Technical Journal

A Comparative Study of Supervised Machine Learning Techniques for Water Leakage Prediction in Drill and Blast Excavated Water Tunnels in the Himalayan Region

Jyoti Thapa[1], Biplove Ghimire[2,*], Bikram Bhusal[2]

[1]*Post Graduate Student, M.Sc. in Structural Engineering, School of Engineering, Pokhara University, Pokhara, Nepal*

[2]*Post Graduate Student, M.Sc. in Rock and Tunnel Engineering, Pashchimanchal Campus, IoE, TU, Nepal*

**Corresponding email: <biplovegmeray@gmail.com>*

Received: January 15, 2024 Revised: May 15, 2024 Accepted: June 15, 2024

Abstract

In the Himalayan region, drill and blast (DB) method excavated water tunnels often pass through complex geological rock mass conditions which were formed with frequent tectonic movements. Due to these tectonic movements, the rock mass conditions in the Himalayas are highly faulted, folded, jointed, sheared, and fractured. The geological formations are the pathway for the water ingress and leakage out in the water tunnel. This water leakage in the tunnel causes a complicated geological hazard, which significantly increases tunnel instability and can lead to a delay in completion time and finally increase the cost of the tunnel project. Therefore, an efficient water ingress/leakage prediction model is essential to mitigate these challenges. In this research, various field data such as rock mass properties, topography, and permeability data sets were collected from the Nilgiri-II Hydroelectric project water tunnel. These real-time field datasets have been used for comprehensive assessment and to predict the water ingress/leakage in the water tunnel by using four supervised machine learning (ML) approaches such as Support Vector Regression (SVR), Decision Tree (DT) regression, K-Nearest Neighbors (KNN) and Random Forest (RF) regression models. It was observed that the KNN shows the best regression performance of 93% followed by RF of 92%, DT of 86%, and SVR of 66%. Therefore, all these machine learning approaches show good performance in predicting water ingress/leakage based on real field data except the SVR model.

Keywords: *Himalayan Region, Machine Learning, Permeability, Rock Mass, Water leakage*

1. Introduction

The tunnel is an artificial underground passage constructed without disturbing the ground surface by various excavation methods. The drill and blast (DB) method is a versatile and widely used underground construction work. This method has low initial investment, and it is very beneficial for the full face and staged excavation, construction of any shape and size of the underground opening, varying geological conditions, and low to high-strength rock with a fast start-up time. However, this method has some limitations such as low advance rate, high explosive cost, high overbreak, and high rock mass disturbance, high explosive cost (Katuwal & Adhikari, 2023; Kolymbas, 2005; Tatiya, 2013; Zou, 2016). This DB method is mostly used in underground construction work Nepal Himalayas. Due to the high rock mass disturbance and high overbreak of this method, there is a higher possibility of water ingress/leakage in a water tunnel. Therefore, this increases the water tunnel instabilities during construction, and it requires careful planning, skilled personnel, and adherence to safety protocols to ensure successful tunnel construction for hydropower infrastructure (Katuwal et al., 2023).

Nepal is highly vulnerable to active tectonic zones due to its location lies between the Indian plate on the south and the Eurasian plate on the north. Frequent collision of these plates has resulted in frequent and massive earthquakes which has a significant effect on geology. Thus, the geology of Nepal and the Himalayas is very young. Due to this tectonic movement rock mass in the Himalayas is highly fractured and deeply weathered which requires considerable temporary rock support to be installed during excavation (Basnet, 2013; Chaudhary, 2022; Panthi, 2006; Panthi & Shrestha, 2018). The geological environment comprises not only the rock mass, but it is also accompanied by water. So, tunneling is not only the art of understanding the rock mass function but it is also the activity of rock mass and the water present in underground. The ingress of water in the tunnel significantly affects the overall stability and integrity of the tunnel (Katuwal et al., 2024; Panthi & Nilsen, 2010). Therefore, an efficient water ingress prediction model is essential to mitigate the challenges and enhance the stability of the water tunnel in the Himalayan region.

In the past, various approaches have been applied for leakage prediction in the tunnel including analytical, empirical, and numerical applications (Holmøy, 2008; Panthi & Nilsen, 2010). These approaches show results in acceptable limits. However, it has always been a challenging task to accurately predict the amount of water ingress/loss from the water tunnel valley. To minimize the limitations of these approaches, a comprehensive data-based approach is essential. A detailed study of rock mass properties, topographical structures, and permeability factors was conducted in a data-driven approach with the application of machine-learning approaches (Katuwal et al., 2024; Mahmoodzadeh et al., 2023).

In Himalayan geology, for the prediction of water ingress/leakage, various approaches such as analytical, empirical, and numerical applications have been mostly applied. In this study, various robust supervised machine learning approaches such as Support Vector Regression (SVR), Decision Tree (DT) regression, K-Nearest Neighbors (KNN), and Random Forest (RF) regression models are used to predict water ingress/leakage. Based on the comparison between the performance of these ML approaches, the efficient and best prediction model is selected to predict the water ingress/leakage in the water tunnel in the Himalayan geological region. So, this research focuses on a thorough investigation of these data with AI learning for predicting the leakage obtained from the Nilgiri-II Hydroelectric Project.

2. Project Background

2.1 Salient Features of Nilgiri-II

Scheme Type of Nilgiri-II is RoR (Cascade development) having a gross head of 789.75 m and a mean annual discharge of 17.15 m3/s with a capacity of 71 MW. Intake has been installed at the side with several openings one of which has the size of 1.5 m x 3 m. The descending basin is designed as an intermittent flushing system with a single bay and double hooper to trap the silt and sediments flowing from the intake. The headrace tunnel (HRT) is approximately 4.25 kilometers long and has an inverted

D-shape with a 10.5 square meters cross-section. A surge tank of height 36.25 m was built to avoid excessive water hammer pressure. To facilitate the tunnel and shaft construction, four Adit tunnels of cross-section of 10.5 square meters were excavated.

2.2 Project Geology

This project area is located near the Main Central Thrust (MCT), which is one of the major fault systems of the Himalayas. Rock mass in this region is mainly characterized by Jurassic metamorphic and sedimentary rocks consisting coarse coarse-grained mica gneiss with garnet, kyanite or silimanite, migatite, and quartzite around Nilgiri-II. The rock types in the project area are mainly banded gneiss with inter-bedded bands of quartz. The rock mass contains light grey to light green, fine to coarse grain, and medium to thick banded gneiss. It contains three joints and one random joint with medium to high persistency i.e. 1 to 10 meters and open to tight aperture i.e. 0-2 mm infilled with sand silt and clay. The rock quality designation (RQD) is fair, and the surface is rough and planar. Groundwater condition is seen as wet to dripping.

In this project area, the foliation planes are generally striking towards the northwest direction and dipping towards the northeast. The area is also influenced by several shear bands and weak zones are represented by crushed zones. Glacial-fluvial deposits were also observed at the inlet of Nilgiri-I and in the vicinity of Nilgiri II which was cemented strongly where no support or any other further treatment was required. Figure 1 illustrates the detailed description of the geological conditions of this project area.

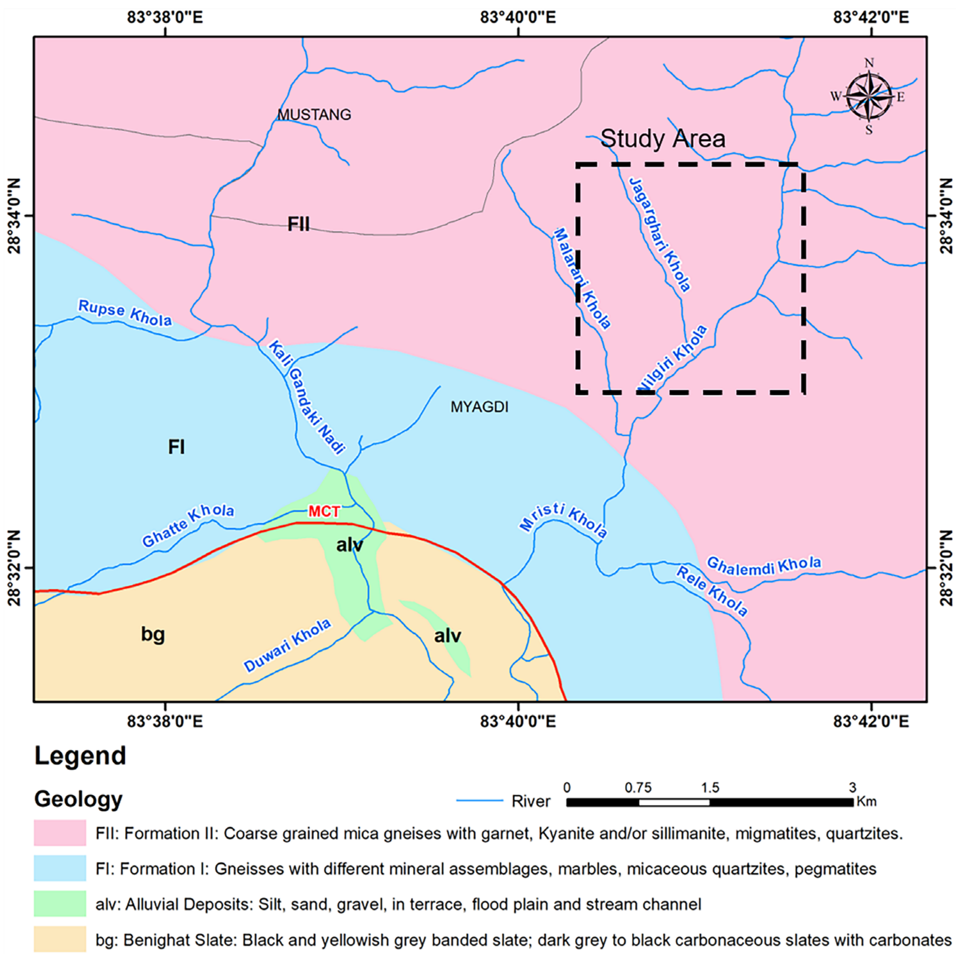

Figure 1: Geology of the Nilgiri-II HEP

3. Research Methodology

Figure 2 is a flowchart that shows the stepwise process involved in the research process, from data collection to model evaluation and selection. Firstly, the required field data such as rock mass quality, topography, and water inflow were collected from the Nilgiri-II Hydroelectric project water tunnel. The selected data sets may not in as per requirements, thus, the selected data sets are preprocessed with cleaning, transferring, feature selection, data correlation analysis, distribution, and data standardization. After processing, the collected field datasets are split into training and testing sets. The selected four supervised machine learning models such as Support Vector Regression (SVR), Decision Tree (DT) regression, K-Nearest Neighbors (KNN), and Random Forest (RF) regression were trained with their optimal hyperparameter. These train regressions models were tested with the tested dataset and the output results were checked using various statistical indicators such as R-squared (R2), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (RRMSE), Mean Absolute Percentage Error (MAPE), Mean Relative Error (MRE) and Variance Accounted For (VAF). Finally, the efficient and best water ingress/leakage prediction model was recommended with the comparison of performance indicators of each regression model. The details of each step are discussed in the following sections.

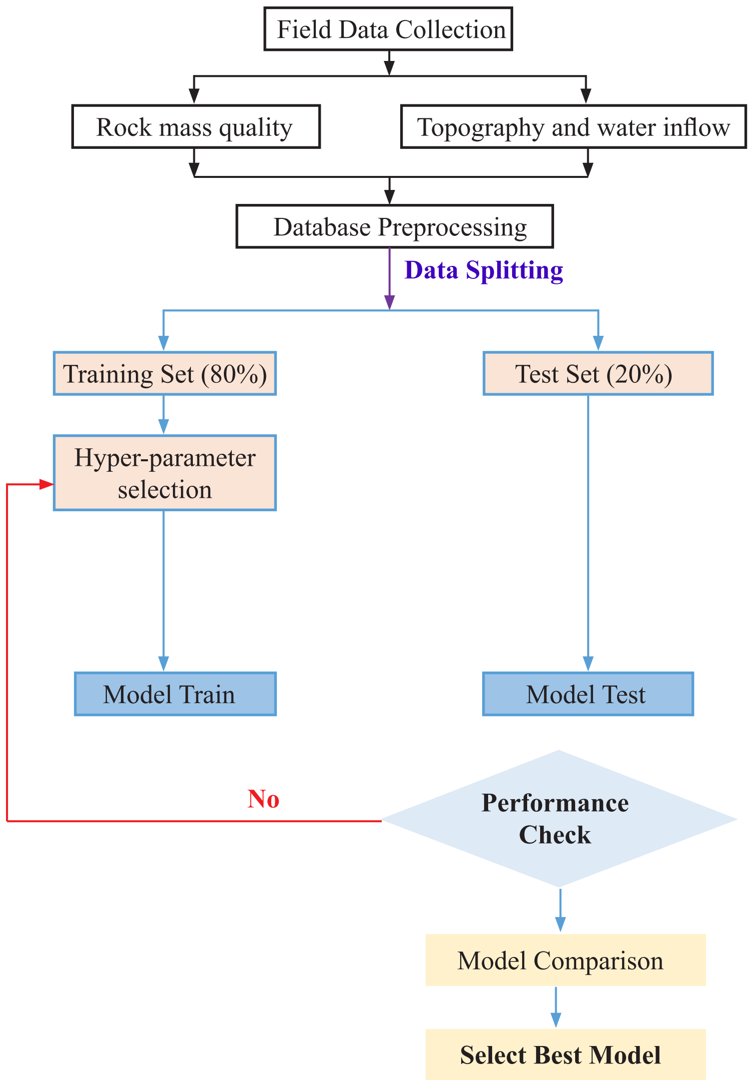

Figure 2: Workflow for water inflow prediction

4. Machine Learning Techniques

In predicting water leakage in tunnels by using machine learning approaches is significantly influenced by the nature, volume, and characteristics of field data various applications can be used. In this study, four supervised machine learning models as Support Vector Regression (SVR), Decision Tree (DT), K-Nearest Neighbour (KNN), and Random Forest (Rf) are selected to predict the water ingress/leakage in 4.25 Km of headrace tunnel at Nilgiri-II HEP. The appropriate input features and target variables are selected as per the collected field data. The stepwise research methodology has been applied to predict the water ingress/leakage. All four models were trained and tested to evaluate the parameters such as R-squared (R2), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (RRMSE), Mean Absolute Percentage Error (MAPE), Mean Relative Error (MRE), and Variance Accounted For (VAF) and were compared to select the best models among three by assigning the weightage.

4.1 Database Study

In this research, rock mass parameters like Rock Quality Designation (RQD), Number of Joints (Jn), Joint Roughness (Jr), Joint Alteration (Ja), and Q-classification value (Q) were collected. In addition, topographical features such as Hydrostatic head (Hst), shortest perpendicular distance to the valley side (D), and water permeability properties (fa) on the valley side of the tunnel were collected. These independent features were selected as input parameter to determine the target variable (water inflow). Both features have been summarized below in Table 1 and Table 2 with their descriptive statistics for better interpretation. Based on these field parameters, the prediction of the specific leakage (q) was estimated by utilizing Panthi's semi-empirical approach (Panthi & Nilsen, 2010). The initial potential leakage was estimated considering the dry season, natural groundwater lowers down to the level of the headrace tunnel, and fluid flowing in HRT will produce the head and will govern the stretch of water leakage on the valley side of the topography (Basnet & Panthi, 2020).

Table 1: Field datasets for rock mass quality

Description	Count	mean	Std	min	25%	50%	75%	max
RQD	211	50.3	18.8	10	40	55	65	85
Jn	211	11.3	1.5	6	12	12	12	15
Jr	211	1.6	0.3	1	1.5	1.5	2	2
Ja	211	2.8	1.6	1	2	2	4	12
Jw	211	1	0.1	0.5	1	1	1	1
SRF	211	2.2	2.3	1	1	1	2.25	10
Q	211	3.1	3.3	0.01	0.63	2.03	5	18.89

Table 2: Field datasets for topography and water inflow

Description	Count	mean	Std	min	25%	50%	75%	max
Hst	211	23.6	11.9	0.25	13.15	26.31	35.4	36.36
D	211	163.1	72.2	62.1	105.86	140.29	224.525	305.14
fa	211	0	0	0	0.02	0.03	0.05	0.14
q	211	6.5	7.7	0.17	1.375	3.29	8.81	33.68

4.2 Correlation Analysis

Correlation analysis is a statistical tool that has been used to evaluate the strength and direction of multiple feature variables and dependent variables. This correlation coefficient measures that quantifies the strength and direction of the relationship between variables which ranges from -1 to +1. A correlation coefficient of 1 indicates a high correlation relationship, -1 indicates a negative relationship and 0 indicates no relationship. A correlation between variables is visualized in the form of a heat map with color coding inside each cell aid to illustrate the interrelation between independent and dependent variables.

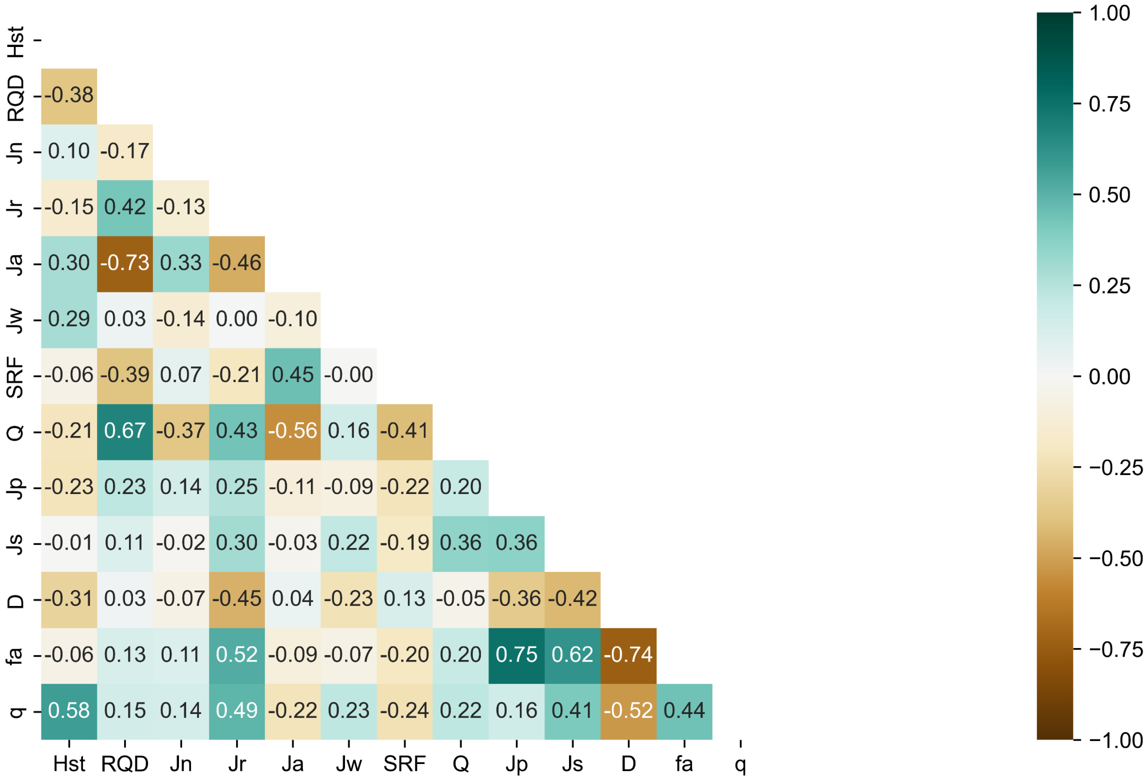

Figure 3: Correlation matrix of datasets

Figure 3 illustrates that the water ingress/leakage is positively correlated with Hst, RQD, Jn, Jr, Jw, Q, Jp, Js, and fa with correlation coefficients 0.58, 0.15, 0.14, 0.49, 0.23, 0.22, 0.16, 0.41, and 0.44 respectively. Likewise, negatively correlated with Ja, SRF, and D with 0.22, 0.24, and 0.52 respectively. These correlations indicate that the selected input features have good correlations with the target variable.

4.3 Data Distribution

The selected input and target viable data distribution is essential to visualize the patterns of datasets. For this purpose, the box plot is used for rock mass, topographical, and specific leakage data, and the violin plot for understanding key summary statistics. Initially, observed datasets were plotted on a box plot which summarized the median, quartiles, and potential outliers of the data as shown in Figure 4. Outliers can be easily identified as points beyond the whisker of the box and the histogram plot displays the data distribution for Jn, Jr, and Ja, which is present in Figure 4. The advantage of a box plot and histogram plot is that it is simple and effective making them a quick and informative visualization.

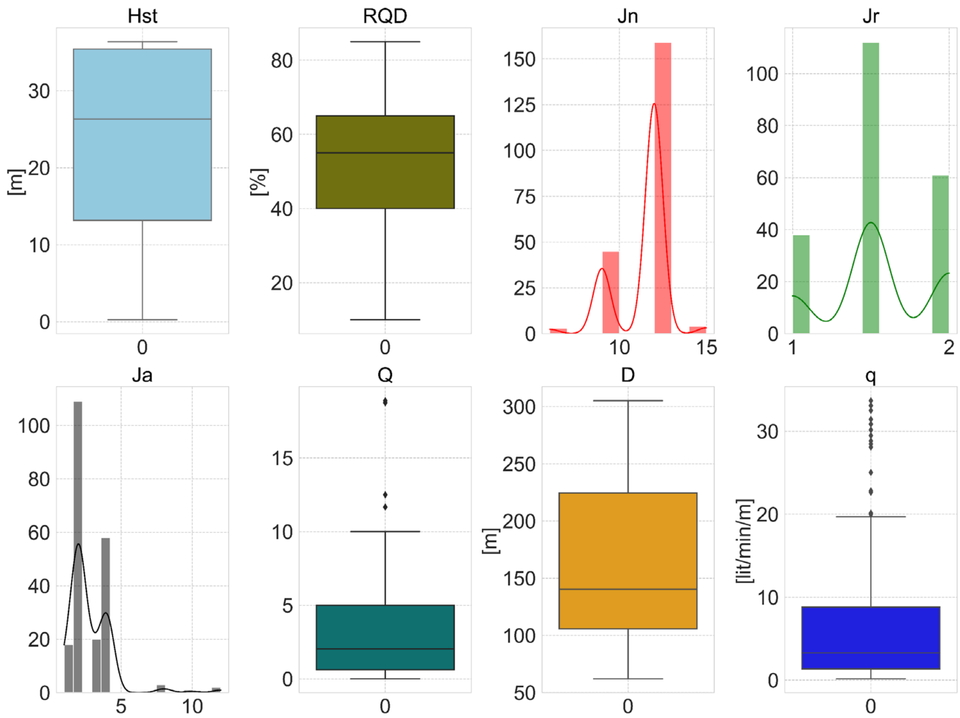

Figure 4: Multivariate data visualization of selected variables

Violin plots offer more detail about the data distribution compared to box plots. These selected datasets are distributed in violin plots as seen in Figure 5. Violin plots combine the aspects of box plots and kernel density estimates, providing insights into the distribution shape. The width of the violin plot at different points represents the estimated probability density of the data. Figure 5 illustrates the distribution of the statistical quartile summary of the filed data set. In these figures, a wider region of density plot indicates the more frequent occurrence and vice versa.

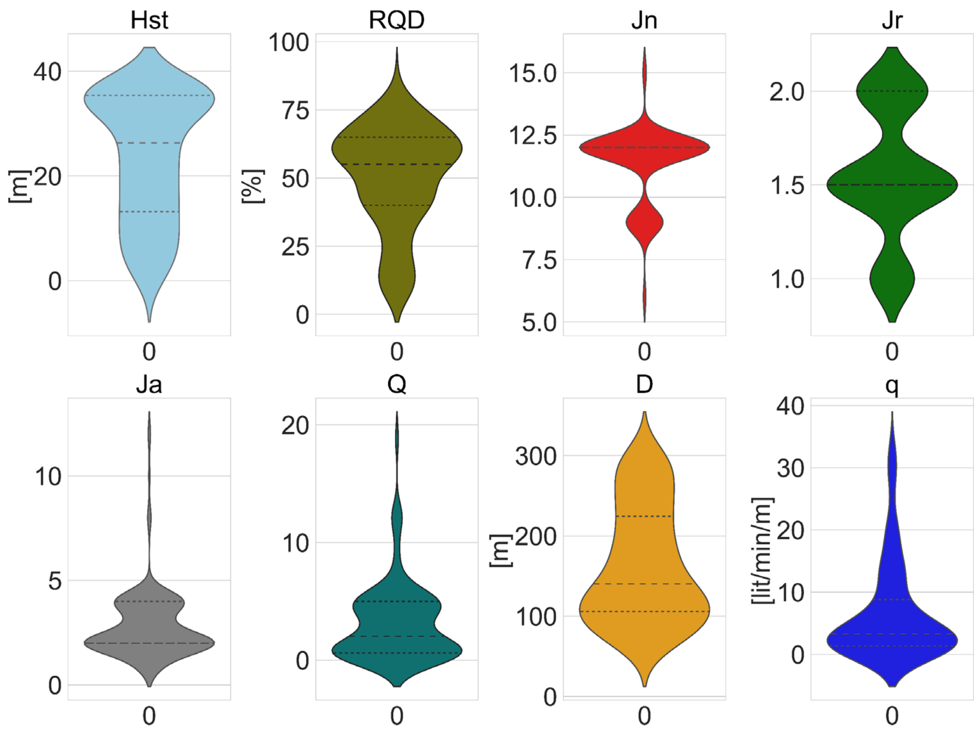

Figure 5: Data distribution in violin plot

A pair plot is a graphical representation that visualizes the correlation between pairs of variables in selected datasets, which is useful to explore the data distribution and correlation between multiple variables simultaneously. This plot is also known as a scatter plot matrix. In this study, this plot is used to identify the patterns, trends, and correlations of selected features with the target variable. A pair plot was generated for the datasets of hydrostatic head (Hst), RQD, Q, D, and q which allows us to interpret the pairwise relationship between variables in the dataset, making it easy to identify patterns, correlations, and distributions. Additionally, a pair plot is more advantageous for exploring the overall structure of a data set before diving into a more detailed analysis. Unusual observations or outliers can be visually identified in scatterplots, aiding in the detection of data points that may have a significant impact on analysis. These benefits can be best illustrated in Figure 6, illustrates the different types of variables related to each other according to the rock mass classes (RMC). For example, to illustrate the correlation between the data of RQD and Q, for the increasing value of RQD, Q value also increases simultaneously for the RMC-II. Similarly, the higher the value of the Hydrostatic head (Hstatic), the more the leakage will be for the RMC-IV. The relationship between Hstatic and q demonstrates that, for RMC-IV, q has been increasing since the hydrostatic head in the water tunnel increased. In RMC-II, q has been dropping indistinguishably while Q has been increasing.

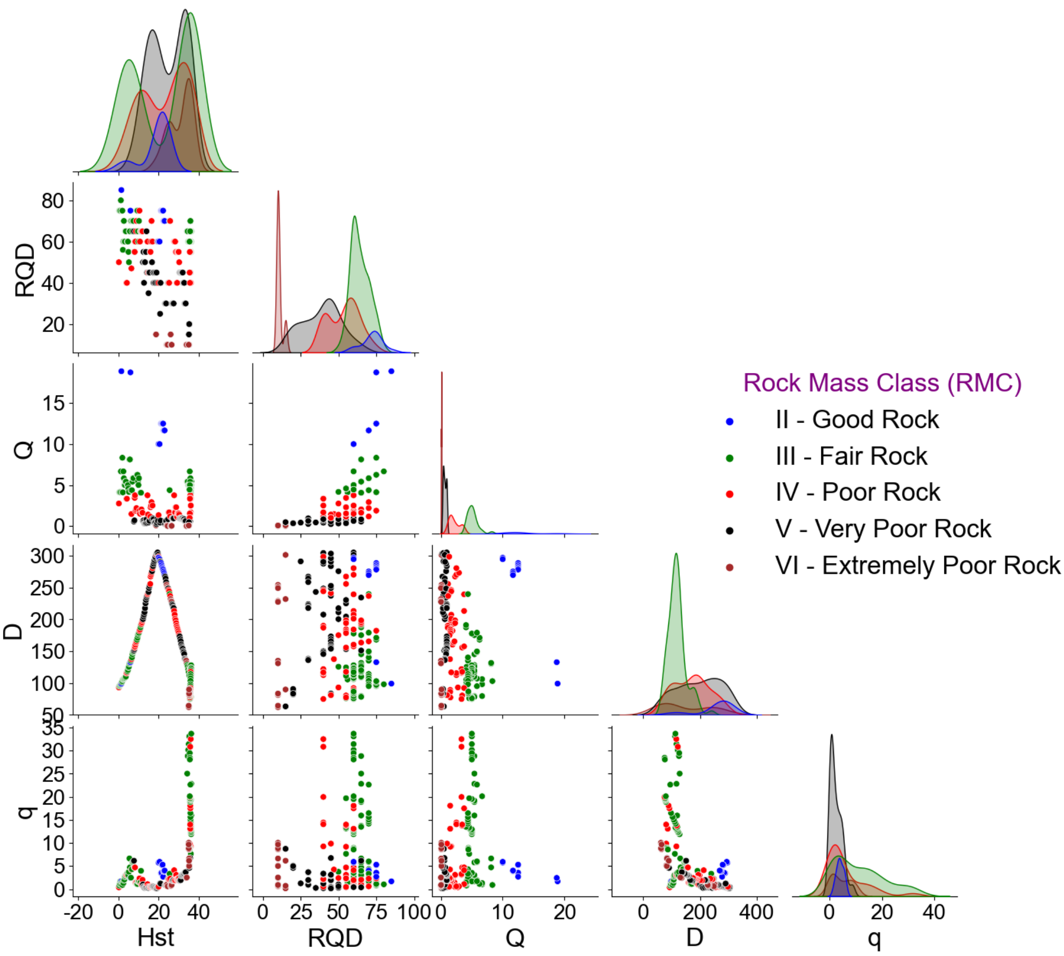

Figure 6: Leakage assessment data relationship

After this data distribution, the selected database contains outlier in some parameters. In tunneling, the removal of outlier may increase the risk and instability in underground tunnel structure. Thus, these outliers were considered during the machine learning regression analysis.

4.4 Data Normalization

Data normalization is also a data pre-processing step in machine learning to scale and transform features within a uniform scale. Normalization ensures that different features are on a similar scale, preventing one feature from dominating others. From Table 1 and Table 2 for the datasets of rock mass, topography, and water leakage, the range of value for each parameter was defined using descriptive statistics. This indicates that the extent of input data is in different scales. Therefore, in this study min-max data normalization method is used to scale the selected datasets. This process increases the prediction performance of machine learning models.

4.5 Statistical Analysis of Selected Model

Statistical analysis is an essential component for detailed understanding and forecasting effective leakage models for hydropower projects headrace tunnels. In this study, various statistical parameters such as R-squared (R2), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (RRMSE), Mean Absolute Percentage Error (MAPE), Mean Relative Error (MRE) and Variance Accounted For (VAF) are used and these are calculated using equation 1 to 8 respectively.

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

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}[{}_i^aY - {}_i^pY] \tag{2}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}[{}_i^aY - {}_i^pY]^2 \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}({}_i^aY - {}_i^pY)2} \tag{4}$$

$$\text{RRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{{}_i^aY - {}_i^pY}{{}_i^aY}\right)^2} \tag{5}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{{}_i^aY - {}_i^pY}{{}_i^aY}\right| * 100\% \tag{6}$$

$$\text{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{{}_i^aY - {}_i^pY}{{}_i^aY}\right| \tag{7}$$

$$\text{VAF} = 1 - \left|\frac{\text{var}({}_i^aY - {}_i^pY)}{\text{var}({}_i^aY)}\right| * 100\% \tag{8}$$

Where actual and predicted values of variables are represented by (_1^a)Y and (_1^p)Y respectively, and n in the equations is the total number of datasets that are used in selected machine learning models.

5. Water Inflow Prediction Model

In this research, water intrusion in the headrace tunnel of Nilgiri-II is assessed and predicted using the Support Vector Regression (SVR), Decision Trees (DT), K-Nearest Neighbors (KNN) models, and Random Forest (RF) in Anaconda version 3.6 using Python computing.

5.1 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a machine learning algorithm used for regression and classification tasks. It belongs to the family of Support Vector Machines (SVM) and is particularly effective for handling non-linear relationships between input features and target variables. In SVR hyperplane represents the relationship between the input datasets and target variable. In this paper target variable is water leakage from the headrace water tunnel of 4.2 Kilometers. It links the capacities of support vector machines to hydropower tunnel modeling and prediction, assisting rock engineers in making educated decisions for efficient handling and mitigation of water intrusion issues (Katuwal et al., 2024). Figure 7 illustrates actual and predicted inflow by the application of the SVR model.

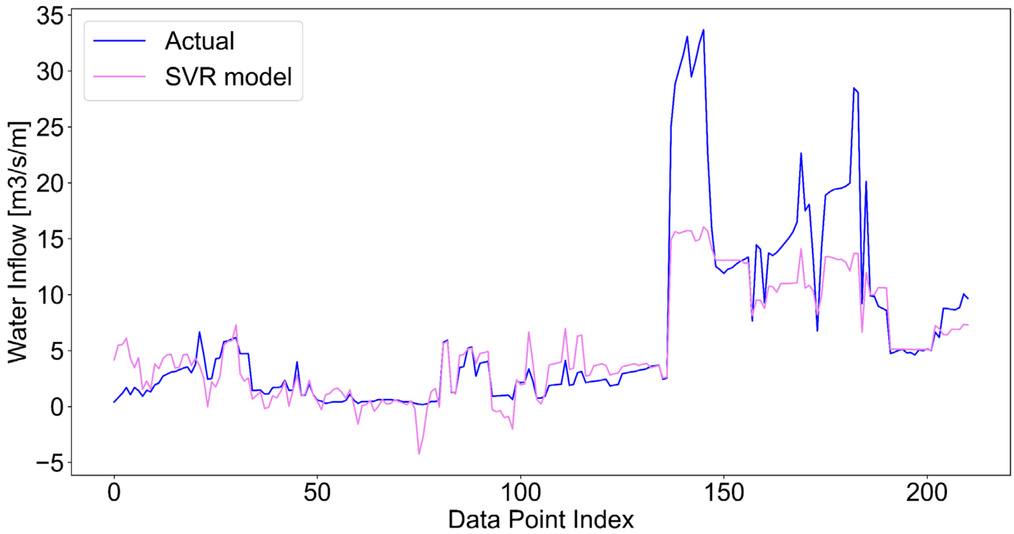

Figure 7: SVR model for water leakage prediction

Similarly, Table 3 depicts the statistical indices R2, MAE, MSE, RMSE, RRMSE, MAPE, MRE, and VAF which are evaluated as 0.66, 2.82, 24.16, 4.91, 0.59, 116.06, 69.70, and 66.47, respectively. This indicated that the SVR models have poor performance for accurately predicting the ingress/leakage of water as shown in Figure 8.

Table 3: A summary of statistical indices of the SVR model

R2	MAE	MSE	RMSE	RRMSE	MAPE	MRE	VAF
0.66	2.82	24.16	4.91	0.59	116.06	69.70	66.47

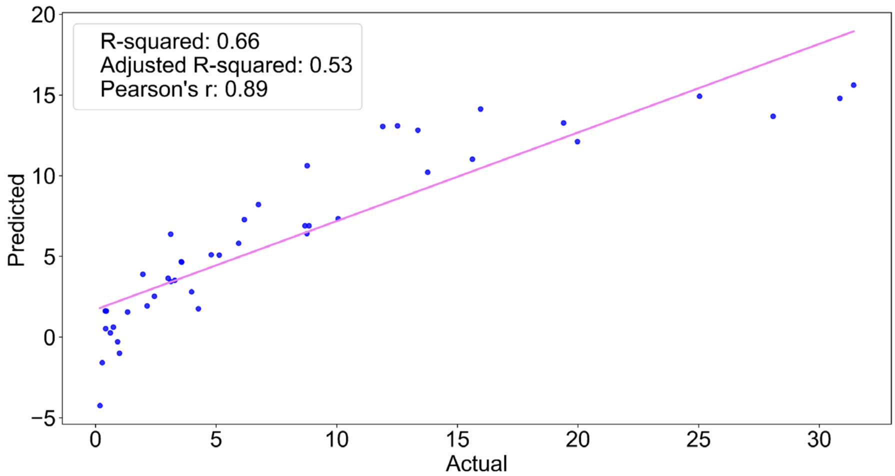

Figure 8: Relationship between actual and predicted leakage by SVR model.

5.2 Decision Tree (DT)

Decision trees can be useful in predicting water leakage in tunnels through a process of analyzing various factors that may contribute to indicating potential water ingress/leakage. These various factors have been described above in data sets as rock mass, topography, and permeability factors. The effectiveness of a decision tree model must depend on the quality and representatives of the training data. Figure 9 illustrates the comparison between the actual field data set and the water inflow prediction by using this DT model.

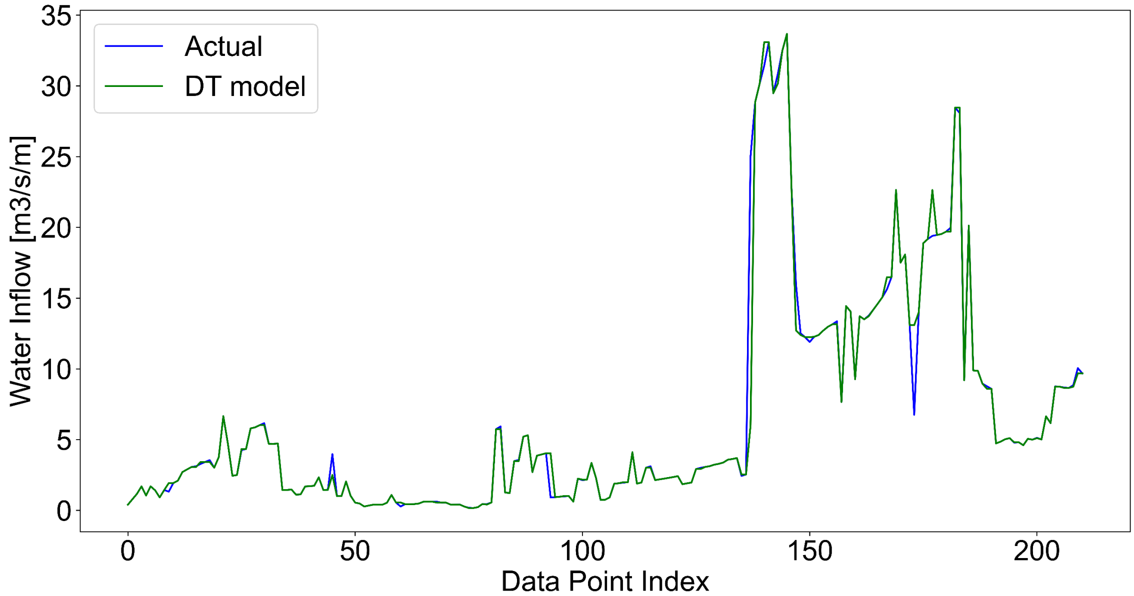

Figure 9: DT model for water leakage prediction

Table 4: A summary of statistical indices of the DT model

R2	MAE	MSE	RMSE	RRMSE	MAPE	MRE	VAF
0.86	1.02	10.22	3.19	0.39	19.23	-10.19	85.81

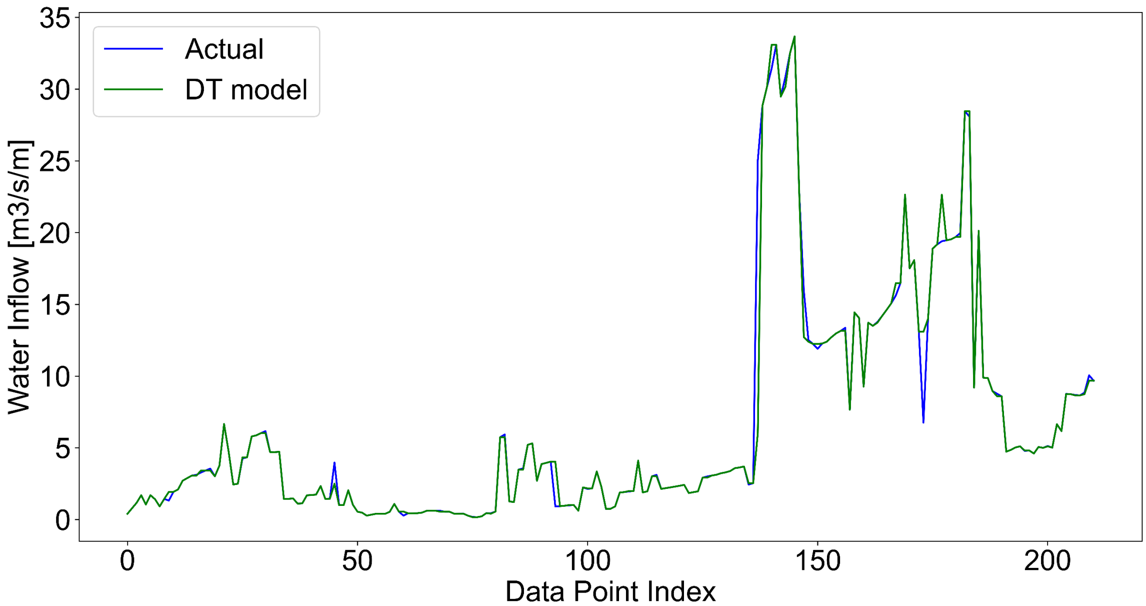

Figure 10: Relationship between actual and predicted leakage by DT model.

Similarly, Table 4 visualizes the statistical indices R2, MAE, MSE, RMSE, RRMSE, MAPE, MRE, and VAF which are evaluated as 0.86, 1.02, 10.22, 3.19, 0.39, 19.23, -10.19, and 85.81, respectively. This indicated that the DT models are capable of accurately predicting the ingress/leakage of water. The statistical indices and Figure 10 indicate that this model presents a good correlation with the features and dependent variables.

5.3 K-Nearest Neighbour (KNN)

KNN is based on the idea that data points with similar features tend to belong to the same class or have similar values. The algorithm classifies a new data point by comparing it to the k-nearest neighbor in the training data set. "K' represents the number of neighbors considered when making a prediction. Figure 11 illustrates the comparison between the actual field data set and the KNN prediction model for water inflow in the case tunnel project. Table 5 represents the statistical parameters R2, MAE, MSE, RMSE, RRMSE, MAPE, MRE, and VAF which are calculated as 0.93, 1.31, 5.17, 2.27, 0.27, 24.22, -6.57, and 92.81, respectively. This model shows a decent association with the characteristics and dependent variables, according to the statistical indices. All of the statistical results and Figure 12 shows that the KNN models are suitable and have a good capacity for predicting the loss of water in hydropower tunnels.

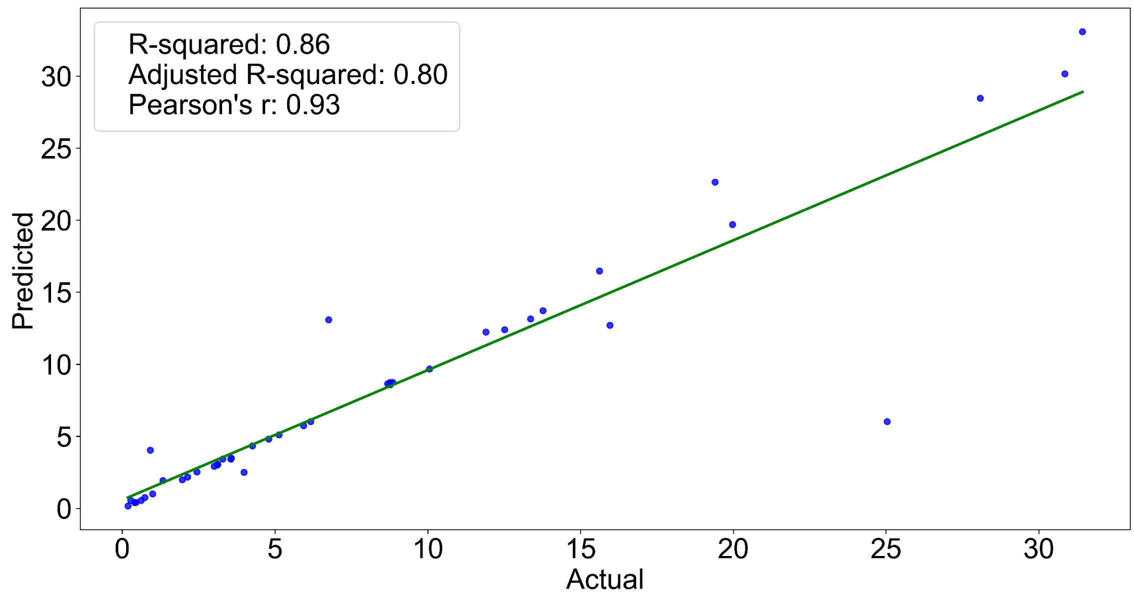

Figure 11: KNN model for water leakage prediction

Table 5: A summary of statistical indices of the KNN model

R2	MAE	MSE	RMSE	RRMSE	MAPE	MRE	VAF
0.93	1.31	5.17	2.27	0.27	24.22	-6.57	92.81

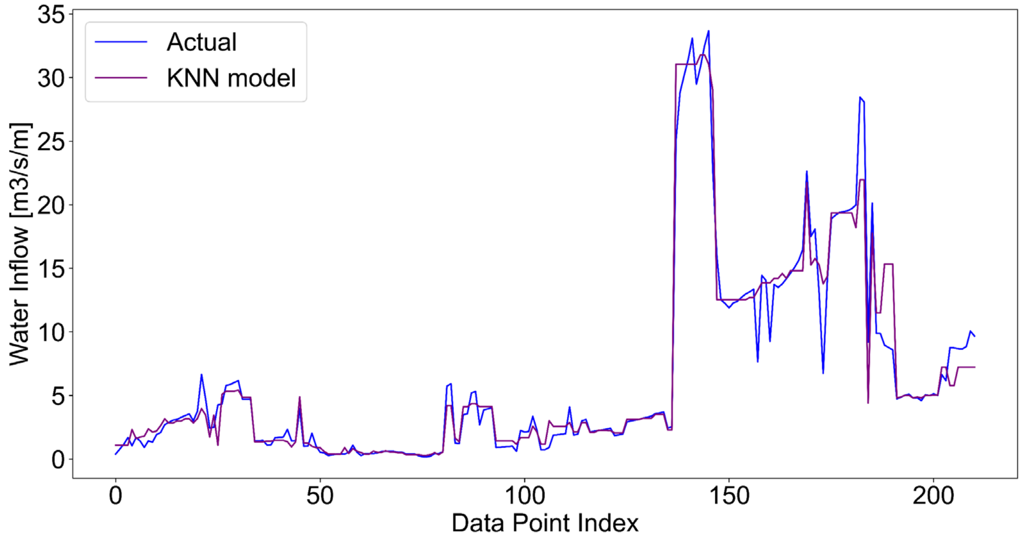

Figure 12: Relationship between actual and predicted leakage by KNN model.

5.4 Random Forest (RF)

Random forest is an ensemble learning method used for both classification and regression tasks. It builds multiple decision trees and merges them to get a more accurate and stable prediction. Each tree is trained on a random subset of the data and features, reducing overfitting and improving generalization. Figure 13 illustrates the comparison between the actual field data set and the RF prediction model for water inflow in the case tunnel project, which shows good fitting with actual water inflow. Also, Table 6 represents the statistical parameters R2, MAE, MSE, RMSE, RRMSE, MAPE, MRE, and VAF which are calculated as 0.92, 0.99, 6.10, 2.47, 0.30, 13.12, -2.26, and 91.52 respectively. The statistical indices and Figure 14 demonstrate that it has a respectable relationship between the attributes and dependent variables in this model. Every one of the findings demonstrates that the KNN models are appropriate and highly capable of forecasting water loss in hydroelectric tunnels.

Random forest builds multiple decision trees and merges their predictions. In addition, robust to overfitting, handles high-dimensional data well, and provides feature importance. So, random forest application also has been chosen to evaluate the prediction and finally compared with other applications. Random forest is often chosen for its versatility, while SVR and KNN may be preferred for certain types of data or problems, and decision trees are chosen for their simplicity and interpretability.

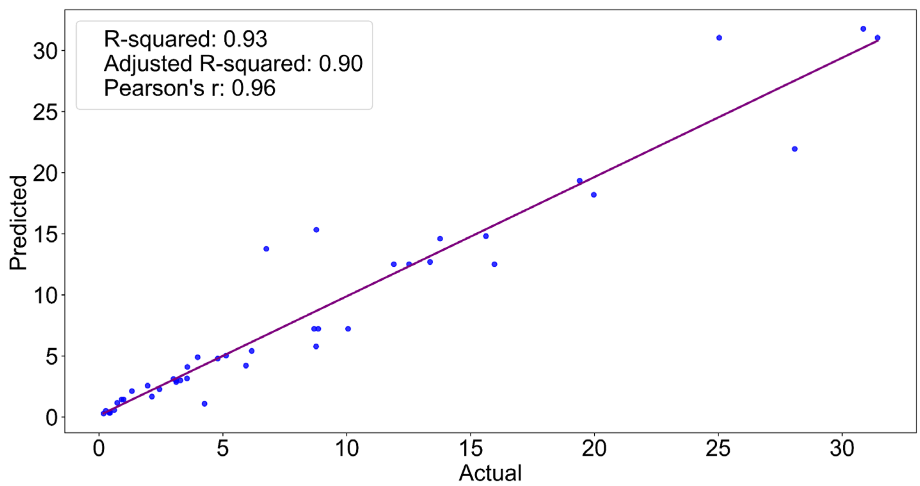

Figure 13: RF model for water leakage prediction

Table 6: A summary of statistical indices of the RF model

R2	MAE	MSE	RMSE	RRMSE	MAPE	MRE	VAF
0.92	0.99	6.10	2.47	0.30	13.12	-2.26	91.52

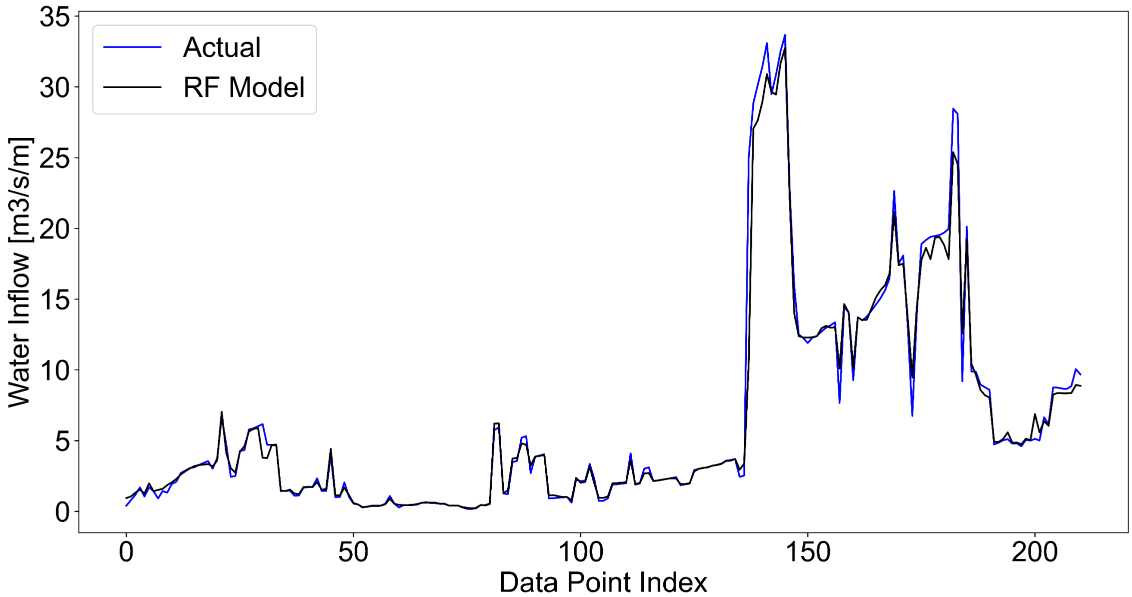

Figure 14: RF model for water leakage prediction

6. Result Comparison of Water Leakage Prediction

In this study, four machine learning regression models such as Support Vector Regression (SVR), Decision Tree (DT) regression, K-Nearest Neighbors (KNN), and Random Forest (RF) regression models were used to predict the water leakage prediction in the 4.2 Km headrace tunnel of Nilgiri-II HEP considering the rock mass properties, topographical, and specific leakage datasets. The target variable (water inflow/leakage) was predicted and the output result from SVR, DT, KNN, and RF regression models were analyzed with statistical measures and loss functions. All these outcomes were compared to actual datasets to measure the accuracy. All the ML applications output shows satisfactory results individually except SVR. To achieve this, statistical indices and loss functions were used to evaluate and categorize the performance of each model as good, better, or best, with respective weightings of 2,4,6 and 8. The overall ranking of each model was determined by adding the assigned weightage for respective statistical parameters. Table 7 illustrates the comparative ranking of selected models.

Particularly, Machine learning models depend upon the nature of the dataset available and the problem at hand. When these four models were compared, KNN resulted in the maximum weightage as shown in Table 7, followed by RF, DT, and SVR models.

Table 7: Water ingress prediction model comparison

Parameter/Model		SVR	DT	KNN	RF
R2	Value	0.66	0.86	0.93	0.92
	Weightage	2	4	8	6
MAE	Value	2.82	1.02	1.31	0.99
	Weightage	2	4	6	8
MSE	Value	24.16	10.22	5.17	6.10
	Weightage	2	4	8	6

RMSE	Value	4.91	3.19	2.27	2.47
	Weightage	2	4	8	6
RRMSE	Value	0.59	0.39	0.27	0.30
	Weightage	2	4	8	6
MAPE	Value	116.06	19.23	24.22	13.12
	Weightage	2	6	4	8
MRE	Value	69.7	-10.19	-6.57	-2.26
	Weightage	2	4	6	8
VAF	Value	66.47	88.81	92.81	91.52
	Weightage	2	4	8	6
Total Weightage		20	34	56	54
Rank		4	3	1	2

7. Conclusions

The real construction datasets such as permeability, topographic characteristics, and rock mass attributes were collected from Nilgiri-II HEP for a comprehensive assessment of water ingress/leakage in a hydropower tunnel in the Himalayan region. These characteristics were utilized in machine learning to forecast water ingress/leakage over the 4.2 km headrace tunnel. To anticipate the leakage/outflow from the valley side of the tunnel, Support Vector Regression (SVR), Decision Tree (DT), K-nearest neighbor (KNN), and Random Forest (RF) models were used. Before modeling, it was looked into how different factors moved the heatmap for the statistical correlation between various inputs in a multicollinear way and the study's findings lead to the following conclusion.

- The specific discharge is selected as the target variable and it has a good positive correlation with Hst, RQD, Jn, Jr, Jw, Q, Jp, Js, and fa with correlation coefficients 0.58, 0.15, 0.14, 0.49, 0.23, 0.22, 0.16, 0.41, and 0.44 respectively. Likewise, negative correlation with Ja, SRF, and D with 0.22, 0.24, and 0.52 respectively. Therefore, these parameters significantly influenced the quantity of water ingress/leakage in the water tunnel in the Himalayan region.
- The selected machine learning (ML) models establish quite good water ingress/leakage prediction capabilities individually except the SVR model.
- Based on the statistical indicators, the KNN regression model shows the highest R-squared value of 93%. Moreover, the RF regression model shows 92%, the DT regression model shows 86% and the SVR model shows 66 % prediction capabilities.

This study suggests that the real-time data-driven machine learning approach has good capabilities to predict the water ingress/leakage in the water tunnel and, thus, can be used in future research work and real tunnel construction projects.

References

Basnet, C. B. (2013). Evaluation on the Squeezing Phenomenon at the Headrace Tunnel of Chameliya Hydroelectric Project, Nepal. Theses at NTNU. *Norwegian University of Science and Technology, Trondheim, Norway.*

Basnet, C. B., & Panthi, K. K. (2020). Leakage potential through a shotcrete lined high-pressure headrace tunnel - An analysis on a case from Nepal. *Rock Mechanics for Natural Resources and Infrastructure Development- Proceedings of the 14th International Congress on Rock Mechanics and Rock Engineering*, ISRM 2019, September, 722–728.

Chaudhary, B. (2022). Assessment on plastic deformation at the powerhouse cavern and tailrace tunnel of andhikhola hydroelectric project tailrace tunnel of andhikhola hydroelectric project background. Theses at NTNU. Norwegian *University of Science and Technology, Trondheim, Norway.*

Holmøy, K. H. (2008). Significance of geological parameters for predicting water leakage in hard rock tunnels. Ph.D. Theses at NTNU 2008:291. *Norwegian University of Science and Technology, Trondheim, Norway.*

Katuwal, T. B., & Adhikari, S. (2023). Fracture Mechanics Induced by Drilling and Blasting in Underground Openings: Prevailing Practices and Studies. *Technical Journal*, 3(1), 134–146.

Katuwal, T. B., Panthi, K. K., & Basnet, C. B. (2023). Challenges associated with the construction of vertical and inclined shafts in the Himalayan Region. *15th ISRM Congress 2023 & 72nd Geomechanics Colloquium, Schubert & Kluckner (eds.) © ÖGGAt: Salzburg Congress*, Austria, pp. 1199-1204.

Katuwal, T. B., Panthi, K. K., Basnet, C., & Adhikari, S. (2024). Leakage Prediction and Post-Grouting Assessment in Headrace Tunnel of a Hydropower Project. Yan, J., Celestino, T., Thewes, M., & Eberhardt, E. (Eds.), *Tunnelling for a Better Life (1st ed.). Proceedings of the ITA-AITES World Tunnel Congress on CRC Press*, WTC 2024, Shenzhen, pp. 3044-3052.

Kolymbas, D. (2005). Tunnelling and Tunnel Mechanics:A rational approach to tunnelling. *In Springer Science & Business Media.*

Mahmoodzadeh, A., Ghafourian, H., Hussein Mohammed, A., Rezaei, N., Hashim Ibrahim, H., & Rashidi, S. (2023). Predicting tunnel water inflow using a machine learning-based solution to improve tunnel construction safety. *Transportation Geotechnics*, 40, 100978.

Panthi, K. (2006). Analysis of Engineering Geological Uncertainties Related to Tunnelling in Himalayan Rock Mass Conditions. Ph.D. Theses at NTNU 2006:41. *Norwegian University of Science and Technology.*

Panthi, K. K., & Nilsen, B. (2010). Uncertainty analysis for assessing leakage through water tunnels: A case from Nepal Himalaya. *Rock Mechanics and Rock Engineering*, 43(5), 629–639.

Panthi, K. K., & Shrestha, P. K. (2018). Estimating Tunnel Strain in the Weak and Schistose Rock Mass Influenced by Stress Anisotropy: An Evaluation Based on Three Tunnel Cases from Nepal. Rock Mechanics and Rock Engineering, 51(6), 1823–1838.

Tatiya, R. R. (2013). Surface and Underground Excavations: Methods, Techniques and Equipment (2nd edition). London: *CRC Press.*

Zou, D. (2016). Theory and technology of rock excavation for civil engineering. *Springer.*