which might be two or more, has a number of neurons [50]. The weights of a network are correlated with the strength of linkages between layers [47]. Each neuron's associated weights linearly alter the input vectors, which serve as the arguments for the nonlinear activation function (transfer function) of each neuron. In neural networks, back-propagation (BP) and multilayer pre-propagation (MLPP) are the two major techniques [3], [51], [53]. The weights are updated using this approach to ensure that the loss functions produce the least amount of mistake (loss) possible. To meet the termination criteria, this training procedure is performed numerous times [54]. The term BP refers to the circumstance in which the gradient for nonlinear multilayer networks (the networks used to solve the bulk of engineering problems) is calculated[55]. The sigmoid transfer function accepts the input values and displays them as a 0–1 interval regardless of the starting input interval [56]. Figure 4 shows the overview of the ANN network used in this study.

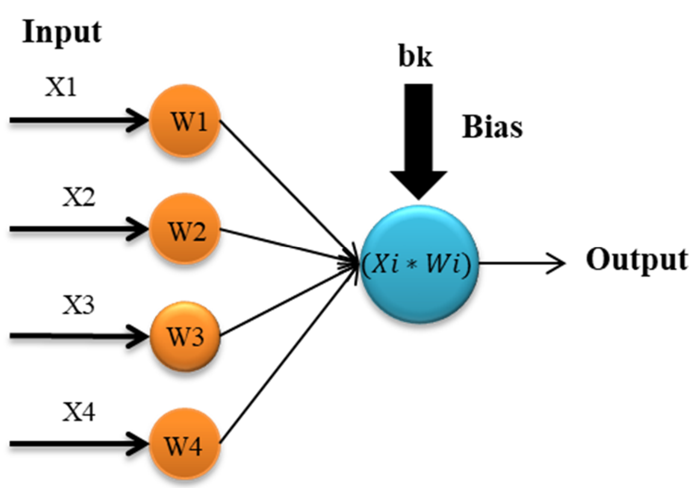


Figure 4. ANN that used in this study

3.2.2. Particle Swarm Optimization (PSO) Based on the ANN

Particle swarm optimization (PSO), a method for optimum continuous problems, was first proposed by Kennedy and Eberhart [57] . PSO is a nonlinear method that draws its inspiration from social systems like schools of fish [58]. In actuality, PSO is dependent on the quantity of randomly created particles [59]. Another phase in the iterative process of PSO is the search for an optimal value goal. At this point, the particles modify their location in response to their own and other particles' experiences [60], [61]. In order to reach the ideal position, each particle follows its personal best position (PBEST) as well as the collective best position (GBEST) among other particles [62], [63]. Each particle tends to move toward its PBEST as well as GBEST throughout the training process, based on a new velocity term and the distance between its best positions throughout the learning phase [64], [65]. Each particle's new position in the subsequent iteration is influenced by the new velocity value [59], [66]. The position-update formula for particles used in this paper is [67]:

$$V_i^{t+1} = wV_i^t + c_1r_1(p_{best,i}^t - X_i^t + c_2r_2(g_{best,i}^t - X_i^t)X_i^{t+1} = X_i^t + V_i^{t+1} \quad (2)$$

Where V_i^{t+1} and V_i^t represents the velocity of particle i at iteration t and t þ 1; X_i^{t+1} and X_i^t represent positions of particle i; w, c1 and c2 are the inertia parameter, the cognitive influence parameter and the social influence parameter; r1 and r2 are random values between 0 and 1; $p_{best,i}^t$ and $g_{best,i}^t$ represent the best position of a particle.

The PSO-ANN model's prediction process is depicted in Figure 5 [68], [69].

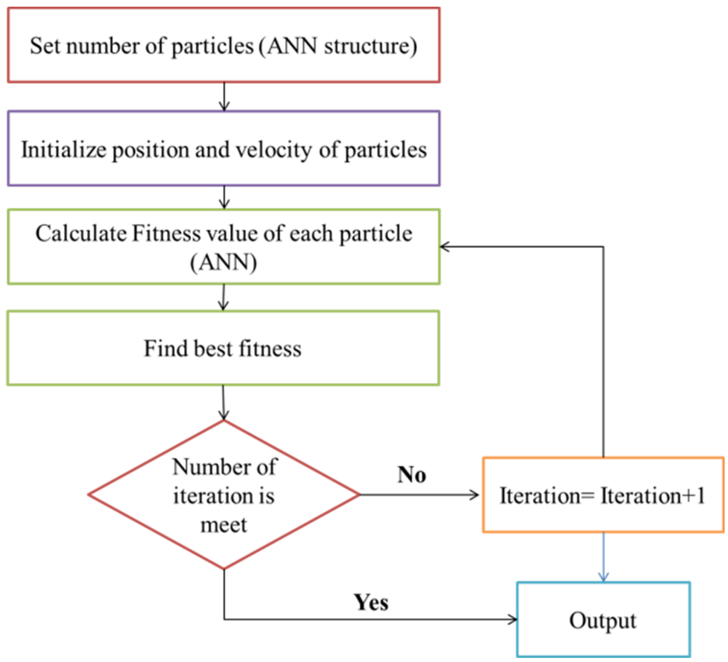


Figure 5. Flow chart of PSO-ANN's model [68], [69]

3.2.3 Gene expression programming algorithm (GEP)

Gene expression programming algorithm refers to one of the state-of-the-art methods developed in the field of artificial intelligence [70]. It is actually a more developed version of GA and GP [71]. GEP, which is made up of many components, offers appropriate answers for various issues [72]. The expression tree utilized in this method, which employs two primary chromosomes, demonstrates its ability to overcome the constraints of the preceding two (GA and GP) [45], [50]. In GEP, encodings are frequently represented by strings that were written in the Karva programming language can behave in an alien-like manner [73]. It is intriguing that GEP's models, which build connections between dependent and independent components, may be represented by mathematical equations. In the field of engineering, the creation of models that can generate equations is essential and valuable[74]. Such techniques are effective alternatives to ANN models for problem-solving [75]. The experts in this sector have been prompted by these problems to improve these techniques. Numerous mathematical operations, such as, +,-, sin, etc., are expressed and applied to variables in GP [76]. In order to analyze the issue, they can be combined to form a mathematical set. Each gene on chromosomes with many genes depicts a sub-ET with a head and a tail [77]. Such symbolic chromosomes ought to be represented as variously sized and shaped trees, or "expression trees" [54]. These points are examined in light of the models' control mechanisms and the degree of compatibility between them [78]. There are several varieties of these functions, and each type may be specified using various standards. The root mean square error (RMSE), mean absolute error (MAE), and root relative square error (RRSE) are some of these functions [79]. The best chromosomes chosen by the roulette wheel approach for the first process are put into the next structure in the event that the termination requirement is not satisfied (i.e., achieving the maximum repetition or adequate fitness value) [80]. Then, according to the ratio of the existing chromosomes, the three most important genetic operators—mutation, transfer (RIS, IS, and gene transfer), and reconstruction (one point, two points, and gene reconstruction)—are employed [77]. What does GP stand for? Thus, the remaining chromosomes are replaced with new ones, and this process continues until all of the conditions for termination have been met. Figure 6 depicts the flowchart for the gene expression programming strategy.

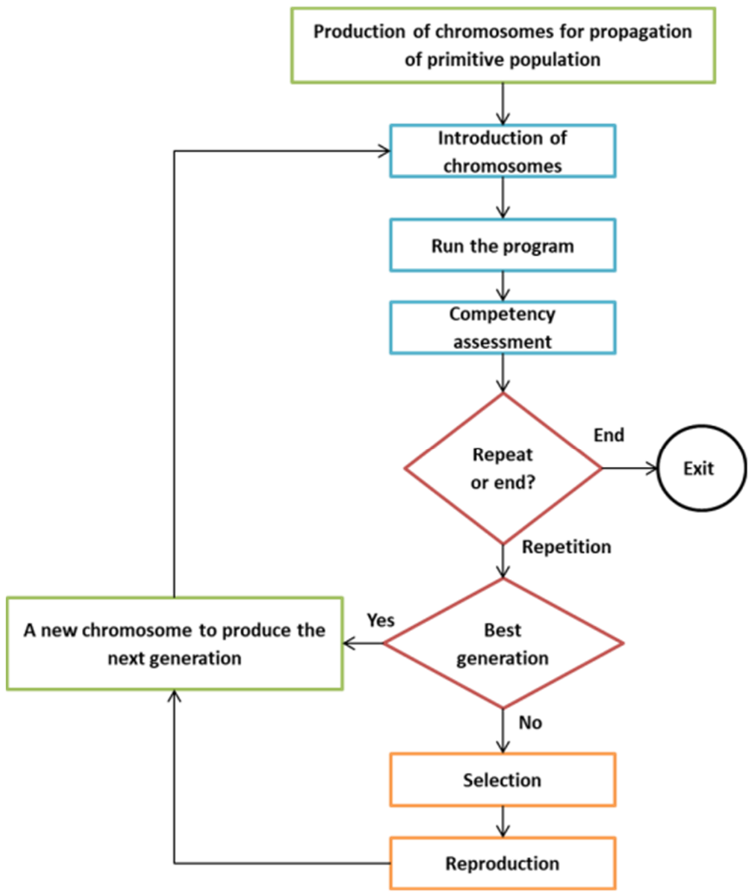
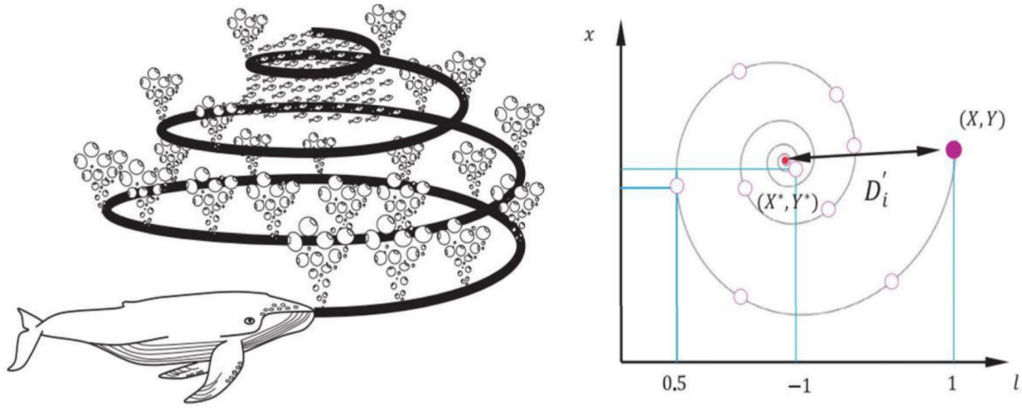


Figure 6. Flowchart of gene expression programming algorithm [77]

3.2.4 Whale optimization algorithm (WOA) Based on the GEP

The whale optimization algorithm, a ground-breaking stochastic optimization method, was developed in 2016 by Mirjalili and Lewis [81]. The two main phases of humpback whales' hunting behavior (exploration and exploitation) were the inspiration for this algorithm [82]. When they locate their meal, humpback whales hunt for tiny schools of fish and then dive 10 to 15 meters below the surface to force the fish to the top [46], [83]. In order to concentrate the fish and surround them, the whales release a large number of spiral air bubbles, which imprison the fish and prevent them from swimming [84]. The whale commander finally sends out a signal to start the attack [85]. The term "bubble web feeding behavior" refers to their cunning method of detecting and pursuing prey. Humpback whale behavior during the design process is depicted in Figure 7 and the trend view of the whale optimization method is shown in Figure 8.



(a) (b)

Figure 7. Humpback whale activity during WOA design [85]

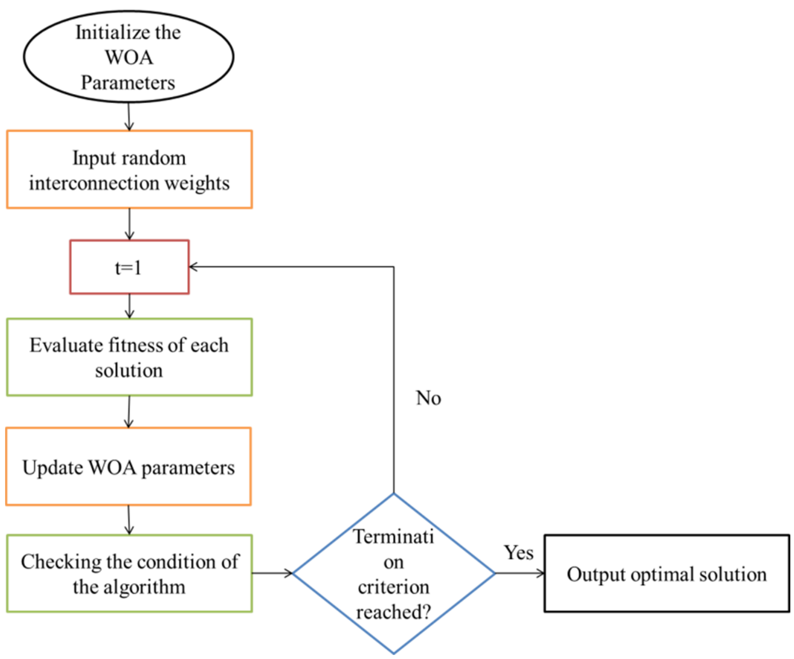


Figure 8. An illustration of the whale optimization technique [84]

The mathematical representation of WOA is established using this approach as follows:

a) The method of shrinking encirclement:

The whale uses the following equations to update its location as it circles the prey during the hunting mechanism in order to hunt its target [46].

$$\vec{D} = \left|\vec{C}.\vec{X}^*(g) - \vec{X}(g)\right| \tag{3}$$

$$\vec{X}(g+1) = \vec{X}^*(g) - \vec{A}\vec{D} \tag{4}$$

Where the distance between the humpback whale and its prey is represented by $\vec{D}$, $\vec{X}$ denotes the position vector, and $\vec{X}^*$ indicates the position vector of the optimal solution obtained until the iteration time. In each iteration, if there is a recovery solution, $\vec{X}^*$ should be updated. The vectors A* and C* are calculated as follows:

$$\vec{A} = 2\vec{a}\vec{r} - \vec{a} \tag{5}$$

$$\vec{C} = 2\vec{r} \tag{6}$$

Where a* decreases linearly from 2 to 0, and r* is a random vector in the range [0, 1].

b) Helical update position:

The WOA conducts a spiral equation between the prey location and the sperm whales' position to replicate the latter in their spiral motion as the whales' spiral toward the target prey to update their position. The following is the spiral equation in mathematics:

$$\vec{X}(g+1)=\vec{D}_2 . e^{bl} . \mathrm{Cos}(2\pi l)+\vec{X}^*(g) \tag{7}$$

Where D2 is the distance that separates the whale as well as its prey, which is first determined using the following equation: where b stands for a constant that characterizes the form of the logarithmic spiral, l is an integer randomly selected from the range [1, 1], and

$$\vec{D}=\left|\vec{X}^*(g)-\vec{X}(g)\right| \tag{8}$$

WOA starts by producing an initial set of search agents at random, just like other evolutionary algorithms. The algorithm employs the probability coefficient (p) to decide between the long-distance and logarithmic pathways under the assumption that the likelihood of the contraction and helix mechanisms are equal during the optimization phase. The following is how this idea is quantitatively modeled [84]:

$$\vec{X}(g+1)=\begin{Bmatrix}\vec{X}^*(g)-\vec{A}.\vec{D} \to \forall P \prec 0.5\\ \vec{D}_2 . e^{bl} . \mathrm{Cos}(2\pi l)+\vec{X}^*(g) \to \forall P \geq 0.5\end{Bmatrix} \tag{9}$$

Instead of depending on the best search agent X found so far, WOA changes the search agent's location at random if the random value of A is larger than 1. The following equations are used to mathematically represent this strategy, which is used to guarantee a worldwide search [84]:

$$\vec{D}=\left|C.\vec{X}_{rand}-\vec{X}\right| \tag{10}$$

$$\vec{X}(g+1)=\vec{X}_{rand}-\vec{A}.\vec{D} \tag{11}$$

In this study, the weight and bias coefficients of the hidden layer are optimized during the training of the GEP network using the whale method.

4. Results

Out of the 448 data utilized in the modeling, 336 data (75% of the total) were used for training, while the remaining 112 data (25% of the total) were provided for testing and assessing the modeling. The chosen percentages of the data were in the opposite direction of the best guess based on the methodologies chosen for estimating the water flow into the tunnel. Randomly selected data from all data in order to better estimate the ability in supervised learning created the best value for the lowest possible value of modeling error, which was determined after preparing and comparing 1100 models of algorithms, because choosing the right number for model learning, especially in Prediction algorithms can be unique for each study, which may change according to the parameters and the purpose of modeling [86].

4.1. Hybrid Neural Network-Particle Swarm Algorithm (ANN-PSO) model

Engineering challenges that are both linear and nonlinear can be solved with ANN. The hybrid particle swarm approach is applied with the neural network model in this study. The neural network models are presented in this part so that the new GEP models may be contrasted with the findings of the neural network models. 75% of the entire data, or 336 examples, were sent to the training part to be used in the model creation process, and 25% were given to the testing area in order to create the necessary networks. RMSE is often regarded as a crucial design factor for artificial neural networks. It is regarded as the main requirement for network training process termination. The values collected from the system (network) and the measured values may be used to calculate the RMSE value. It should be noted that the best fit model is found when RMSE = 0. Additionally, the value of the R determination coefficient was applied. It was in charge of figuring out the relationship between the expected value and the measured value. R is at its optimum when it equals 1. The present study's prediction models were assessed using two metrics: Rand RMSE. The model performed best when the number of neurons was fixed at 8, and the iteration value was set at 250. When BP is considered as a local learning process, the ANN optimum search approach could come up with a poor result. In order to improve the performance of ANN, biases and weights may be modified using PSO. There is occasionally a high likelihood of convergence when taking into account the local minimum of the ANN. PSO, however, is able to locate the global minimum. PSO-ANN therefore benefits from the search features of its methodology. PSO looks for global minima in the search space, and ANN utilizes them to discover the greatest performance with the least amount of system error. The efficiency of PSO-ANN may be influenced by a small number of factors, including swarm size, inertia weight, and coefficient speed. The inertial weight in this investigation is set at 1 utilized in the PSO-ANN model after being chosen. To build PSO-ANN models, many C1 and C2 combinations were taken into consideration through parametric study. While C1 = C2 = 2 was identified as the optimal model based on the lowest system error. As a result, the variables are listed in the PSO system as the best C1 and C2. Different values of SS from 50 to 400 with an incremental step of 50 and a total of 500 iterations were taken into consideration to identify the swarm size (SS) and the maximum number of iterations (IMax). Eight PSO-ANN models were therefore created in order to forecast water leaks. On the other hand, from iteration 1 to iteration 400, RMSE values of SS steadily declined. The findings of the RMSE remain constant after 400 iterations. Figure 9 shows the optimal ANN architecture. Figure 10 shows the performance of PSO-ANN model.

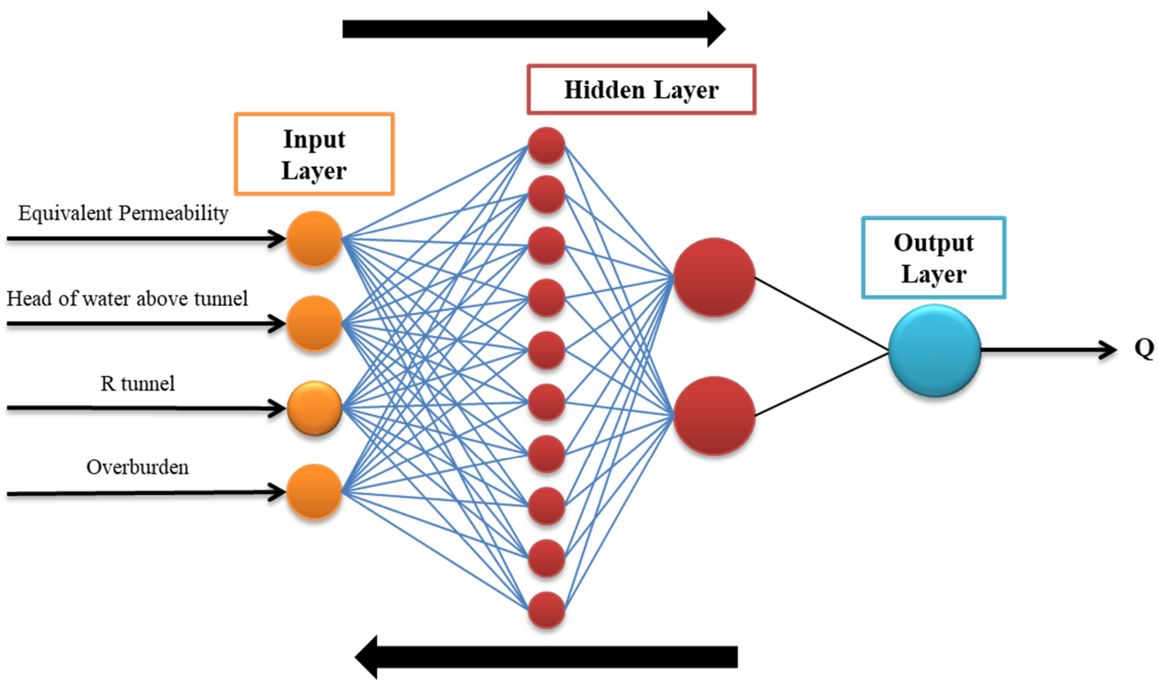


Figure 9. Optimum ANN architecture

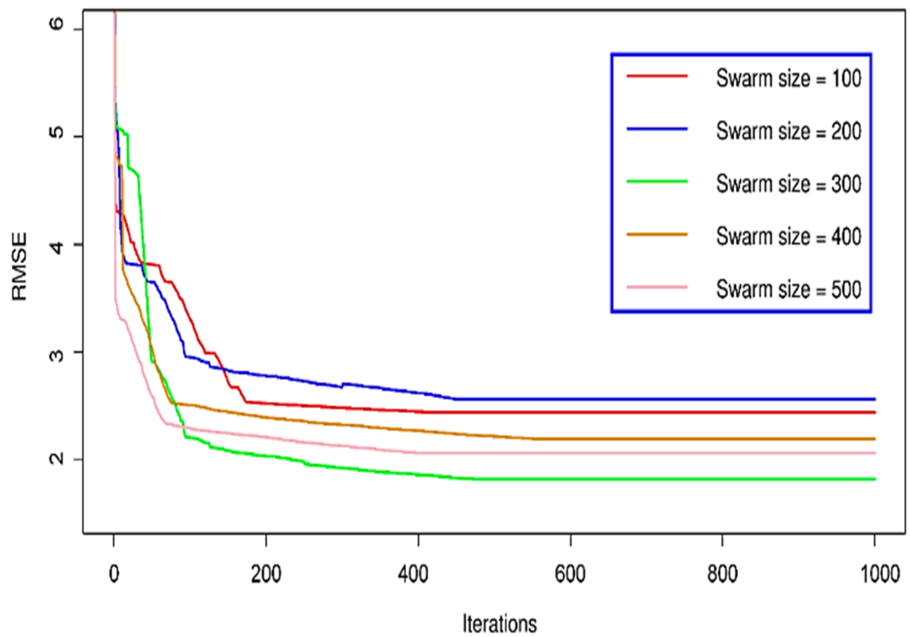


Figure 10. Performance of PSO-ANN model

4.2. Algorithm of gene expression

The water seepage into the tunnel is predicted using the GeneXpro Tools v5 program. Table 5 contains a list of the factors taken into account for prediction in the gene expression algorithm. The gene expression tree technique is depicted in Figure 11 as a simple way to explain gene expression modeling using programming.

Table 5. Variables considered in gene expression programming modeling

Variable	Number
Population size for training	336.000
Population size for the test	112.000
Complexity before simplification	77.000
Complexity after simplification	28.000
Gene transfer rate	0.040
Inversion ratio	0.040
IS Transfer rate	0.040
RIS transfer rate	0.040
Combination rate of genes	0.300
Single point compound rate	0.002
Two-point compound rate	0.002
Gene size	30.000
Head size	8.000
Tail size	20.000
Connector function	Avg
Mutation rate	0.001
Chromosome length	35.000
Number of genes	14.000

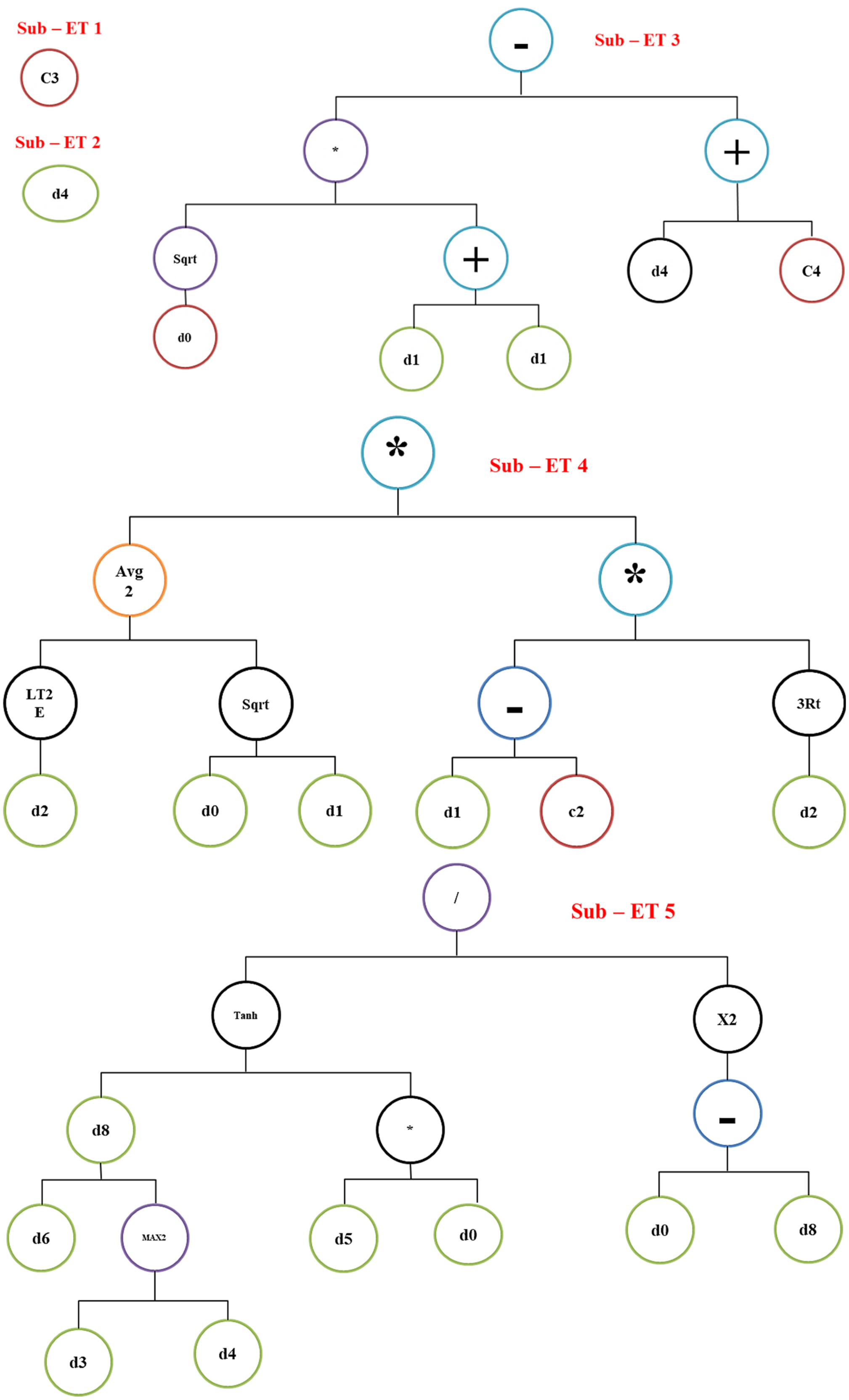


Figure 11. Expression tree of gene expression algorithm

In this stage, the GEP prediction models are implemented once the PSO-ANN network's findings have been obtained with the goal of creating an equation to forecast water loss into the tunnel. It will be discussed what the values are and how to use them to run GEP models and display the connections quantitatively. The following steps were taken to apply GEP in this study:

1. The fitness coefficient was chosen as a gauge for each chromosome's fitness occurrence in the first stage. A frequent fit function employed in the GEP modeling method is RMSE. Nevertheless, many modes may be employed to more thoroughly analyze the models' performance, depending on the characteristics of the situation. As a result, the following criteria were used to assess each chromosome's fitness:

2. The allocation of two crucial components, the set of terminals (T) and functions (F), to the structure of chromosomes, which resulted in a combination of them, was the second stage. The terminal set is thought of as the independent variables, and the function set is often specified with reference to the crux of the issue. Trigonometry and mathematical operations are applied as follows in this study:

$$RMSE' = \frac{1}{1+RMSE} \times 1000 \quad (12)$$

$$F = \{+,-,\times,.,\mathrm{Sin},\mathrm{Cos},Arc\ \mathrm{Tan},\tanh,sqrt\} \quad (13)$$

3. The third stage introduced and applied GEP structural parameters to the system. For each chromosome, the parameter of the number of genes for the designated ET components was added. Trial and error is the most effective method for obtaining optimal values for GEP structural parameters. The analytical method began by raising the values of the aforementioned GEP parameters, after which the GEP model's accuracy of prediction was assessed. To forecast the compressive strength of composite columns, several GEP models with various parameter settings have been developed and put into practice. Finally, after repeatedly repeating these procedures, the number values discovered for this part are 40, 5, and 3 accordingly.

4. The pace of genetic operators was chosen in the fourth phase. At this point, given the values recommended by earlier researchers, a few more GEP models were built through the process of trial and error. Table 3 displays the parameters of the GEP's obtained values. The relational functions of addition (+), subtraction (-), division (/), and multiplication (*) are only a few examples. R and RMSE were used as performance to assess how well GEP models predicted outcomes. In this part, a number of GEP model parameters were investigated to ascertain their effect on the models' performance. This allows for the comparison and performance assessment of several mathematical formulae for the prediction of uniaxial compressive strength in composite columns. Based on the R values, model number 4 was ultimately chosen as the best model.

In Table 6, various error criteria in the performance of the gene expression algorithm are compared. Figure 12 shows the performance of GEP for estimating Q of Amir Kabir tunnel.

Table 6. Comparison of different error criteria in the performance of gene expression algorithm

Description	Training	Testing
Fitness	1.72	1.55
MSE	170.680	155.860
RMSE	11.640	12.410
MAE	10.390	11.420
RSE	4.56	4.77
RRSE	0.27	0.37
Correlation coefficient	0.96	0.96
R-square	0.98	0.97

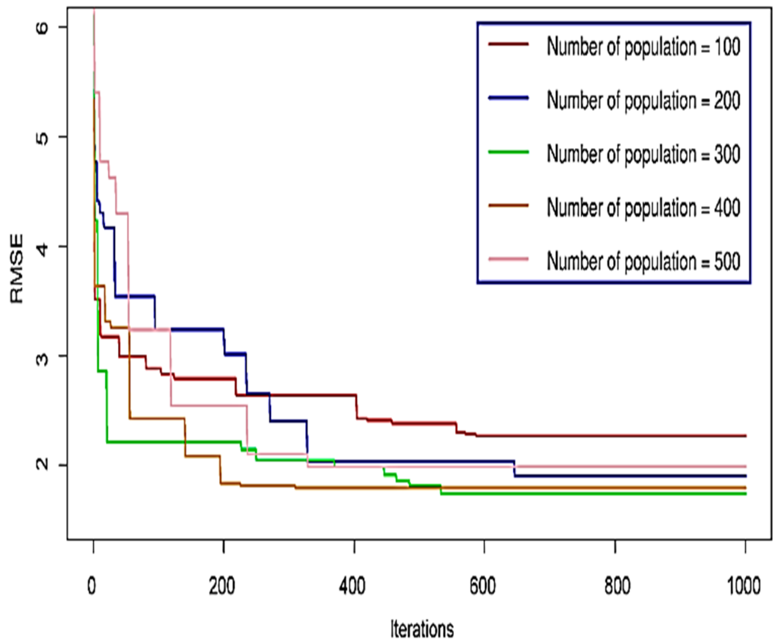


Figure 12. GEP performance for estimating Q of Amir Kabir tunnel

4.3. Whale Optimization Algorithm (WOA) based on GEP

The whale optimization algorithm (WOA) is created in this part to improve the predictions of tunnel water seepage. Utilizing the chosen functions, the WOA algorithm was tested. As you can see, the algorithm's written code does a good job at identifying the minimum. For optimization, a selective prediction model (GEP-based tree) was applied. In reality, the WOA approach takes the GEP equation into account as a cost function. Different models of the WOA algorithm (with various parameters) were created, and each model was implemented by varying the optimization algorithm's parameters. The WOA algorithm's best parameters were discovered through a series of analyses. Table 7 lists the ideal conditions for the WOA algorithm's performance in order to solve this issue optimally. The table contains the proposed parameters. Different design patterns may be used in various situations to get the best results. WOA may therefore improve model inputs to provide the greatest results with the least amount of error. WOA may therefore be presented as a powerful optimization technique for predicting water leaks. Figure 13 shows the performance of WOA-GEP to estimate Q of Amir Kabir tunnel.

Table 7. Variables considered in whale algorithm modeling

Variable	Count of search personnel	Maximum number of repetitions	A	L	b
Number	150	70	linearly decreased from 5 to 0	[-1,+1]	1

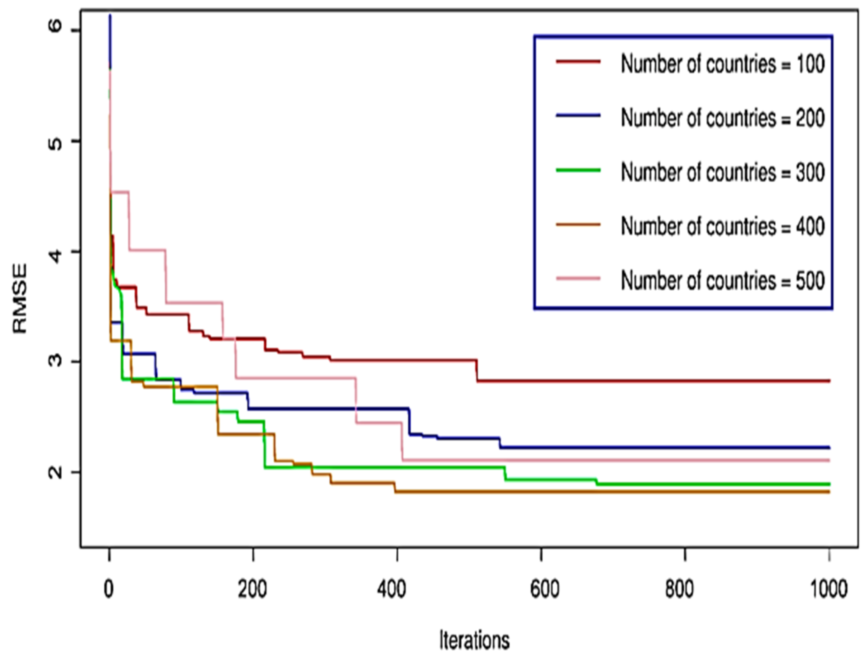


Figure 13. WOA-GEP performance for estimating Q of Amir Kabir tunnel

The ideal values of the input parameters are displayed in Table 8. As can be observed, this algorithm's written code has a good performance at identifying the minimum. As a result, it may be used with the research settings discovered in the preceding section.

Table 8. Optimum values of input parameters

Parameter	Initial value	Optimal value
Equivalent Permeability	4.560	4.540
Head of water above tunnel	251.810	250.810
R tunnel	2.350	2.350
Overburden	320.650	320.450
Q	0.012	0.011

5. Discussion

In this part, the methods used in this study have been compared and statistically the superiority of the methods has been determined. After completing the modeling, the correlation coefficient of all eight models was estimated and compared with each other. According to Figure 14, the gene expression optimization algorithm by the whale algorithm has a better match in training and evaluation in order to predict water seepage into the tunnel.

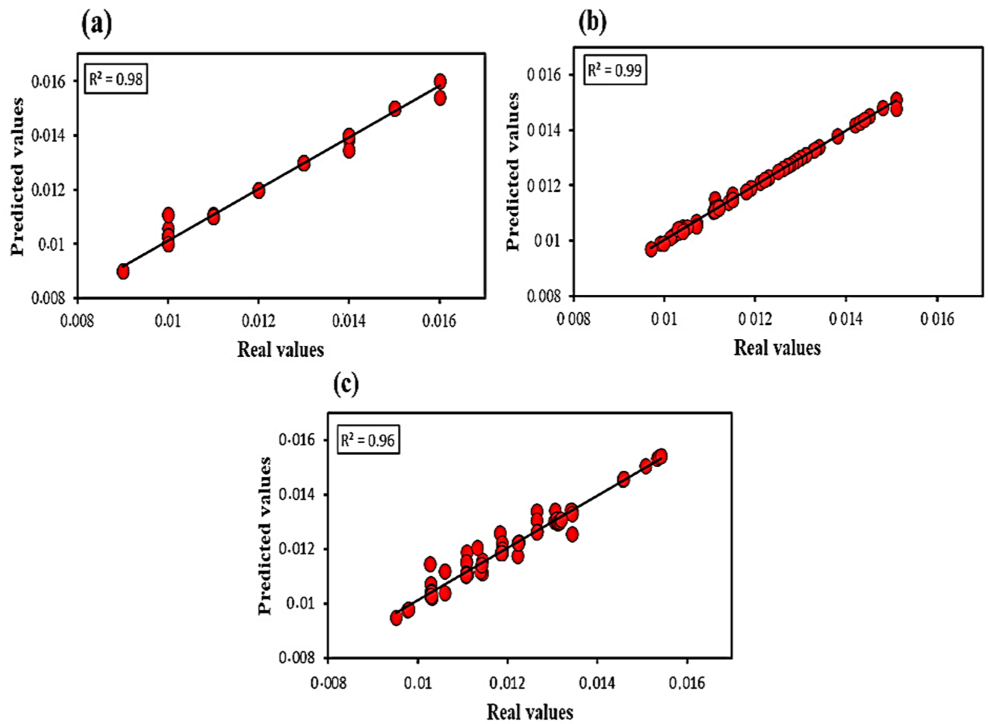


Figure 14. Comparison of the R2 index resulting from the comparison of the models of the selected algorithms in train state a) GEP b) WOA-GEP c) ANN-PSO

Comparing the outcomes with one another and with the actual data serves as the foundation for rating the performance of the models in this study. In this respect, five statistical indicators—the root mean square error [87], the mean absolute value of the error [88], the variance of the error [89], and the mean square error [90] are used to compare the outcomes from the models that have been provided (Eq. 14 to Eq.17). And R2 standards were looked upon.

$$RMSE = \frac{\sum \sqrt[2]{(X_{ir} - X_{ip})^2}}{n} \tag{14}$$

$$\mathrm{VAF} = 100(1 - \frac{\mathrm{var}(X_{ir} - X_{ip})}{\mathrm{var}(X_{ir})}) \tag{15}$$

$$MAE = \frac{\sum (X_{ir} - X_{ip})}{n} \tag{16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_{ir} - X_{ip})^2 \tag{17}$$

In the aforementioned equations, n stands for Xir and Xip, or the actual and expected values, respectively, as well as the total number of observations during the forecast period. Each model's error requirements are modest, which suggests that the values they forecast are more similar to actual values [53]. Additionally, the following newly developed engineering index, the α10-index, has been utilized to assess the dependability of the enlarged AI models:

$$a10 - index = \frac{m10}{M} \tag{18}$$

Where M is the number of datasets and m10 is the number of samples having measured or anticipated values for rates (range between 0.9 and 1.1). It is vital to note that a perfect prediction model requires an α10-index with a value of one. Table 9 compares the statistical indices of the models that were given. According to this table, it can be concluded that the optimization expression model of the whale algorithm makes accurate predictions about water seepage into the tunnel and may be a useful tool for predicting the water flow into the tunnel.

Table 9. Calculation of the overall error created in the methods

Method	WOA-GEP	GEP	ANN-PSO
RMSE Train	1.710	11.640	6.510
RMSE Test	1.820	12.410	7.320
MAE Train	0.002	10.390	3.670
MAE Test	0.003	11.420	4.554
VAF Train	99.210	92.560	92.520
VAF Test	99.110	91.860	91.550
MSE Train	6.650	170.680	73.790
MSE Test	6.620	155.860	72.880
a10-index Train	1.000	0.863	0.891
a10-index Test	1.000	0.850	0.870
R^2 Train	0.990	0.981	0.963
R^2 Test	0.981	0.972	0.952

6. Sensitivity analysis

Multivariate sensitivity analysis considers the impact of different variables in the modeling process and encompasses a wide range of data through Monte Carlo simulation [91]. To begin, a uniform distribution is fitted to each input variable of the model individually. For instance, in this study, separate estimations of groundwater seepage into tunnels are performed for each variable. Subsequently, random numbers are generated using Monte Carlo simulation, utilizing the fitted distributions and the available data for each variable [92]. In the next step, the influence of each variable is assessed by calculating the objective function, which is the sum of squared errors between the observed values and the modeled values[93]:

$$f_h = \sum_{i=0}^{k} \left[x_{0,h} - x_{c,h}(i) \right]^2 \tag{10}$$

Here, f_h represents the value of the objective function, $x_{0,h}$ represents the observed value from modeling, and $x_{c,h}(i)$ represents the actual value at time h at random value i. By evaluating the objective function, the sensitivity analysis impact factor is calculated:

$$\lambda_h = \frac{f_h}{x_{0,h}} \tag{11}$$

After calculating the impact variable (λ_h), the factors determining whether the factors are independent or dependent (γ) are calculated [91]:

$$\gamma = \sum_{h=0}^{i_{FC.\max}} \lambda_h \tag{12}$$

Once the relation index in Equation 12 is calculated, it is possible to assess the influence of each variable in the modeling process. This results in three possible situations:

Insensitive variable: $\gamma \le 1$

Sensitive variable: $1 < \gamma \le 100$

Critical variable: $\gamma > 100$

Table 10 presents the results of the multivariate sensitivity analysis conducted in this research.

As can be seen, the head of water above tunnel parameter has a critical effect on modeling. The rest of the parameters are also sensitive parameters based on multi-parameter sensitivity analysis calculations.

Table 10. Results of multivariate sensitivity analysis

Parameters	Sensitivity value	variable status
Equivalent Permeability	87	sensitive
Head of water above tunnel	114	Critical
R tunnel	54	sensitive
Overburden	98	sensitive

7. Validation

In 2012, Farhadian et al. [94] modeled groundwater seepage into Amirkabir tunnel using analytical methods. To validate the models created in this study, the results of the modeling were compared with the output of Farhadian et al. The figure 15 shows the results of comparing the models. As can be seen, the results obtained from the selected model in the present study are consistent with the results obtained from Farhadian's analytical method, and this shows the effectiveness of the model compared to analytical models.

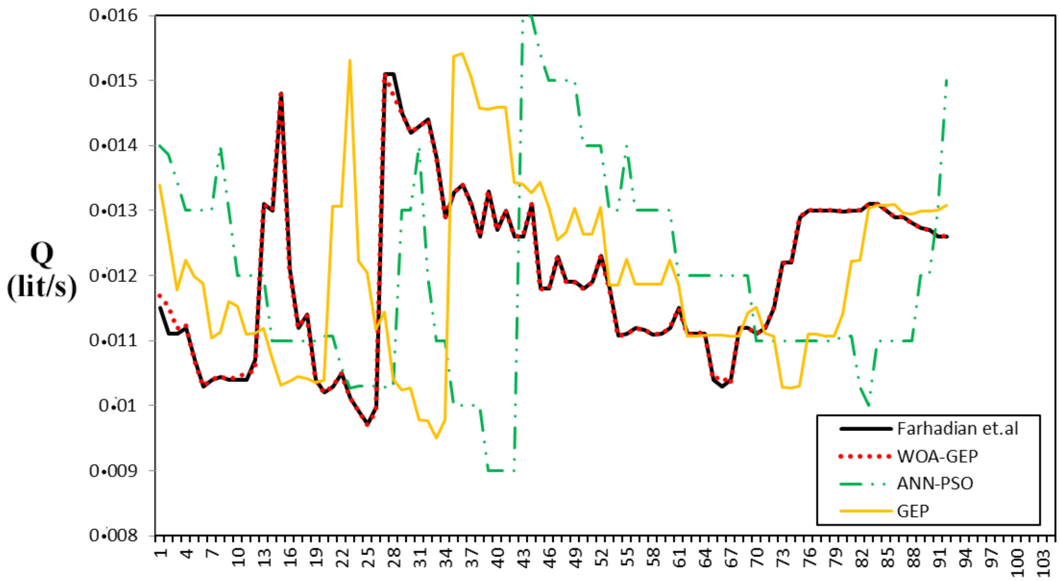


Figure 15. Comparing the results of the modeling done in the present study with the analytical method of Farhadian et al

8. Conclusions

In recent years, the engineering community has placed a heightened emphasis on the precise forecasting of groundwater inflow, especially in applications critical to environmental impact assessment and the design of tunnel seepage systems. Despite the availability of various analytical equations in the technological background addressing this concern, many of these equations were developed under oversimplified assumptions, notably assuming a homogeneous aquifer. This oversimplification is at odds with the intricate and fragmented structure of rock masses typically encountered in tunnel environments, resulting in an inadequate representation of real-world occurrences.

To overcome these limitations and enhance the anticipation of water seepage into Amir kabir tunnel, this study employed two hybrid approaches: the neural network-particle swarm algorithm, gene expression algorithm, and optimization of the selected method using the whale algorithm. The primary objective was to extend and compare machine learning methods to offer more robust insights. Upon thorough comparison, it became evident that the whale optimization algorithm, exhibiting correlation coefficient values of 0.99 for the training set and 0.98 for the test set, demonstrated remarkable accuracy in forecasting the extent of water seepage into the tunnel. This underscores the efficacy of advanced optimization techniques in enhancing the precision of groundwater inflow predictions in complex geological settings.

References

[1]. Rabieian, M., & Qaderi, F. (2024). Optimizing Hybrid Photocatalytic-ozonation for Offshore Produced Water Treatment. *Journal of Mining and Environment*, 15(1), 239-259. doi:10.22044/jme.2023.13081.2376

[2]. Chen, Y., Selvadurai, A. P. S., & Liang, W. (2019). Computational Modelling of Groundwater Inflow During a Longwall Coal Mining Advance: A Case Study from the Shanxi Province, China. *Rock Mechanics and Rock Engineering*, 52(3), 917-934. doi:10.1007/s00603-018-1603-1

[3]. Dadashi, E., Noorzad, A., Shahriar, K., & Goshtasbi, K. (2018). An optimal method to design reinforced concrete lining of pressure tunnels. *Journal of Mining and Environment*, 9(4), 829-837. doi:10.22044/jme.2018.6776.1500

[4]. Ackerman, J. R., Peterson, E. W., Van der Hoven, S., & Perry, W. L. (2015). Quantifying nutrient removal from groundwater seepage out of constructed wetlands receiving treated wastewater effluent. *Environmental Earth Sciences*, 74(2), 1633-1645.

doi:10.1007/s12665-015-4167-3

[5]. Rasmussen, T. C., & Yu, G. (2006). Determination of groundwater flownets, fluxes, velocities, and travel times using the complex variable boundary element method. *Engineering Analysis with Boundary Elements*, 30(12), 1030-1044. doi:https://doi.org/10.1016/j.enganabound.2006.01.017

[6]. Frough, O., Torabi, S. R., & Tajik, M. (2012). Evaluation of TBM Utilization Using Rock Mass Rating System: A Case Study of Karaj-Tehran Water Conveyance Tunnel (Lots 1 and 2). *Journal of Mining and Environment*, 3(2), 89-98. doi:10.22044/jme.2012.86

[7]. Ying, H., Zhu, C., Shen, H., & Gong, X. (2018). Semi-analytical solution for groundwater ingress into lined tunnel. *Tunnelling and Underground Space Technology*, 76, 43-47. doi:https://doi.org/10.1016/j.tust.2018.03.009

[8]. Kim, G. B. (2013). Assessment of water seepage through a geologic barrier surrounding a large reservoir using groundwater levels, soil condition, and a numerical model. *Environmental Earth Sciences*, 69(6), 2059-2072. doi:10.1007/s12665-012-2041-0

[9]. Kamali, A., Shahriar, K., El Tani, M., Aalianvari, A., & Gholami, M. A. (2018). Challenging Estimation of Seepage in Powerhouse Cavern and Drainage Tunnel in Iran. *ISRM European Rock Mechanics Symposium - EUROCK*, May 2018, ISRM-EUROCK-2018-173.

[10]. Wadslin Frenelus, & Hui Peng JZ. (2021). Evaluation methods for groundwater inflows into rock tunnels: a state-of-the-art review Abstract. *International Journal of Hydrology Research*, 5(4), 152-168.

[11]. Zhang, L., Yang, D., Liu, Y., Che, Y., & Qin, D. (2014). Impact of impoundment on groundwater seepage in the Three Gorges Dam in China based on CFCs and stable isotopes. *Environmental Earth Sciences*, 72(11), 4491-4500. doi:10.1007/s12665-014-3349-8

[12]. Wang, Z. H., Wu, S. X., Li, J. L., Sun, W. C., Wang, Z. F., & Liu, P. J. (2023). Surface subsidence and its reclamation of a coal mine locating at the high groundwater table, China. *International Journal of Environmental Science and Technology*. doi:10.1007/s13762-023-04915-8

[13]. 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

[14]. Surinaidu, L., Gurunadha Rao, V. V. S., Srinivasa Rao, N., & Srinu, S. (2014). Hydrogeological and groundwater modeling studies to estimate the groundwater inflows into the coal Mines at different mine development stages using MODFLOW, Andhra Pradesh, India. *Water Resources and Industry*, 7-8, 49-65. doi:https://doi.org/10.1016/j.wri.2014.10.002

[15]. Chandel, N., Gupta, S. K., & Ravi, A. K. (2024). Groundwater Quality Analysis using Machine Learning Techniques: a Critical Appraisal. *Journal of Mining and Environment*, 15(2), 419-426. doi:10.22044/jme.2023.13452.2484

[16]. Ke, X., Wang, W., Xu, X., Li, J., & Hu, H. (2021). A saturated–unsaturated coupling model for groundwater flowing into seepage wells: a modeling study for groundwater development in river basins. *Environmental Earth Sciences*, 80(21), 711. doi:10.1007/s12665-021-10035-8

[17]. Okeke, C. A. U., Azuh, D., Ogbuagu, F. U., & Kogure, T. (2020). Assessment of land use impact and seepage erosion contributions to seasonal variations in riverbank stability: The Iju River, SW Nigeria. *Groundwater for Sustainable Development*, 11, 100448. doi:https://doi.org/10.1016/j.gsd.2020.100448

[18]. Gholizadeh, H., Behrouj Peely, A., Karney, B. W., & Malekpour, A. (2020). Assessment of groundwater ingress to a partially pressurized water-conveyance tunnel using a conduit-flow process model: a case study in Iran. *Hydrogeology Journal*, 28(7), 2573-2585. doi:10.1007/s10040-020-02213-y

[19]. Janipella, R., & Pujari, P. R. (2022). Review on Groundwater Flow and Solute Transport Modelling in India: Recent Advances and Future Directions. *Journal of the Geological Society of India*, 98(2), 278-284. doi:10.1007/s12594-022-1968-3

[20]. Aalianvari, A., Malekitehrani, M., & Soltanomohammadi, S. (2014). Estimation of water flow from the upper reservoir of Azad pumped storage power plant, using geostatistical methods. *Journal of the Geological Society of India*, 83(1), 76-82. doi:10.1007/s12594-014-0009-2

[21]. Aghda, S. M. F., GanjaliPour, K., & Esmaeilzadeh, M. (2019). The Effect of Geological Factors on the Grout Curtain Performance Analysis of Darian Dam Using the Results of Instrumentation Data in the First Impounding. *Journal of the Geological Society of India*, 93(3), 360-368. doi:10.1007/s12594-019-1185-x

[22]. Jiang, C., Han, H., Xie, H., Liu, J., Chen, Z., & Chen, H. (2021). Karst Aquifer Water Inflow into Tunnels: An Analytical Solution. In Y. Zhao (Ed.), *Geofluids*, 2021, 6672878. doi:10.1155/2021/6672878

[23]. Javadi, M., Sharifzadeh, M., & Shahriar, K. (2016). Uncertainty analysis of groundwater inflow into underground excavations by the stochastic discontinuum method: Case study of Siah Bisheh pumped storage project, Iran. *Tunnelling and Underground Space Technology*, 51, 424-438.

doi:https://doi.org/10.1016/j.tust.2015.09.003

[24]. Jiang, Q., Yao, C., Ye, Z., & Zhou, C. (2013). Seepage flow with free surface in fracture networks. *Water Resources Research*, 49(1), 176-186. doi:10.1029/2012WR011991

[25]. Li, X., Li, D., Xu, Y., & Feng, X. (2020). A DFN-based 3D numerical approach for modeling coupled groundwater flow and solute transport in fractured rock mass. *International Journal of Heat and Mass Transfer*, 149, 119179. doi:10.1016/j.ijheatmasstransfer.2019.119179

[26]. Liu, D., Xu, Q., Tang, Y., & Jian, Y. (2021). Prediction of Water Inrush in Long-Lasting Shutdown Karst Tunnels Based on the HGWO-SVR Model. *IEEE Access*, 9, 6368-6378. doi:10.1109/ACCESS.2020.3047626

[27]. Nikvar Hassani, A., Farhadian, H., & Katibeh, H. (2018). A comparative study on the evaluation of steady-state groundwater inflow into a circular shallow tunnel. *Tunnelling and Underground Space Technology*, 73, 15-25. doi:https://doi.org/10.1016/j.tust.2017.11.019

[28]. Jing, L. (2003). A review of techniques, advances, and outstanding issues in numerical modeling for rock mechanics and rock engineering. *International Journal of Rock Mechanics and Mining Sciences*, 39(4), 283-353. doi:10.1016/S1365-1609(03)00013-3

[29]. Nikakhtar, L., & Zare, S. (2020). Evaluation of Underground Water Flow into Tabriz Metro Tunnel First Line by Hydro-Mechanical Coupling Analysis. *International Journal of Geotechnical and Geological Engineering*, 14(1), 1-7.

[30]. Wu, Q., Wang, M., & Wu, X. (2004). Investigations of groundwater bursting into coal mine seam floors from fault zones. *International Journal of Rock Mechanics and Mining Sciences*, 41(4), 557-571. doi:https://doi.org/10.1016/j.ijrmms.2003.01.004

[31]. Lianchong, L., Tianhong, Y., Zhengzhao, L., Wancheng, Z., & Chunan, T. (2011). Numerical investigation of groundwater outbursts near faults in underground coal mines. *International Journal of Coal Geology*, 85(3), 276-288. doi:https://doi.org/10.1016/j.coal.2010.12.006

[32]. Farhadian, H., Katibeh, H., & Huggenberger, P. (2016). Empirical model for estimating groundwater flow into a tunnel in discontinuous rock masses. *Environmental Earth Sciences*, 75(6), 471. doi:10.1007/s12665-016-5332-z

[33]. Li, Q., Ito, K., Wu, Z., Lowry, C. S., & Loheide, S. P. (2009). COMSOL Multiphysics: A novel approach to groundwater modeling. *Ground Water*, 47(4), 480-487. doi:10.1111/j.1745-6584.2009.00584.x

[34]. Li, S., He, P., & Li, L. (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

[35]. 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

[36]. Yang, Z. (2016). Risk prediction of water inrush of karst tunnels based on BP neural network. In *4th International Conference on Mechanical Materials and Manufacturing Engineering*. doi:10.2991/mmme-16.2016.74

[37]. Donglin, D., Wenjie, S., & Sha, X. (2012). Water-inrush assessment using a GIS-based Bayesian network for the 12-2 coal seam of the Kailuan Donghuantuo coal mine in China. *Mine Water and the Environment*, 31(2), 138-146. doi:10.1007/s10230-012-0178-4

[38]. Mahmoodzadeh, A., Mohammadi, M., & Gharrib Noori, K. (2021). Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques. *Automation in Construction*, 127, 103719. doi:https://doi.org/10.1016/j.autcon.2021.103719

[39]. Chen, S., & Dong, S. (2020). A sequential structure for water inflow forecasting in coal mines integrating feature selection and multi-objective optimization. *IEEE Access*, 8, 183619-183632. doi:10.1109/ACCESS.2020.3028959

[40]. Liu, W., & Zhang, H. (2016). PSO algorithm for block sequencing problem in open pit mining. In *Proceedings of the 2015 5th International Conference on Computer Science and Automation Engineering*. doi:10.2991/iccsae-15.2016.83

[41]. Bagatur, T., & Onen, F. (2014). Prediction of flow and oxygen transfer by a plunging water jet with genetic expression programming (GEP) models. *Arabian Journal for Science and Engineering*, 39(6), 4421-4432. doi:10.1007/s13369-014-1092-9

[42]. Hemasian-Etefagh, F., & Safi-Esfahani, F. (2020). Group-based whale optimization algorithm. *Soft Computing*, 24(5), 3647-3673. doi:10.1007/s00500-019-04131-y

[43]. Farhadian, H., Aalianvari, A., & Katibeh, H. (2012). Optimization of analytical equations of groundwater seepage into tunnels: A case study of Amirkabir tunnel. *Journal of the Geological Society of India*, 80, 96-100. doi:10.1007/s12594-012-0122-z

[44]. Koseoglu Balta, G. C., Dikmen, I., & Birgonul, M. T. (2021). Bayesian network-based decision support for predicting and mitigating delay risk in TBM tunnel projects. *Automation in Construction*, 129, 103819. doi:10.1016/j.autcon.2021.103819

[45]. Embaby, A. K., Gomaa, S., Darwish, Y., & Selim, S. (2024). Predicting Gabal Gattar Uranium Content as a Function of Total Gamma-ray and Thorium Contents using an Artificial Neural Network in Northeastern Desert, Egypt. *Journal of Mining and Environment*, 15(1), 175-189. doi:10.22044/jme.2023.13651.2524

[46]. Mostafaei, K., Kianpour, M. N., & Yousefi, M. (2024). Delineation of Gold Exploration Targets based on Prospectivity Models through an Optimization Algorithm. *Journal of Mining and Environment*, 15(2), 597-611. doi:10.22044/jme.2023.13472.2489

[47]. Rezaei, M., Hossaini, M. F., Majdi, A., & Najmoddini, I. (2017). Determination of the height of destressed zone above the mined panel: An ANN model. *International Journal of Mining and Geo-Engineering*, 51(1), 1-7. doi:10.22059/ijmge.2017.62147

[48]. Alimoradi, A., Moradzadeh, A., Naderi, R., Salehi, M. Z., & Etemadi, A. (2008). Prediction of geological hazardous zones in front of a tunnel face using TSP-203 and artificial neural networks. *Tunnelling and Underground Space Technology*, 23(6), 711-717. doi:https://doi.org/10.1016/j.tust.2008.01.001

[49]. Razavi Rad, F., Mohammad Torab, F., & Abdollahzadeh, A. (2016). Estimation of Cadmium and Uranium in a stream sediment from Eshtehard region in Iran using an Artificial Neural Network. *Journal of Mining and Environment*, 7(1), 97-107. doi:10.22044/jme.2016.500

[50]. Saini, A., & Yadav, J. S. (2024). Bearing Capacity of Circular Footing Resting on Recycled Construction Waste Materials using ANN Method. *Journal of Mining and Environment*, 15(1), 97-114. doi:10.22044/jme.2023.13453.2485

[51]. Su, M., Zhang, Z., Zhu, Y., Zha, D., & Wen, W. (2019). Data-driven natural gas spot price prediction models using machine learning methods. *Energies*, 12(9). doi:10.3390/en12091680

[52]. Bayatzadeh Fard, Z., Ghadimi, F., & Fattahi, H. (2017). Use of artificial intelligence techniques to predict the distribution of heavy metals in groundwater of Lakan lead-zinc mine in Iran. *Journal of Mining and Environment*, 8(1), 35-48. doi:10.22044/jme.2016.592

[53]. Moshrefi, S., Shahriar, K., Ramezanzadeh, A., & Goshtasbi, K. (2018). Prediction of the ultimate strength of shale using artificial neural network. *Journal of Mining and Environment*, 9(1), 91-105. doi:10.22044/jme.2017.5790.1390

[54]. Saljoughi, B. S., & Hezarkhani, A. (2024). A Comparative Analysis of Artificial Neural Network (ANN) and Gene Expression Programming (GEP) Data-driven Models for Prospecting Porphyry Cu Mineralization; Case Study of Shahr-e-Babak Area, Kerman Province, SE Iran. *Journal of Mining and Environment*, 15(2), 761-790. doi:10.22044/jme.2023.13852.2573

[55]. Taiwo, B. O., Famobuwa, O. V., & Mata, M. M. (2024). Granite Downstream Production Dependent Size and Profitability Assessment: an application of Mathematical-based Artificial Intelligence Model and WipFrag Software. *Journal of Mining and Environment*, 15(2), 497-515. doi:10.22044/jme.2023.13731.2543

[56]. Chai, S. H., & Lim, J. S. (2016). Forecasting business cycle with chaotic time series based on neural network with weighted fuzzy membership functions. *Chaos, Solitons and Fractals*, 0, 1-9. doi:10.1016/j.chaos.2016.03.037

[57]. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Neural Networks, 1995 Proceedings, IEEE International Conference*, 4, 1942-1948 vol.4. doi:10.1109/ICNN.1995.488968

[58]. Eil Saadatmand, F., Alimoradi, A., Moosavi, M. M., & Davari, M. A. (2024). Optimizing Deep Learning Models for Shear Wave Velocity Estimation Utilizing Petrophysical Logs: A Case Study on an Oil Reservoir in Southern Iran. *Journal of Petroleum Science and Engineering*, 31-47. doi:10.22107/jpg.2024.436270.1225

[59]. Fathi, M., Alimoradi, A., & Hemati Ahooi, H. R. (2021). Optimizing Extreme Learning Machine Algorithm using Particle Swarm Optimization to Estimate Iron Ore Grade. *Journal of Mining and Environment*, 12(2), 397-411. doi:10.22044/jme.2021.10368.1984

[60]. Arbabsiar, M. H., Farsangi, M. A. E., & Mansouri, H. (2020). A new model for predicting the advance rate of a tunnel boring machine (Tbm) in hard rock conditions. *Rudarsko-Geološko-Naftni Zbornik*, 35(2), 57-74. doi:10.17794/rgn.2020.2.6

[61]. Sharifi, F., Arab Amiri, A. R., & Kamkar Rouhani, A. (2019). Using a combination of genetic algorithm and particle swarm optimization algorithm for GEMTIP modeling of spectral-induced polarization data. *Journal of Mining and Environment*, 10(2), 493-505. doi:10.22044/jme.2019.7902.1655

[62]. Rezaei, M., & Asadizadeh, M. (2020). Predicting Unconfined Compressive Strength of Intact Rock Using New Hybrid Intelligent Models. *Journal of Mining and Environment*, 11(1), 231-246. doi:10.22044/jme.2019.8839.1774

[63]. Jing, L. J., Li, J. B., Zhang, N., Chen, S., Yang, C., & Cao, H. B. (2021). A TBM advance rate prediction method considering the effects of operating factors. *Tunnelling and Underground Space Technology*, 107, 103620. doi:10.1016/j.tust.2020.103620

[64]. Zarean, A., & Poormirzaee, R. (2016). Joint inversion of ReMi dispersion curves and refraction

travel times using particle swarm optimization algorithm. *Journal of Mining and Environment*, 7(1), 67-79. doi:10.22044/jme.2016.492

[65]. Mousavi, S. S., Nikkhah, M., & Zare, S. (2019). A comparative study of two meta-heuristic algorithms in optimizing the cost of reinforced concrete segmental lining. *Journal of Mining and Environment*, 10(1), 95-112.

[66]. Sun, S., Sun, Y., Wang, S., & Wei, Y. (2018). Interval decomposition ensemble approach for crude oil price forecasting. *Energy Economics*, 76, 274-287. doi:10.1016/j.eneco.2018.10.015

[67]. Alimoradi, A., Hajkarimian, H., Hemati Ahooi, H., & Salsabili, M. (2022). Comparison between the performance of four metaheuristic algorithms in training a multilayer perceptron machine for gold grade estimation. *International Journal of Mining and Geo-Engineering*, 56(2), 97-105. doi:10.22059/ijmge.2021.314154.594880

[68]. Wang, X., Lu, H., Wei, X., Wei, G., Behbahani, S. S., & Iseley, T. (2020). Application of Artificial Neural Network in Tunnel Engineering: A Systematic Review. *IEEE Access*, 8, 119527-119543. doi:10.1109/ACCESS.2020.3004995

[69]. Nikakhtar, L., Zare, S., & Mirzaei, H. (2023). Performance Comparison of Particle Swarm Optimization and Genetic Algorithm for Back-analysis of Soil Layer Geotechnical Parameters. *Journal of Mining and Environment*, 14(1), 217-232. doi:10.22044/jme.2023.12452.2260

[70]. Moomivand, H., Amini Khoshalan, H., Shakeri, J., & Vandyousefi, H. (2022). Development of new comprehensive relations to assess rock fragmentation by blasting for different open pit mines using GEP algorithm and MLR procedure. *International Journal of Mining and Geo-Engineering*, 56(4), 401-411. doi:10.22059/ijmge.2022.339174.594951

[71]. Bastami, R., Bazzazi, A. A., Shoormasti, H. H., & Ahangari, K. (2020). Predicting and minimizing the blasting cost in limestone mines using a combination of gene expression programming and particle swarm optimization. *Archives of Mining Sciences*, 65(4), 835-850. doi:10.24425/ams.2020.135180

[72]. Monjezi, M., Dehghani, H., Shakeri, J., & Mehrdanesh, A. (2021). Optimization of prediction of flyrock using linear multivariate regression (LMR) and gene expression programming (GEP)—Topal Novin mine, Iran. *Arabian Journal of Geosciences*, 14(15). doi:10.1007/s12517-021-07772-2

[73]. Kadkhodaei, M. H., & Ghasemi, E. (2019). Development of a GEP model to assess CERCHAR abrasivity index of rocks based on geomechanical properties. *Journal of Mining and Environment*, 10(4), 917-928. doi:10.22044/jme.2019.8141.1684

[74]. Jahed, D., Vali, A., Ahmad, S., & Mohd, F. (2017). Uniaxial compressive strength prediction through a new technique based on gene expression programming. *Neural Computing and Applications*. doi:10.1007/s00521-017-2939-2

[75]. Yang, B., Zhang, W., & Wang, H. (2019). Stock Market Forecasting Using Restricted Gene Expression Programming. De Maio C, ed. *Computational Intelligence and Neuroscience*, 2019, 7198962. doi:10.1155/2019/7198962

[76]. Dehghani, H., Velicković, M., Jodeiri, B., Mihajlović, I., Nikolić, D., & Panic, M. (2022). Determination of ozone concentration using gene expression programming algorithm (GEP)- Zrenjanin, Serbia. *International Journal of Mining and Geo-Engineering*, 56(1), 1-9. doi:10.22059/ijmge.2021.313278.594874

[77]. Hashmi, M. Z., & Shamseldin, A. Y. (2014). Advances in Water Resources Use of Gene Expression Programming in regionalization of flow duration curve. *Advances in Water Resources*, 68, 1-12. doi:10.1016/j.advwatres.2014.02.009

[78]. Shirani Faradonbeh, R., Jahed Armaghani, D., & Abd Majid, M. Z. (2016). Prediction of ground vibration due to quarry blasting based on gene expression programming: a new model for peak particle velocity prediction. *International Journal of Environmental Science and Technology*, 13(6), 1453-1464. doi:10.1007/s13762-016-0979-2

[79]. Xue, X., & Deng, C. Prediction of creep index of soft clays using gene expression programming. *Soft Computing*, 27, 16265–16278. doi:10.1007/s00500-023-08053-8

[80]. Armaghani, D. J., & Asteris, P. G. (2021). A Comparative Study of ANN and ANFIS Models for the Prediction of Cement-Based Mortar Materials Compressive Strength. *Neural Computing and Applications*, 33. Springer London. doi:10.1007/s00521-020-05244-4

[81]. Turgut, M. S., Sağban, H. M., Turgut, O. E., & Özmen, Ö. T. (2021). Whale optimization and sine–cosine optimization algorithms with cellular topology for parameter identification of chaotic systems and Schottky barrier diode models. *Soft Computing*. 25(2), 1365-1409. doi:10.1007/s00500-020-05227-6

[82]. Tang, C., Sun, W., Xue, M., Zhang, X., Tang, H., & Wu, W. (2022). A hybrid whale optimization algorithm with artificial bee colony. *Soft Computing*, 26(5), 2075-2097. doi:10.1007/s00500-021-06623-2

[83]. Li, Z., Yazdani Bejarbaneh, B., Asteris, P. G., Koopialipoor, M., Armaghani, D. J., & Tahir, M. M. (2021). A hybrid GEP and WOA approach to estimate the optimal penetration rate of TBM in granitic rock mass. *Soft Computing*, 25(17), 11877-11895. doi:10.1007/s00500-021-06005-8

[84]. Nabavi, Z., Mirzehi, M., Dehghani, H., & Ashtari,

P. (2023). A Hybrid Model for Back-Break Prediction using XGBoost Machine learning and Metaheuristic Algorithms in Chadormalu Iron Mine. *Journal of Mining and Environment*, 14(2), 689-712. doi:10.22044/jme.2023.12796.2323

[85]. Mirjalili, S., & Mirjalili, A. L. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, 95.

[86]. Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Chapter 12. Machine Learning in Transportation Data Analytics, *Elsevier Inc*. doi:10.1016/B978-0-12-809715-1.00012-2

[87]. Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. doi:10.5194/gmd-7-1247-2014

[88]. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82. doi:10.3354/cr030079

[89]. Jahanmiri, S., Asadizadeh, M., Alipour, A., & Nowak, S. (2021). Predicting the Contribution of Mining Sector to the Gross Domestic Product (GDP) Index Utilizing Heuristic Approaches. *Applied Artificial Intelligence*, 00(00), 1-23. doi:10.1080/08839514.2021.1997225

[90]. Marmolin, H. (1986). Subjective Mse Measures. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-16(3), 486-489. doi:10.1109/tsmc.1986.4308985

[91]. Lamboni, M., Monod, H., & Makowski, D. (2011). Multivariate sensitivity analysis to measure the global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4), 450-459. doi:https://doi.org/10.1016/j.ress.2010.12.002

[92]. Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., & Monod, H. (2009). Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, 113(3), 312-320. doi:https://doi.org/10.1016/j.fcr.2009.06.007

[93]. Chen, L., & Huang, H. (2024). Global sensitivity analysis for multivariate outputs using generalized RBF-PCE metamodel enhanced by variance-based sequential sampling. *Applied Mathematics and Computation*, 126, 381-404. doi:https://doi.org/10.1016/j.apm.2023.10.047

[94]. Farhadian, H., Aalianvari, A., & Katibeh, H. (2012). Optimization of analytical equations of groundwater seepage into tunnels: A case study of Amirkabir tunnel. *Journal of Geological Society of India*, 80(1), 96-100. doi:10.1007/s12594-012-0122-z

توسعه مدل‌های بهینه‌سازی مبتنی بر درخت GEP، ازدحام-عصبی و نهنگ برای ارزیابی نشت آب زیرزمینی به داخل تونل: مطالعه موردی

شیرین جهانمیری، علی عالی انوری* و ملیحه عباس زاده

دانشکده مهندسی معدن، دانشگاه کاشان، کاشان، ایران

ارسال ۲۰۲۴/۰۹/۱۳، پذیرش ۲۰۲۴/۰۴/۱۲

* نویسنده مسئول مکاتبات: ali_aalianvari@kashanu.ac.ir

چکیده:

جریان آب زیرزمینی موضوع مهمی در حوزه‌های هیدرولوژی، مهندسی هیدرولیک، هیدروژئولوژی، مهندسی سنگ و رشته های مرتبط است. تونل‌های حفاری شده در زیر سطح آب زیرزمینی، به ویژه، با خطر ذاتی نشت آب زیرزمینی در طول فرآیند حفاری و مراحل عملیاتی بعدی مواجه هستند. جریان آب زیرزمینی که اغلب به عنوان خطرات زمین‌شناسی نادر تلقی می‌شود، می‌تواند باعث بی‌ثباتی در سازندهای سنگی اطراف شود و منجر به عواقب شدیدی مانند خسارات، تلفات و هزینه‌های مالی قابل توجه شود. هدف اصلی این تحقیق بررسی کاربرد تکنیک‌های یادگیری ماشین برای شناسایی دقیق‌ترین روش پیش‌بینی نشت آب تونل است. پیش‌بینی اتلاف آب به داخل تونل در مرحله پیش‌بینی، از یک معادله درختی مبتنی بر برنامه‌ریزی بیان ژن (GEP) استفاده کرد. نتایج حاصل از GEP با نتایج به‌دست‌آمده از یک مدل ترکیبی شامل بهینه‌سازی ازدحام ذرات (PSO) و شبکه‌های عصبی مصنوعی (ANN) مقایسه شد. الگوریتم بهینه‌سازی نهنگ (WOA) در مرحله بهینه‌سازی انتخاب و توسعه داده شد. الگوریتم بهینه‌سازی نهنگ در مقایسه با روش‌های فوق، عملکرد برتری را نشان داد و دقیقاً حجم آب از دست رفته در تونل را با ضریب همبستگی ۰/۹۹ پیش‌بینی کرد. این امر بر اثربخشی تکنیک‌های بهینه‌سازی پیشرفته در افزایش دقت پیش‌بینی جریان آب زیرزمینی و کاهش خطرات بالقوه مرتبط با فعالیت‌های تونل‌زنی تاکید می‌کند.

کلمات کلیدی: نشت آب به تونل، آب‌های زیرزمینی، بهینه‌سازی، الگوریتم‌های فراابتکاری.