Journal of Engineering Issues and Solutions

Ensemble Machine Learning and Model Interpretability for Leakage Prediction in Hydraulic Tunnels

Biplove Ghimire*

Post Graduate Student, M.Sc. in Rock and Tunnel Engineering, Paschimanchal Campus, IOE, Tribhuvan University, Nepal

**Corresponding email: biplovegmeray@gmail.com*

Received: May 5, 2025; Revised: July 5, 2025; Accepted: July 12, 2025

https://doi.org/10.3126/joeis.v4i1.81570

ABSTRACT

Drill-and-blast tunnel construction in the Himalayan region often encounters complex geological and hydrogeological conditions, leading to significant water leakage that impacts project cost and stability. This study aims to enhance leakage prediction accuracy using ensemble machine learning techniques. Initial leakage estimates were made using Panthi's semi-empirical approach for the Nilgiri-II Hydropower Project. A dataset comprising rock mass quality, topography, and permeability features was used to train four ensemble models: Bagging, Boosting (XGBoost), Voting, and Stacking. Among these, Bagging outperformed others with an R^2 of 0.99, followed by Voting and Stacking (both $R^2 = 0.97$). Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE) plots were used to interpret model predictions and identify key influencing features such as hydrostatic head (Hstatic), distance to the valley side (D), and joint parameters. These results demonstrate that ensemble learning, particularly bagging, is highly effective in modeling water leakage in challenging Himalayan tunnel environments.

Keyword: *water leakage, rock mass, ensemble machine learning, k-fold CV.*

1. Introduction

The drill-and-blast method is the most widely used technique for tunnel excavation. Over time, advancements in explosives, delay systems, and the understanding of rock fragmentation have significantly improved this method. It is applicable across a wide range of geological conditions (Satici, 2006). There are two sides to the coin: due to vibrations caused by blasting and overbreak increase the possibility of water ingress and leakage in the water tunnel. Therefore, these possibilities 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 & Panthi, 2023).

Geologically, Nepal is situated between the Eurasian tectonic plate on the north and the Indian plate on the south. In the past, multiple collisions between these plates have resulted in large earthquakes up to

7.9 magnitude in the latest records. These earthquakes significantly affect the geology; in Nepal's case, the young and fragile geology exacerbates the impact. 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 (Panthi, 2006). The geological environment includes not only the rock mass but also water, which together influence tunnel conditions. These hydro-geological combinations during the excavation of the tunnel makes the advancement difficult which has a significant effect on tunnel integrity and overall stability. Therefore, to lessen the difficulties and improve the rigidness of the hydraulic tunnel in the Himalayan region, an effective water ingress prediction model is crucial (Singh & Goel, 2011).

There are various approaches analytical, empirical, semi-empirical, and numerical approaches for predicting the leakage in hydraulic tunnels which also have been used in the past. Predicting leakage in the water tunnel on the valley side is however a challenging task. Initially, Panthi's semi-empirical approach has been utilized for specific leakage prediction. To minimize the limitations of this approaches data-driven broad study is vital for the accurate prediction of leakage. Features like rock mass properties and topographical and permeability properties were utilized for the prediction of leakage using an ensemble machine learning approach.

In this study, ensemble machine learning approaches such as Bagging, Boosting, Voting and Stacking were used instead of individual supervised machine learning methods for predicting leakage. Subsequently, statistical indices were applied to differentiate the performance of each ensemble approach, where each model interpretation included PDP (Partial Dependence Plots) and ICE (Individual Conditional Expectation) plots (Molnar, 2022).

2. Project Background

2.1 Salient features

Nilgiri Khola Hydropower-II is located at around 2400 m elevation in Annapurna Rural Municipality-4, Myagdi district, approximately 110 km northwest of Pokhara in the Higher Himalayas. Originating from the 7061 m Nilgiri Mountain, the project comprises Nilgiri I (41 MW) and Nilgiri II (76 MW), totaling 117 MW. Nilgiri-II is a Run-of-River (RoR) scheme with a mean annual discharge of 17.15 m^3/s and a gross head of 789.75 m. It features three 1.5 m × 3 m side intakes, a double-bay intermittent flushing desanding basin, and a 2.25 km inverted D-shaped HRT (10.5 m^2 cross-section). A 36.25 m tall surge tank mitigates water hammer pressure, and four Adit tunnels have been constructed.

2.2 Geology

The project area lies near the Main Central Thrust (MCT), a major Himalayan fault system. The rock mass mainly consists of Jurassic metamorphic and sedimentary rocks such as marble, mica gneiss with garnet, kyanite or sillimanite, migmatite, quartzite, and banded gneiss with quartz bands. Rocks are light grey to green, fresh to slightly weathered, and infilled with sand, silt, and clay. It contains three joints and one random joint with medium to high persistency (1–10 m) and open to tight apertures (0–>5 mm). Joint surfaces are rough-planar and stable. RQD is fair, and groundwater ranges from dry to wet/dripping. Foliation strikes NW and dips NE. Shear bands and crushed zones indicate structural weakness. Well-cemented glacial-fluvial deposits near Nilgiri-I and Nilgiri-II require no treatment. Figure 1 illustrates geological details (Thapa, Ghimire & Bhusal, 2024).

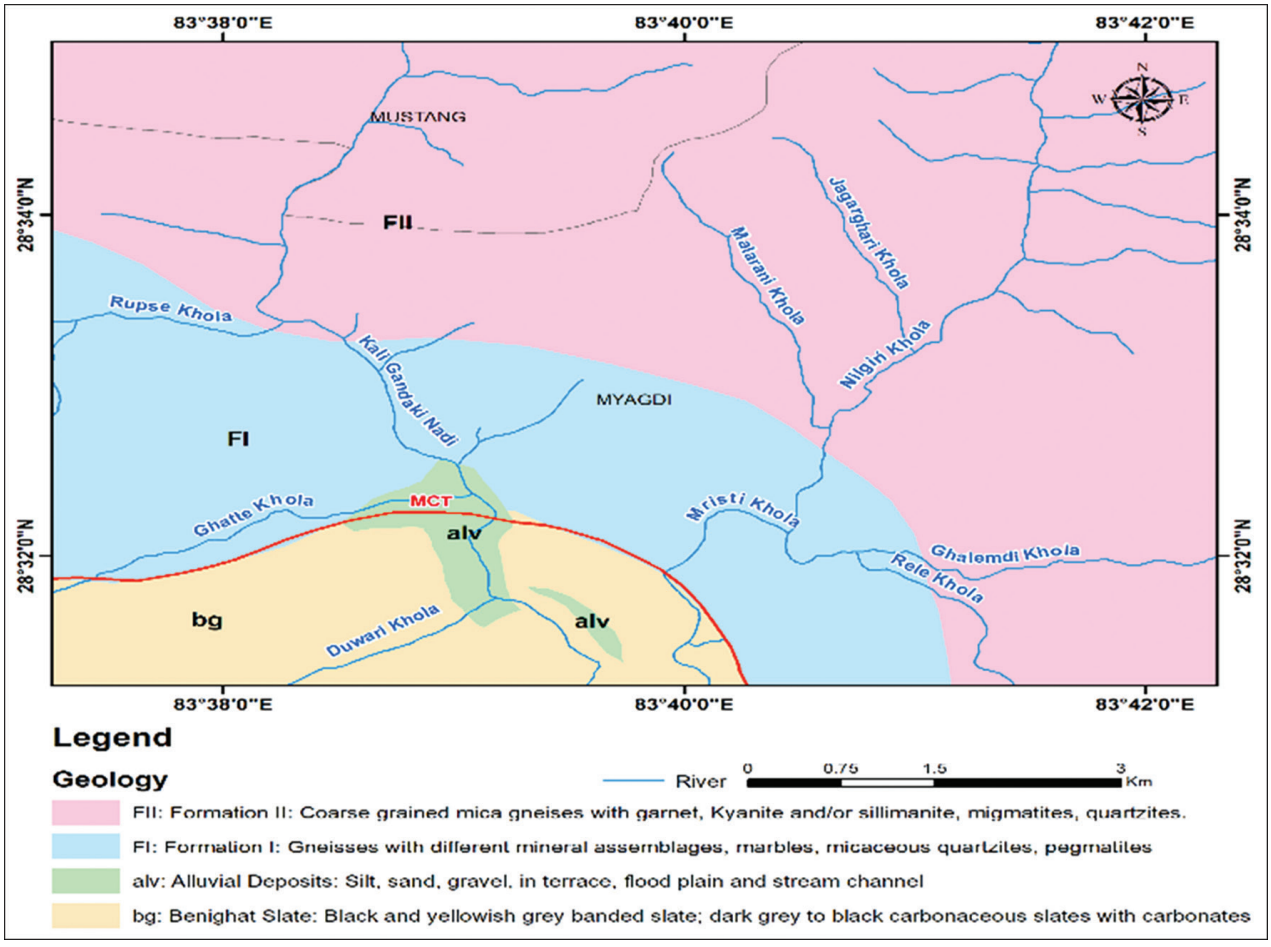

Figure 1: Geology of the project

3. Research Methodology

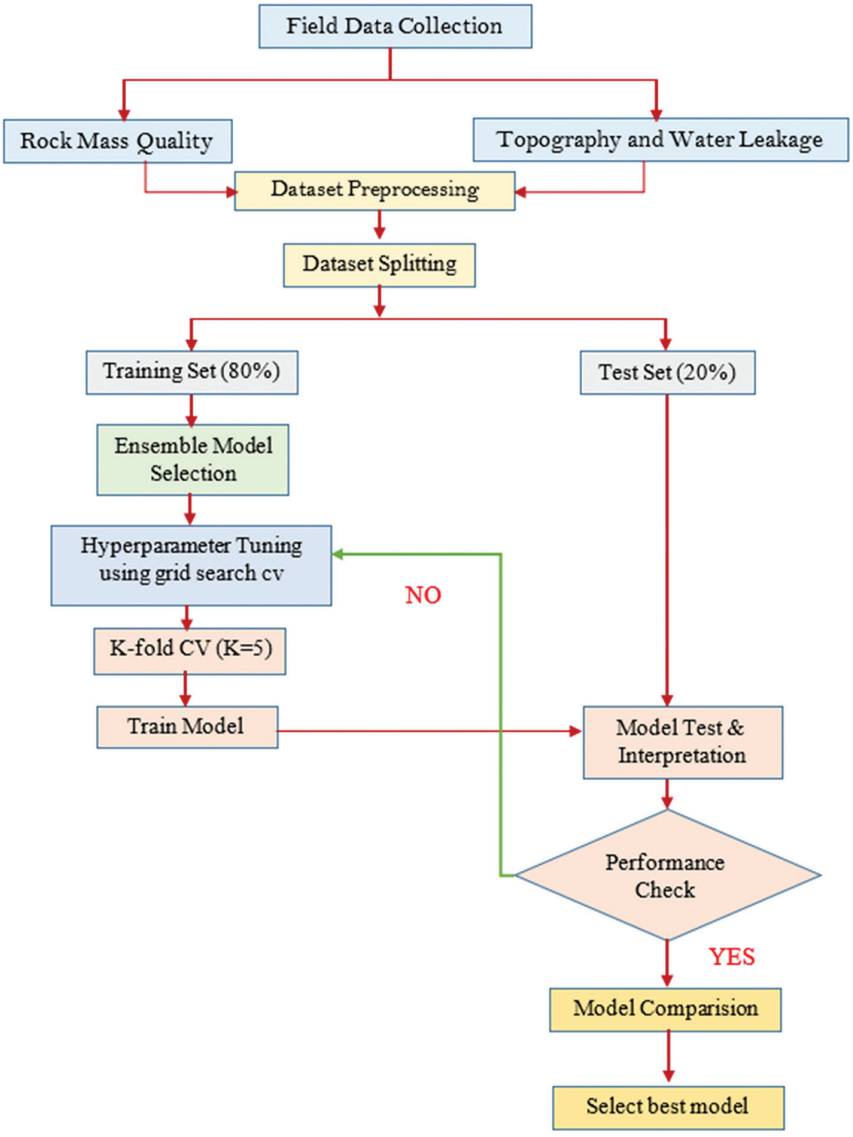

Figure 2: Workflow for water leakage prediction

4. Machine Learning Technique in Tunnel Engineering

The integration of machine learning (ML) into geotechnical and hydrogeological modeling has significantly advanced predictive capabilities in recent years. Traditional empirical and analytical methods often struggle with the high variability and non-linear nature of rock mass behavior and groundwater flow dynamics. In contrast, ML techniques offer robust tools to model complex interactions between geological parameters and engineering outcomes (Zhang et al., 2021).

Several studies have demonstrated the effectiveness of ML in predicting geotechnical parameters such as rock mass classification (Ghaboussi et al., 1998), tunnel convergence (Moayedi & Rezaei, 2019), and water inflow in tunnels (Yilmaz et al., 2020). For instance, Moayedi and Rezaei (2019) applied Artificial Neural

Networks (ANNs) and Support Vector Machines (SVMs) to predict tunnel wall convergence, achieving higher accuracy than conventional regression models.

In the context of water leakage prediction , Yilmaz et al. (2020) used Random Forest (RF) and Gradient Boosting Decision Trees (GBDT) to estimate seepage in underground structures. Their results showed that ensemble models outperformed single learners due to their ability to handle high-dimensional, noisy, and imbalanced datasets typical of geotechnical environments.

More recently, Zhang et al. (2021) applied XGBoost and LightGBM to predict groundwater inflow in deep tunnels under complex hydrogeological conditions. They emphasized the importance of feature selection and hyperparameter tuning in improving model generalization, especially when data are sparse or collected from heterogeneous sources.

5. Partial Dependence Plot (PDP) and Individual Conditional Expectation (ICE)

PDP and ICE plots assist in interpreting complex machine learning models by offering insights into how each feature influences the target variable. PDPs display the average effect of a feature on the predicted outcome by marginalizing over the values of all other features, offering a global interpretation (Friedman, 2001). This helps model users to understand overall trends and offers a relationship between specific features and predicted outcomes. However, because PDPs rely on averaging, they can sometimes obscure heterogeneous behavior across individual data instances, especially when features are correlated (Molnar, 2022).

In contrast, ICE plots reveal the variation in predictions at the individual instance level by showing one line per observation, indicating how predictions change as a single feature varies (Goldstein et al., 2015). This provides a specific view for users to analyze how individual features impact the model output, capturing local pattern and interactions that may not evident in PDP. In general, ICE plot helps in diagnosis of model biases, pattern of individual data points highlighting potential heterogeneity in data. Together, PDP and ICE plots are crucial for understanding the relationship between features and prediction, improving model transpency and accountability and debugging and refining machine learning models.

6. Database Study

According to Panthi (2006), managing water leakage under full hydrostatic pressure is critical in designing unlined or shortcrete-lined tunnels, with an allowable limit of 1.5 L/min/m. Leakage (equation-1) is mainly influenced by joint characteristics , degree of jointing, hydrostatic pressure, and the shortest distance from the tunnel to the surface (Panthi & Basnet, 2021).

$$q = fa * H * \frac{Jn * Jr}{Ja} \quad (1)$$

This study collected rock mass parameters such as RQD, Jn, Jr, Ja, and Q-value, along with topographic features including hydrostatic head (Hst), shortest distance to the valley side (D), and permeability factor (fa). These attributes are summarized in Tables 1 and 2. Panthi's semi-empirical method was used to estimate specific leakage (q), assuming groundwater descends to the tunnel level during the dry season and HRT flow governs leakage along the valley side (Panthi & Basnet, 2021).

Table 1: Field dataset 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 (no.s)	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 dataset for topography and permeability

Description	Count	mean	Std	min	25%	50%	75%	max
Hst (m)	211	23.6	11.9	0.25	13.15	26.31	35.4	36.36
D (m)	211	163.1	72.2	62.1	105.86	140.29	224.525	305.14
fa(l/min/m^2)	211	0	0	0	0.02	0.03	0.05	0.14
q (m^3/s/m)	211	6.5	7.7	0.17	1.375	3.29	8.81	33.68

6.1 Correlation Analysis

Correlation analysis measures the direction and strength of relationships between features and a dependent variable, using a correlation coefficient ranging from -1 to +1. A value of 1 indicates a strong positive correlation, -1 a strong negative correlation, and 0 no correlation. Heat maps are commonly used to visualize these relationships with color-coded cells. As shown in Figure 3, water ingress or leakage shows positive correlations with RQD (0.15), Jn (0.14), Jr (0.49), Jw (0.23), Q (0.22), Jp (0.16), and Js (0.41), and negative correlations with Ja (−0.26), SRF (−0.35), and D (−0.51) (Thapa, Ghimire & Bhusal, 2024). These values indicate a notable relationship between the target and input features.

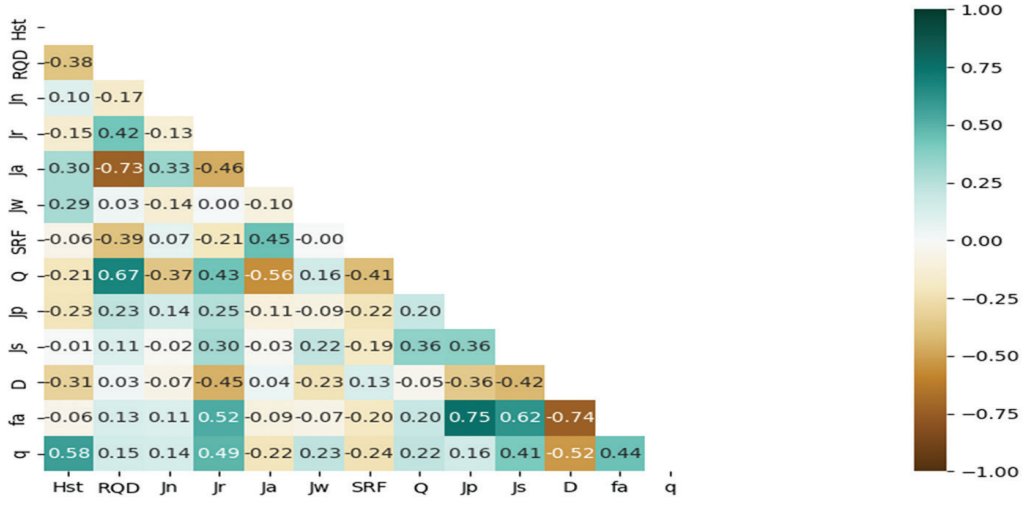

Figure 3: Correlation matrix of datasets

6.2 Data Distribution:

The histogram and KDE plot (Figure 4) present the distribution of specific water leakage values (q) across the tunnel alignment. While this figure initially served to depict basic statistical spread, its shape reveals meaningful implications for leakage behavior. The distribution is notably right-skewed, indicating that most tunnel sections experience low to moderate leakage, whereas a few locations exhibit significantly high leakage rates. This pattern suggests that leakage is not uniformly distributed, but rather influenced by localized geological weaknesses such as intensely fractured zones, thin overburden near valley sides, or highly permeable joint systems. The presence of outliers in the upper range underscores the importance of detecting high-risk sections during design and support planning. Additionally, the multimodal nature of the KDE curve hints at the presence of different leakage regimes, possibly due to transitions in rock mass type or groundwater conditions. This reinforces the necessity of using ensemble machine learning models capable of handling complex, nonlinear patterns in the data to make reliable predictions.

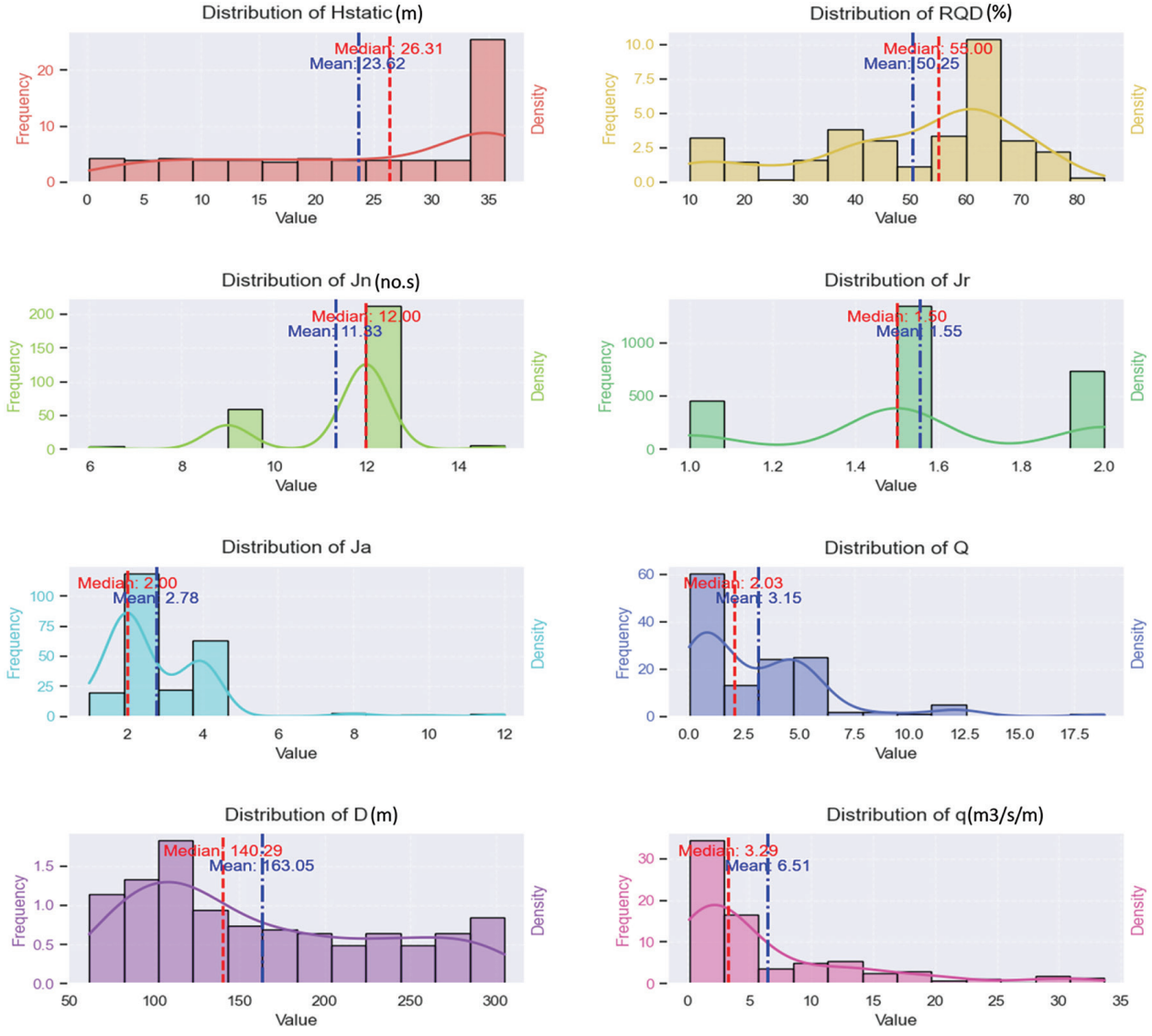

Figure 4: Histogram and Kernel Density Estimation (KDE) plot

7. Hyper-parameter Tuning

Hyper-parameter tuning is the process of selecting the optimal set of model parameters to enhance performance (Bengio, 2012; Courville et al., 2016). These parameters, set prior to training, govern how the model learns (Courville, 2016). Common tuning methods include Grid Search CV and Random Search CV, both effective depending on data characteristics (Bergstra & Bengio, 2012). While Random Search is faster and suited for large datasets, Grid Search often yields more precise results. Despite its longer runtime, Grid Search CV was employed in this study for higher accuracy (Hutter et al., 2019). A 5-fold cross-validation (K=5) was used during tuning (Kohavi, 1995) and the optimal hyperparameters are presented in table 3.

Table 3: Best hyperparameter selection for ML model

Model	Optimal Hyperparameter
Random Forest	'min_samples_leaf': 1, 'min_samples_split': 2
Decision Tree	'max_depth': None, 'max_features': None 'min_samples_leaf': 1, 'min_samples_split': 2
K-Nearest Neighbour	leaf_size': 20, 'n_neighbors': 5, 'p': 1, 'weights': 'distance'
Support Vector Machine	'C': 10, 'degree': 3, 'epsilon': 0.1, 'gamma': 1, 'kernel': 'poly'
Bagging	max_depth': 10, 'bootstrap': False, 'max_features': 1.0, 'max_samples': 0.7, 'n_estimators': 100
Boosting	'colsample_bytree': 1.0, 'learning_rate': 0.1, 'max_depth': 5, 'min_child_weight': 3, 'n_estimators': 150, 'subsample': 0.6

8. Base and Meta-model Selection

Base models perform initial predictions. These base models or learners are used in the bagging, voting, stacking, and blending ensemble approaches. In this study, the base model can be a Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree. Their individual performances, measured in terms of R^2 score and Mean Squared Error (MSE), are presented in Figure 5. In bagging, the same model is used, whereas in voting, we have to use a different algorithm. Similarly, in stacking and blending, we have to combine different base learners before the final prediction is carried out by meta-model.

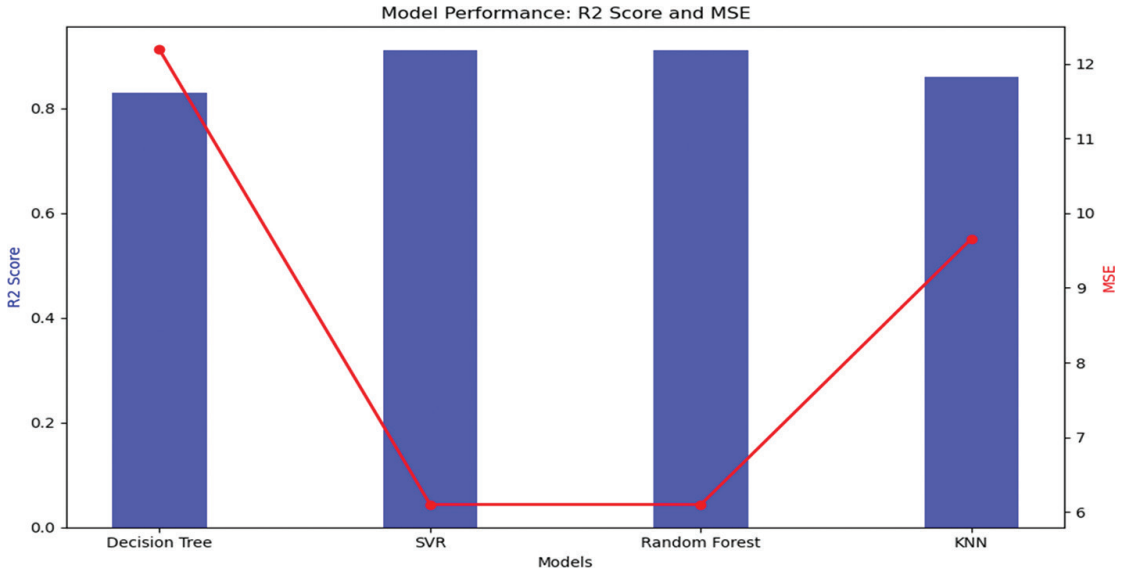

Figure 5: A comparision of base model and meta-model result

9. Statistical Analysis of Selected Model

Developing accurate leakage models for headrace tunnels in hydropower projects requires a deep understanding and application of statistical analysis. In this study, Equations 2 to 9 are used to compute various statistical metrics, including R-squared (R^2), 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) (Thapa, Ghimire&Bhusal,2024).

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}[{}_i^aY - {}_i^pY] \quad (3)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[{}_i^aY - {}_i^pY]^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}({}_i^aY - {}_i^pY)2} \quad (5)$$

$$RRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{{}_i^aY - {}_i^pY}{{}_i^aY})^2} \quad (6)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{{}_i^aY - {}_i^pY}{{}_i^aY}\right| * 100\% \quad (7)$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{{}_i^aY - {}_i^pY}{{}_i^aY}\right| \quad (8)$$

$$VAF = 1 - \left|\frac{\text{var}({}_i^aY - {}_i^pY)}{\text{var}({}_i^aY)}\right| * 100\% \quad (9)$$

10. Results and Discussion

In this study, ensemble techniques like Bagging, xgBoosting, Voting, and Stacking analyzed the given datasets with the optimal hyperparameters from the table 3. As discussed earlier this hyperparameter for improved model performance was evaluated using grid search cv with k-fold cross-validation. For ensemble techniques like voting and stacking, SVM is taken as the meta-model. The prediction and outcomes of different approaches have been discussed below.

10.1 Bagging

Bagging uses SVM as base learners with hyperparameter tuning, and employs bootstrap sampling without replacement, distinguishing it from other ensembles. The model fits tunnel leakage data well, with predicted values (red line) closely matching actual values (blue line). An R^2 of 0.92 confirms strong predictive power, explaining 92% of variance and demonstrating robustness. Overall, bagging with SVM improves accuracy, reduces overfitting, and is effective for complex regression tasks like leakage prediction in tunnels, highlighting the value of ensemble methods for superior performance.

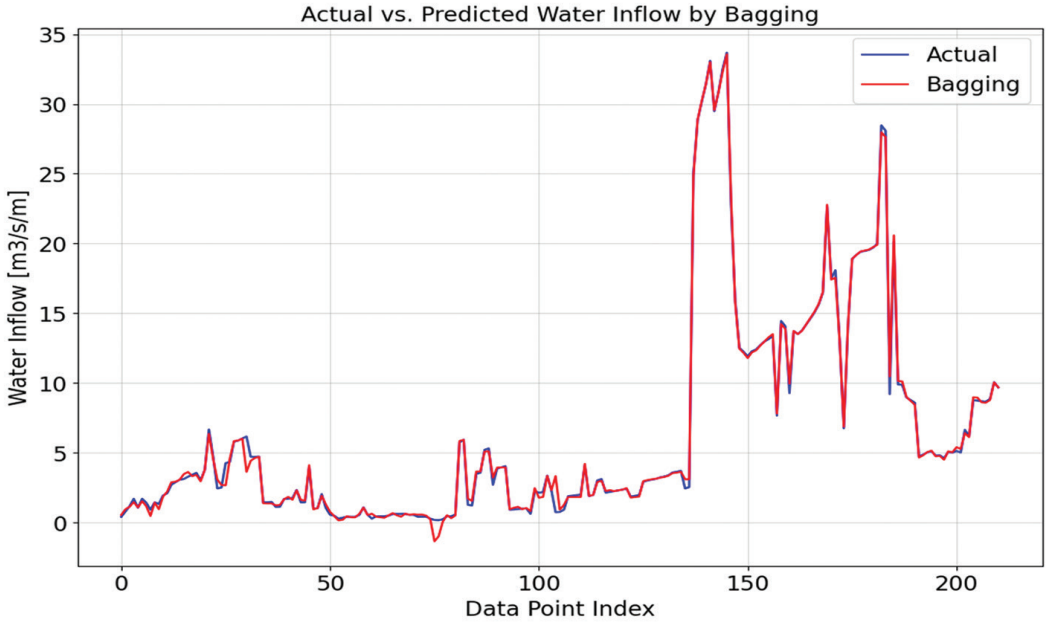

Figure 6: Comparison of actual and predicted water leakage values using bagging model

Table 4: A summary of statistical indices of bagging model

R^2	MSE	MAE	RMSE	RRMSE	MAPE	MRE	VAF
0.9936	0.453	0.3214	0.6769	8.26%	36.83%	10.51%	99.37%

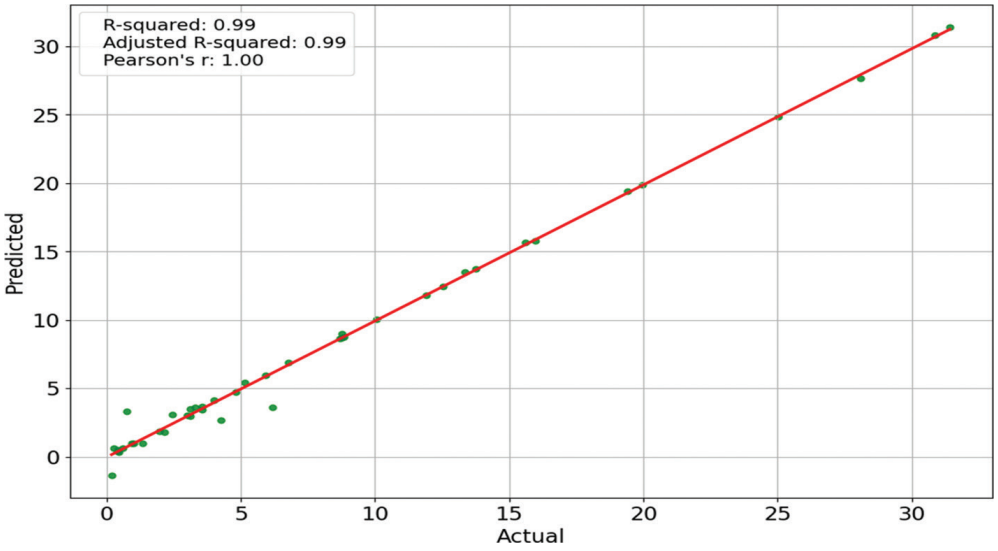

Figure 7: Relationship between actual and predicted datasets for bagging model

10.1.1 Model Interpretation for Bagging

Figure-8 presents PDP and ICE plots for a bagging regressor, showing how features impact predictions of the target variable q. Features like 'Hstatic' and 'Jp' have a strong positive effect, with increasing PDPs and consistent ICE curves, while 'Jr' shows a positive trend with more individual variability. Conversely, 'Js', 'D', and 'fa' display flat PDPs and clustered ICE lines, indicating minimal influence. These plots reveal feature importance and the nature of their relationships with q, aiding model interpretation.

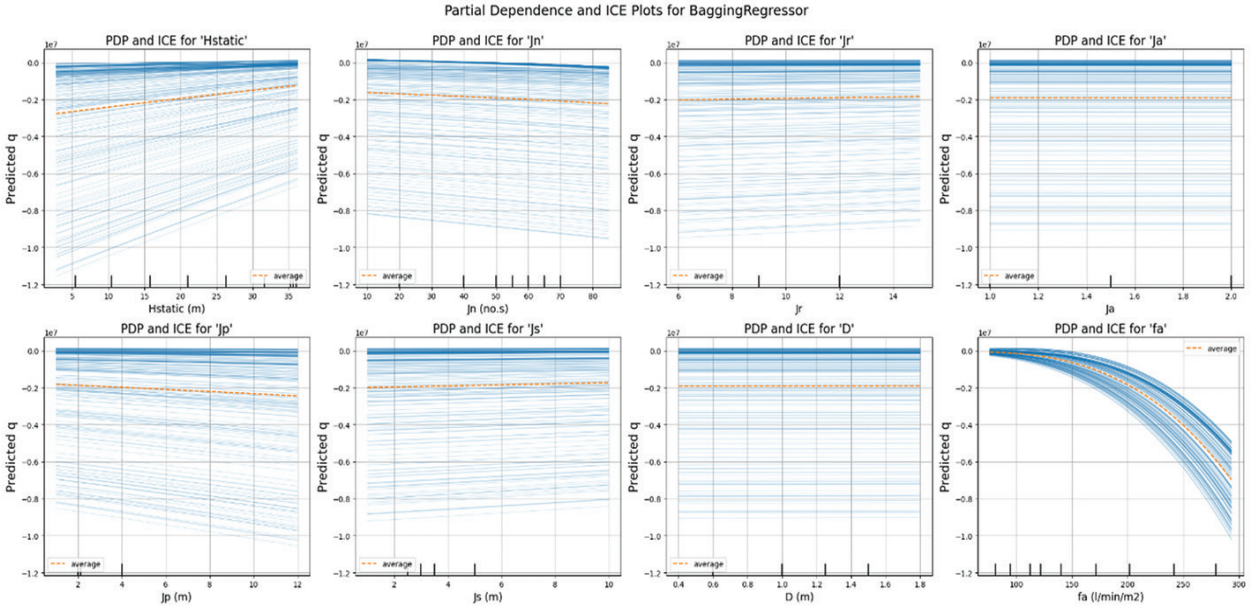

Figure 8: PDP and ICE plot for bagging model

10.2 Boosting

Boosting is an ensemble method that sequentially adds weak learners, typically shallow decision trees, to improve performance by reducing variance and refining predictions. It assigns higher weights to models or samples that minimize errors. In predicting leakage, the Xgboosting model closely matches actual leakage trends, with an R^2 of 0.88, explaining 88% of the variance. This demonstrates strong predictive accuracy, making Xgboosting a reliable and effective approach for modeling leakage patterns.

Table 5: A summary of statistical indices for the boosting model

R^2	MSE	MAE	RMSE	RRMSE	MAPE	MRE	VAF
0.8813	8.55	1.04	2.92	35.66%	13.65%	4.83%	88.58%

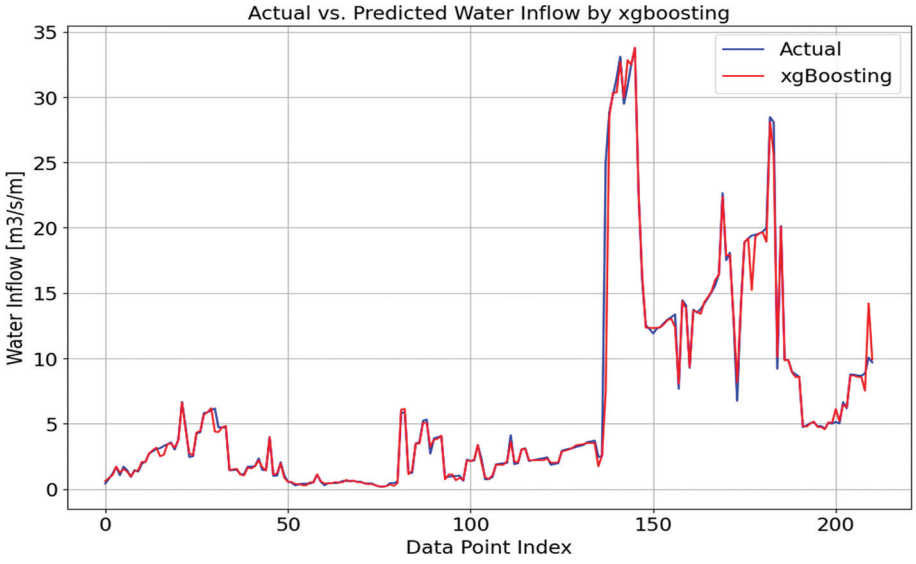

Figure 9: Comparison of actual and predicted water leakage values using xgBoost model

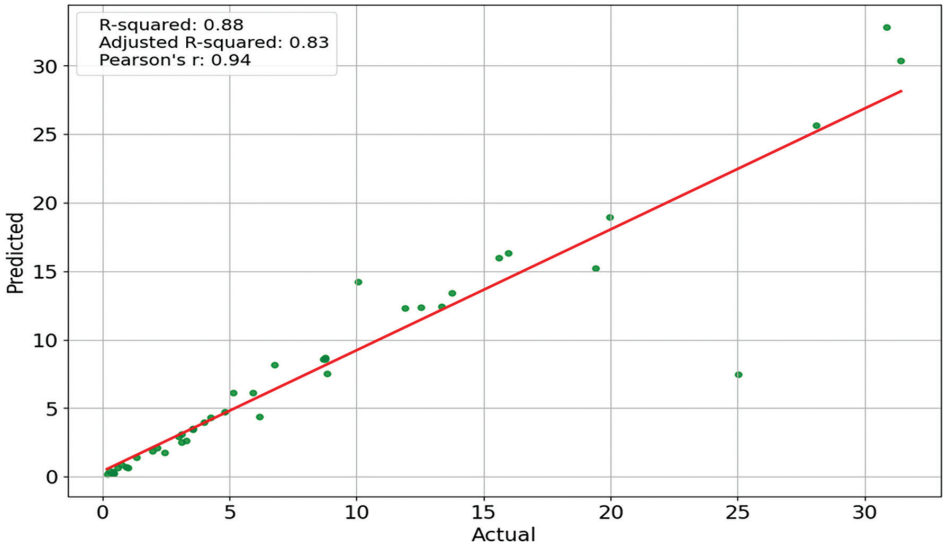

Figure 10: Relationship between actual and predicted datasets for xgboosting

10.2.1 Model Interpretation for Xgboosting

Figure 10's Partial Dependence Plots (PDP) and Individual Conditional Expectation (ICE) plots reveal how features affect boost's predictions for q. Hstatic shows a strong negative effect above 15, while D has a consistent positive correlation. Similarly, fa exhibits a threshold effect with a sharp drop at very low values before stabilizing. ICE plots highlight individual variation: Jr's lines cluster tightly, indicating consistent effects, whereas Jn and Ja show scattered lines, reflecting diverse impacts across samples. These insights emphasize the need to consider both average trends and sample-specific responses in model interpretation.

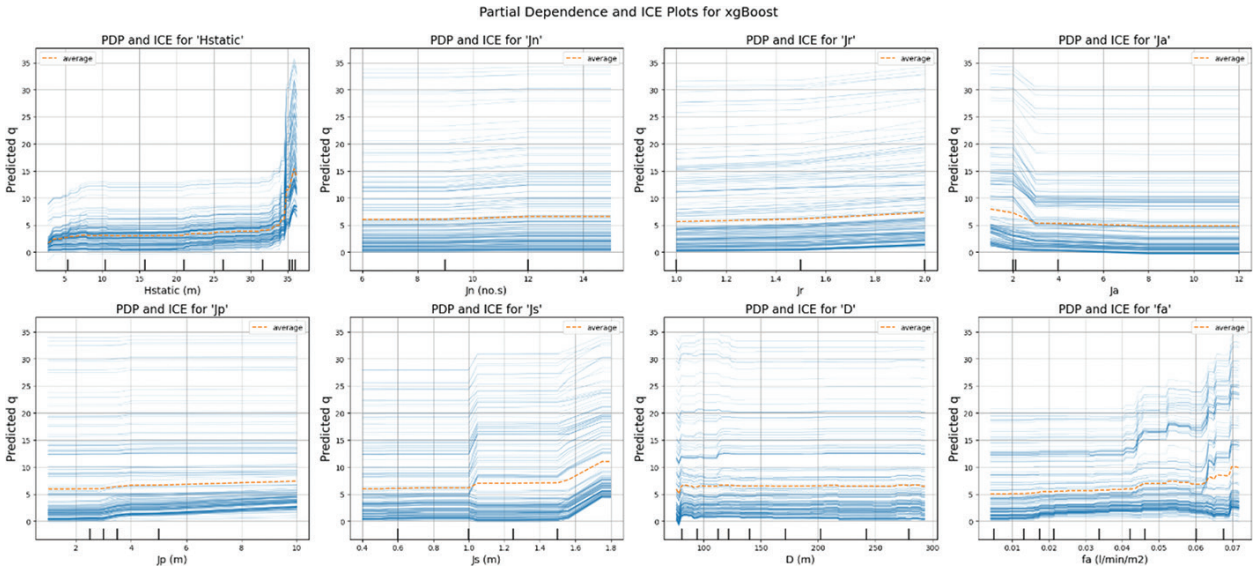

Figure 11: PDP and ICE plot for xgboosting

10.3 Voting

This approach combines Random Forest, K-Nearest Neighbor, Decision Tree, and Support Vector Machine, leveraging diverse models to reduce individual errors. The voting ensemble shows strong performance, with an R^2 of 0.97 and VAF of 97.45%, explaining most variance in the target. Low error metrics MSE of 1.88, RMSE of 1.37, and MAE of 0.57 indicate accurate predictions. Relative errors (RRMSE 16.72%, MAPE 14.52%) are reasonable, and a slight underestimation bias (MRE -3.25%) is observed. Overall, the voting ensemble delivers reliable, precise predictions by effectively minimizing errors and capturing data variability.

Table 6: A summary of statistical indices by voting

R^2	MSE	MAE	RMSE	RRMSE	MAPE	MRE	VAF
0.97	1.88	0.57	1.37	16.72%	14.52%	-3.25%	97.45%

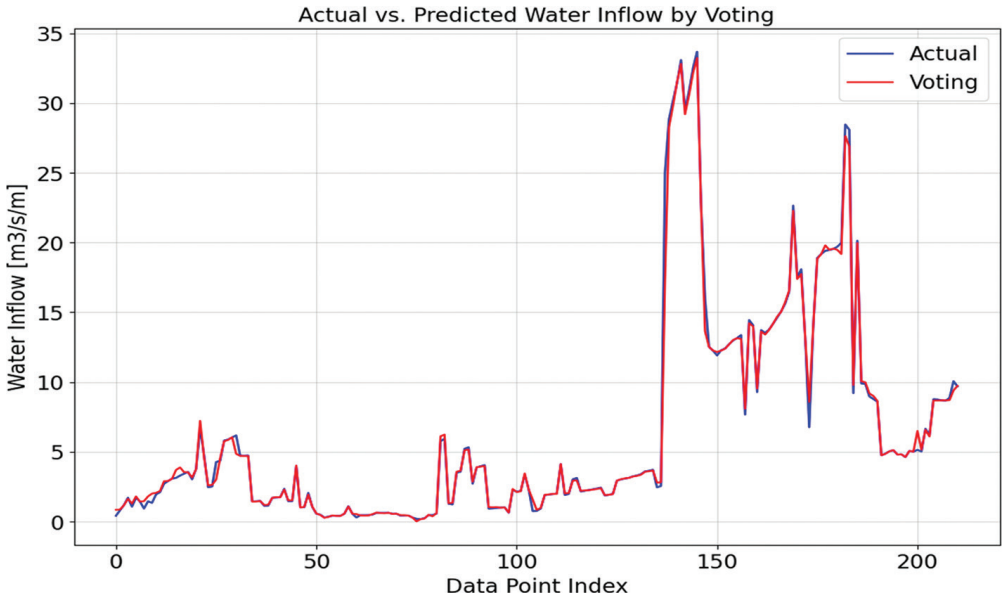

Figure 12: Comparison of actual and predicted water leakage values using voting model

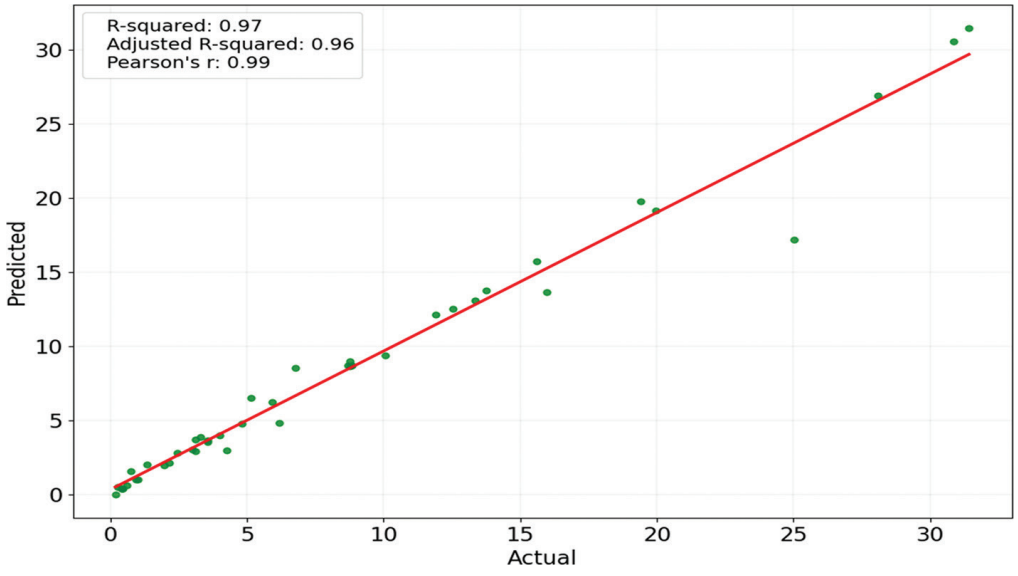

Figure 13: Relationships between actual vs. predicted datasets for voting

10.3.1 Model Interpretation for Voting

The PDP and ICE plots for the voting model show that feature D has a strong negative impact on the predicted target q, with higher D values leading to lower q. In contrast, Jn, Jr , Jp, and fa have nearly flat plots, indicating minimal influence. Features Hstatic and Ja exhibit weak negative trends, while Js shows negligible effect. Overall, D is the most influential feature, with others having little to weak impact on predictions.

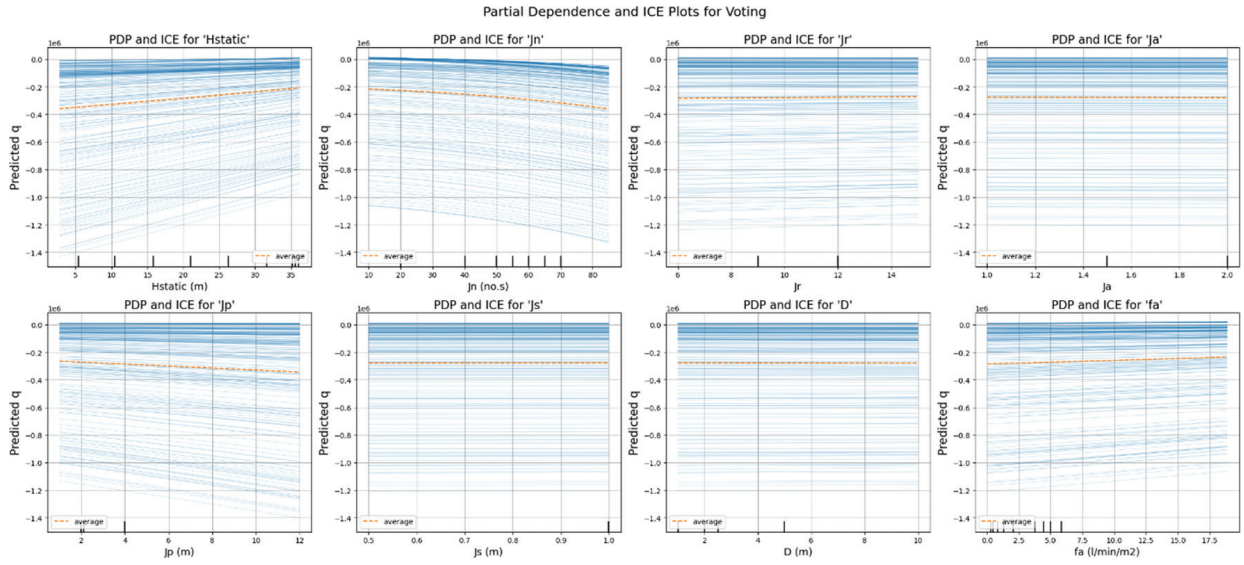

Figure 14: PDP and ICE plot for voting

10.4 Stacking

This stacking method uses Random Forest as a meta-model and combines K-Nearest Neighbors, Decision Tree, and Support Vector Machine as base learners. K-fold cross-validation (K=10) was applied to maximize accuracy, distinguishing stacking from blending and enhancing predictive performance. With an R^2 of 0.970, the model explains 97% of the variance, indicating excellent fit. Low error values—MSE of 2.159, MAE of 0.576, and RMSE of 1.469—show predictions closely match actual values. The RRMSE of 17.92% and MAPE of 14.53% confirm strong accuracy, while a slight underestimation is noted with an MRE of -5.06%. The VAF of 97.07% further supports the model's effectiveness in capturing data variance, demonstrating high precision and reliability.

Table 7: A summary of statistical indices of stacking model

R^2	MSE	MAE	RMSE	RRMSE	MAPE	MRE	VAF
0.97	2.15	0.57	1.46	17.92%	14.53%	-5.06%	97.07%

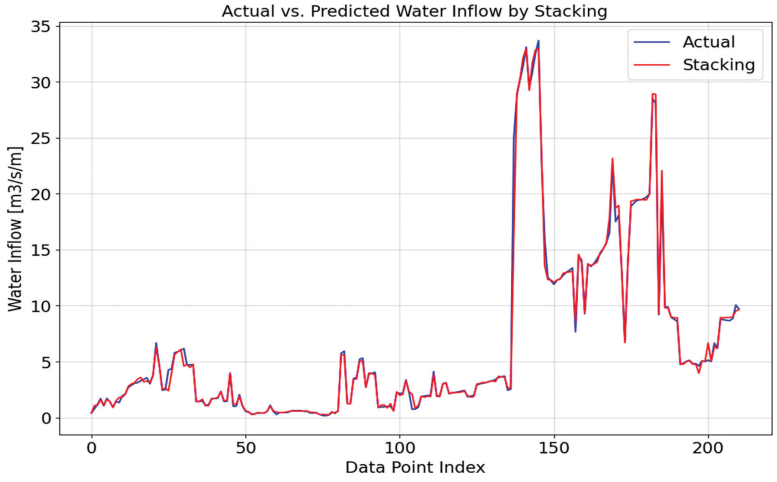

Figure 15: Comparison of actual and predicted water leakage values using stacking model

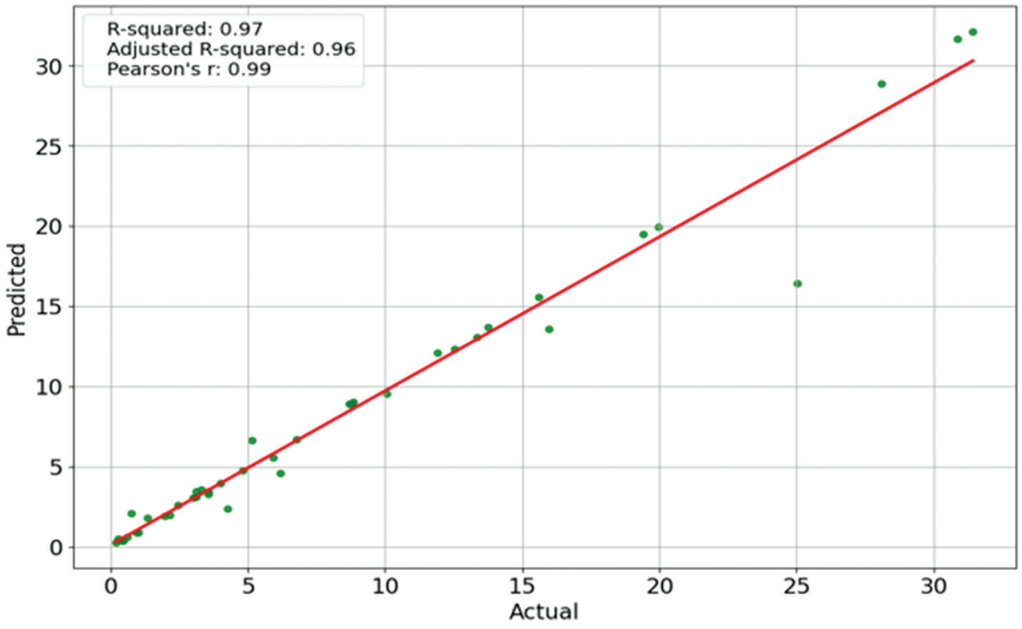

Figure 16: Relationships between actual vs. predicted datasets for stacking

10.4.1 Model Interpretation for Stacking

In the stacking model, Hstatic (figure-17) shows a strong positive influence on predicted rock mass quality (q), especially beyond its midpoint, highlighting its importance. Jr and fa have moderate positive effects, supporting better classification. Conversely, Ja and D exhibit clear negative impacts, indicating quality degradation. Features like Jn, Jp, and Js show minimal influence. Overall, the plots provide valuable global and local insights into how geotechnical factors drive model predictions.

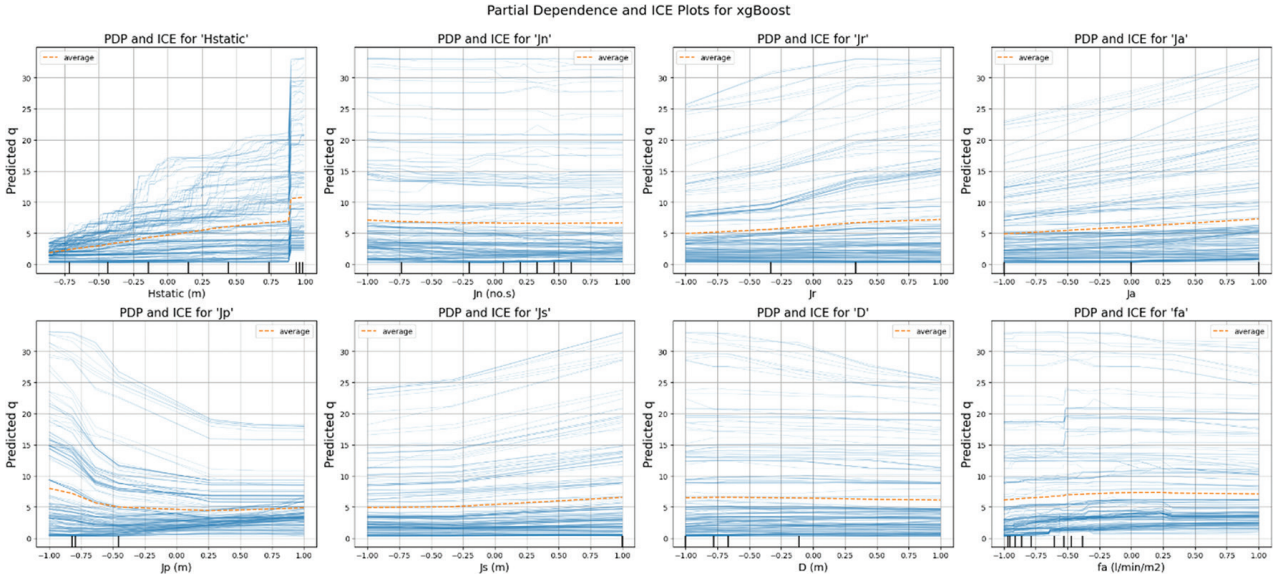

Figure 17: PDP and ICE plot for the stacking model

11. Result Comparison

To determine the best model among Bagging, Boosting, Voting, and Stacking, we analyzed various statistical indices that measure different aspects of model performance. Bagging consistently outperforms the other models in most metrics, including R^2 (0.99), MSE (0.45), MAE (0.32), RMSE (0.67), RRMSE (8.26%), and VAF (99.73%), all of which indicate high accuracy and minimal error. These metrics are critical for assessing how well a model explains variance and minimizes prediction errors, making Bagging the top choice when absolute accuracy is the priority. However, Boosting performs exceptionally well in MAPE (13.65%), which measures percentage-based errors, making it suitable for scenarios where relative accuracy is more important than absolute error. On the other hand, Voting and Stacking show balanced performance, particularly in MRE, where they exhibit less underestimation compared to Bagging and Boosting. Despite their balanced nature, these models lag behind Bagging in terms of overall accuracy metrics like R^2 and MSE. Therefore, while Boosting, Voting, and Stacking have niche strengths, Bagging emerges as the best overall model due to its superior performance across the most robust evaluation metrics, making it the most reliable choice for leakage modeling tasks.

Table 8: Comparison of statistical indices by different ensemble model

S.N	Statistical Indices	Bagging	Boosting	Voting	Stacking
1.	R^2	0.99	0.88	0.97	0.97
2.	MSE	0.45	8.55	1.88	2.15
3.	MAE	0.32	1.04	0.57	0.57
4.	RMSE	0.67	2.92	1.37	1.46
5.	RRMSE	8.26%	35.66%	16.72%	17.92%
6.	MAPE	36.83%	13.65%	14.52%	14.53%
7.	MRE	10.51%	4.83%	-3.25%	-5.06%
8.	VAF	99.73%	88.58%	97.45%	97.07%

12. Conclusion

This study employed four ensemble machine learning techniques—Bagging, Boosting (XGBoost), Voting, and Stacking—to predict water leakage in the headrace tunnel of the Nilgiri-II Hydropower Project. Input features included rock mass characteristics, hydrostatic head, permeability, and topographic variables. Among the models, Bagging outperformed all others, achieving the highest accuracy (R^2 = 0.99, VAF = 99.73%) and lowest prediction error, as confirmed by multiple statistical indices.

The study demonstrates that ensemble models offer significant advantages over individual learners by effectively capturing nonlinear patterns and handling geological variability. Partial Dependence Plots (PDPs) and ICE plots revealed that features like hydrostatic head (Hstatic) and joint properties (Jp, Jr, D) play a critical role in influencing leakage predictions.

From a practical standpoint, the results offer valuable insights for tunnel design, particularly in the Himalayan region, where unpredictable water ingress can compromise structural integrity. Accurate predictions can inform support design, grouting strategies, and even early warning systems during excavation.

However, this study has limitations. The model is based on a single project dataset, which may limit generalizability. Additionally, deep learning approaches, though more complex, were not explored and could

offer further improvement in predictive accuracy. Future research should focus on:

- Validating the model across multiple tunnel projects
- Incorporating real-time data for adaptive modeling
- Exploring deep learning frameworks for higher-dimensional feature spaces

Overall, this study highlights the effectiveness and interpretability of ensemble learning for geotechnical applications and sets a foundation for more advanced and scalable tunnel leakage prediction systems.

Acknowledgements

The author sincerely thanks the Nilgiri-II Hydropower Project authorities for providing access to field data used in this study. The dataset is proprietary and subject to confidentiality; therefore, it is not publicly available. However, it may be shared upon reasonable request and with prior approval from the project stakeholders.

REFERENCES

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. Neural Networks, 26, 9–16.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281–305.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. Journal of Computational and Graphical Statistics, 24(1), 44–65.

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automated machine learning: Methods, systems, challenges.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI) (Vol. 2, pp. 1137–1143).

Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.).

Moayedi, H., & Rezaei, S. (2019). Prediction of tunnel wall convergence using artificial intelligence techniques. Geomechanics and Geoengineering, 14(3), 174–183.

Panthi, K. K. (2006). Analysis of hydrogeological and rock engineering parameters in tunnel design: A case study from the Himalaya [Doctoral dissertation, Norwegian University of Science and Technology (NTNU)].

Panthi, K. K., & Basnet, K. (2021). Water leakage estimation in unlined headrace tunnels using semi-empirical methods. Journal of Hydropower Engineering, 38(2), 25–34.

Satici, O. (2006). Advances in drill-and-blast tunnel construction technology. Tunnelling and Underground Space Technology, 21(3–4), 259–267.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman and Hall.

Singh, B., & Goel, R. K. (2011). Tunnelling in weak rocks. Elsevier.

Thapa, R., Ghimire, B., & Bhusal, A. (2024). Geological and hydrogeological characterization of Nilgiri-II tunnel. Nepal Engineers Association Technical Journal. (

Węglarczyk, S. (2018). Kernel density estimation and its application. ITM Web of Conferences, 23, 00037. https://doi.org/10.1051/itmconf/20182300037

Yilmaz, I., Basha, N. H., & Tatar, I. (2020). Estimation of seepage in underground structures using ensemble learning methods. Bulletin of Engineering Geology and the Environment, 79, 1499–1515.

Zhang, Y., Zhu, H., Liu, J., & Chen, Y. (2021). Prediction of groundwater inflow in tunnels using XGBoost and LightGBM. Engineering Geology, 289, 106194.

Zhou, Z.-H. (2012). Ensemble methods: Foundations and algorithms. Chapman and Hall/CRC.