Geomechanics and Engineering, *Vol. 43, No. 3 (2025) 155-166*
https://doi.org/10.12989/gae.2025.43.3.155

Application of stacking regressor to predict rock fracture toughness mode-I

Ibrahim Albaijan*[1], Arsalan Mahmoodzadeh[2], Mokhtar Mohammadi[3] and Hussein Alrobei[1]

[1]*Mechanical Engineering Department, College of Engineering at Al-Kharj, Prince Sattam Bin Abdulaziz University, Al Kharj 16273, Saudi Arabia*
[2]*Center of Research and Strategic Studies, Lebanese French University, Erbil, Iraq*
[3]*Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Region, Iraq*

(*Received July 4, 2024, Revised October 15, 2025, Accepted October 17, 2025*)

Abstract. Predicting the fracture toughness of rocks, particularly under Mode-I loading conditions, is essential for various geotechnical and civil engineering applications. Traditional methods for determining rock fracture toughness (RFT) are often labor-intensive, time-consuming, and prone to inaccuracies due to the inherent variability in rock properties. This study investigates the efficacy of using a stacking regressor, an advanced ensemble learning technique, to predict the Mode-I RFT. In the proposed model, the strengths of multiple base regressors were combined. 400 experimental data points were utilized, obtained using the cracked Chevron notched Brazilian disc (CCNBD) test and comprising six input parameters affecting the Mode-I RFT. The dataset was partitioned into training and validation sets, ensuring rigorous model evaluation. The stacking regressor's meta-model was trained on the outputs of the base models, effectively learning to integrate their predictions to yield a more accurate final prediction. The performance of the stacking regressor was assessed through several statistical metrics. The results demonstrated that the stacking regressor significantly outperforms individual base models, achieving higher predictive accuracy and reliability. A sensitivity analysis using the mutual information test (MIT) method revealed that the uniaxial tensile strength (UCS) exerts the most significant influence on the Mode-I RFT, underscoring its importance in predictive modeling. Furthermore, developing a machine learning-based graphical user interface (GUI) enhanced the practical applicability of the proposed model, making it accessible to engineers and researchers without extensive expertise in machine learning.

Keywords: cracked Chevron notched Brazilian disc test; fracture toughness Mode-I; machine learning; sensitivity analysis

1. Introduction

Rock fracture toughness (RFT) is a critical parameter influencing failure propagation during hydraulic fracturing, rock blasting, caving, tunneling, and related activities (Guha Roy *et al.* 2018, Wang *et al.* 2021). For instance, RFT is pivotal in forming fracture networks through hydraulic fracturing technology in geothermal and petroleum engineering. Moreover, the anisotropic nature of RFT within shale layers significantly affects the control of the artificial fracture propagation field (Liu and Dai 2021), thereby enhancing shale gas recovery (Chandra and Vishal 2021). Consequently, a comprehensive understanding of RFT is indispensable for successfully designing and implementing projects in mining engineering, geomechanics, and civil engineering.

As illustrated in Fig. 1, rock fracture mechanics delineates that rocks fracture under various loading conditions in three distinct modes: tensile opening mode (Mode-I), in-plane shear sliding mode (Mode-II), and out-of-plane shear tearing mode (Mode-III) (Chang *et al.* 2002, Wang *et al.* 2021). Since rocks possess a lower tensile strength than their shear strength, Mode-I fractures occur more frequently. Consequently, Mode-I rock fracture toughness (KIC) is generally considered a fundamental mechanical parameter in rock mechanics.

Due to its critical importance, numerous efforts have been undertaken to measure the KIC value precisely in laboratory settings and the field. The International Society for Rock Mechanics (ISRM) recommends various geometrically diverse rock specimens to assess Mode-I RFT (K_{IC}), including Chevron bending (CB), Chevron-notched short rod (SR), straight-crack semi-circular bending (SCB), cracked Chevron notched Brazilian disc (CCNBD), cracked Chevron notched semi-circular bending (CCNSCB) and cracked straight-through Brazilian disc (CSTBD). Among these, pre-existing artificial fractures are typically categorized into straight-through and Chevron cracks, with the latter being more prevalently utilized in ISRM-endorsed test specimens for measuring K_{IC} (Wang *et al.* 2021). However, accurately measuring RFT in the laboratory presents significant challenges due to the intricate sample preparation process, the requirement for exact and sensitive instrumentation, the brittleness of samples, premature sample failure, and occasionally the scarcity of sufficient core samples. To address these issues, various empirical correlations have been developed in recent years to facilitate the prediction of K_{IC}, offering more straightforward calculation methods. Despite these advancements, although individual geomechanical properties can be correlated with RFT, the data often exhibit

*Corresponding author, Assistant Professor
E-mail: i.albaijan@psau.edu.sa

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

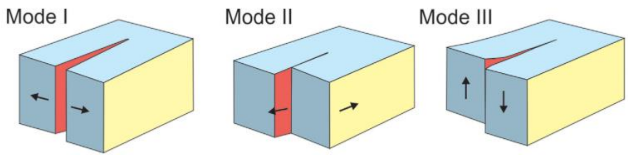

Fig. 1 The basic modes for rock fracturing (Gudmundsson 2014)

random distribution, resulting in weak correlations (Guha Roy *et al.* 2017). Additionally, the type of rock significantly influences these correlations, a factor that empirical relationships have yet to incorporate effectively. These limitations underscore the necessity for developing alternative, more robust methodologies to predict KIC accurately.

Traditional methods for determining RFT are often labor-intensive, time-consuming, and prone to inaccuracies due to the inherent variability in rock properties. Machine learning (ML) techniques have recently emerged as a transformative paradigm in engineering applications and scientific research (Abdelmawla *et al.* 2023, Kamran *et al.* 2022, Lawal *et al.* 2023). ML and data science advancements herald a renaissance in complex data analysis and interpretation, enabling the extraction of patterns from vast, multidimensional datasets beyond human comprehension. Laboratory tests and computational simulations can directly inform ML algorithms, offering ML-based approximate solutions that enhance understanding and prediction in various domains.

Although machine learning (ML) methods have been extensively developed and applied across various engineering disciplines (Hashempour *et al.* 2022, Khan *et al.* 2019, Mahmoodzadeh *et al.* 2020, 2021), their application to problems related to rock fracture toughness (RFT) remains limited (Wang *et al.* 2021). Wang et al. (Wang *et al.* 2021) employed numerical software to simulate different cracked Chevron notched Brazilian disc (CCNBD) rock specimens, generating substantial datasets to train ML models. Mahmoodzadeh *et al.* (2022) generated 250 laboratory data using the CCNBD test on granite rock samples to train and evaluate support vector regression (SVR)-based hybrid models. These research findings indicate that ML-based approaches can yield acceptable and accurate predictions for K_{IC}, demonstrating the potential of these advanced techniques in this specialized area.

This study investigates the efficacy of using a stacking regressor, an advanced ensemble learning technique, to predict the K_{IC}. The proposed model combines the strengths of multiple base regressors to enhance predictive accuracy and robustness. A comprehensive dataset (400 data points) comprising several features compelling on the RFT is used to train and test the model. The stacking regressor's meta-model is trained on the outputs of the base models, effectively learning to integrate their predictions to yield a more accurate final prediction. Finally, the performance of the stacking regressor to predict the K_{IC} is assessed through several standard metrics. Among the most important innovations of this work, the following can be mentioned:

- Innovative use of stacking regressor: The study explores using stacking regressor, an advanced ensemble learning technique, to predict K_{IC}. This method leverages the strengths of multiple base regressors, a novel approach in rock mechanics and geotechnical engineering.
- Integration of multiple base regressors: By combining the outputs of various base models, the research demonstrates how a stacking regressor can integrate these predictions to yield a more accurate and reliable final prediction. This ensemble approach is innovative as it enhances predictive accuracy compared to using individual models.
- Comprehensive dataset utilization: The study utilizes a robust dataset comprising 400 experimental data points obtained using the CCNBD test.
- Practical implications for geotechnical engineering: The study presents a valuable and efficient tool for geotechnical engineers, facilitating more accurate assessments in rock mechanics. Applying stacking regressors can lead to improved decision-making and planning in various geotechnical and civil engineering projects.
- Addressing limitations of traditional methods: The research addresses the limitations of traditional methods for determining K_{IC}, which are often labor-intensive, time-consuming, and prone to inaccuracies. The proposed ML-based approach offers a more efficient and reliable alternative.

These novel contributions highlight the study's significance and potential impact on advancing predictive modeling techniques in rock mechanics, geotechnics, and civil engineering.

2. Methodology

This study uses the stacking regressor (stacked regression or ensemble learning) to predict the K_{IC}. A stacking regressor is an advanced ML technique combining multiple regression models to improve predictive performance. The idea is to leverage the strengths of different models by training a meta-model to make the final prediction based on the outputs of the base models. The procedural steps for the stacking regressor modeling are delineated as follows:

1. Base Models (Level-0 Models):
   - These are the initial regression models that are trained on the original dataset.
   - Different regression algorithms (e.g., linear regression, decision trees, support vector regression) can be base models.
   - Each base model captures different patterns and relationships in the data.
2. Meta-Model (Level-1 Model):
   - The meta-model, or stacking model, is trained on the outputs (predictions) of the base models.
   - It learns to combine these predictions to make a more accurate final prediction.

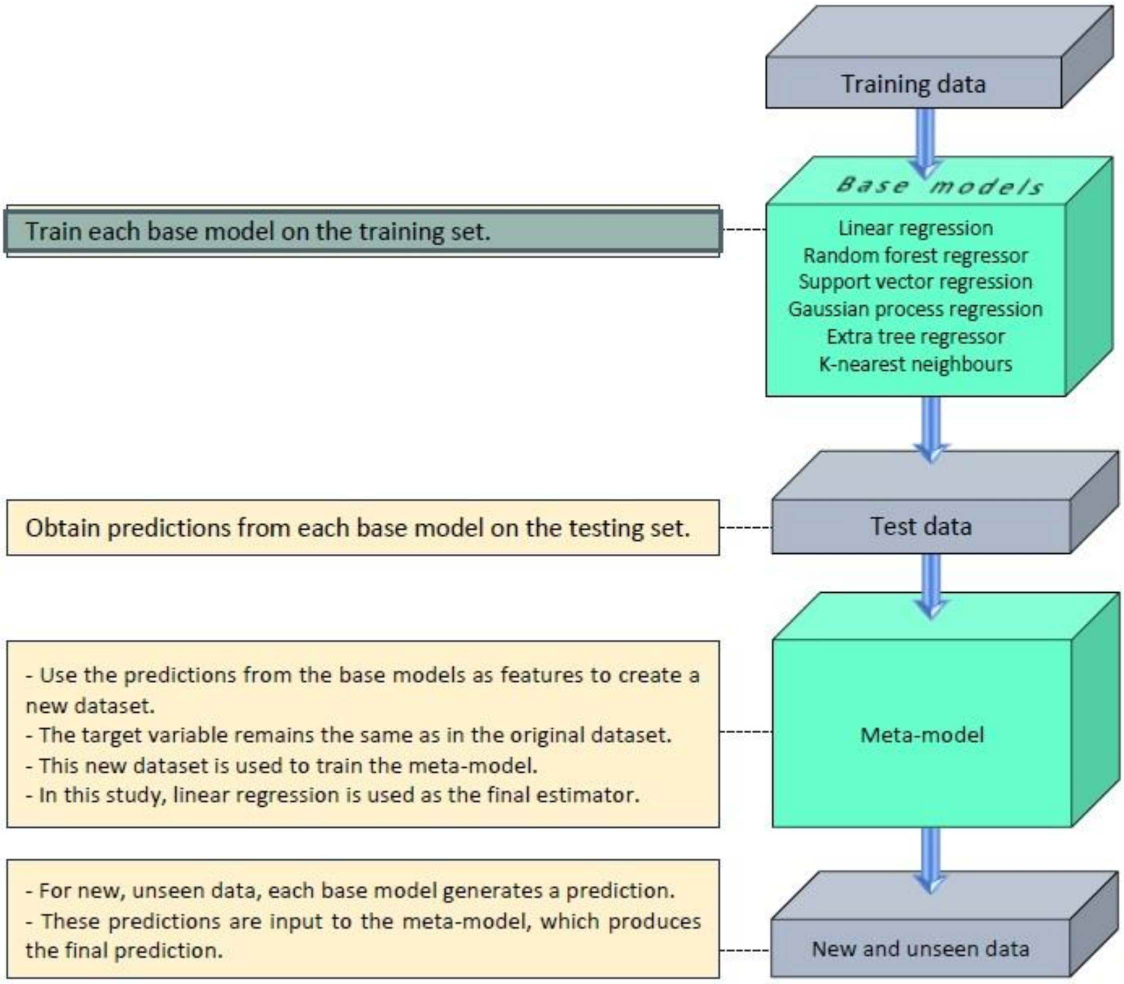

Fig. 2 The whole procedure of the stacking regressor utilized in this study

- Typically, a simple model like linear regression or a more complex model like a neural network can be used as the meta-model.

Some advantages of the stacking regressor are:

- Stacking can often produce more accurate and robust predictions than any single model by combining multiple models.
- It leverages the strengths and compensates for the weaknesses of different models.
- Stacking allows the use of various algorithms, making it highly adaptable to different types of data and problems.
- Using multiple models helps to reduce the risk of overfitting, as the meta-model can smooth out the individual biases of base models.

The Python code employed in this study is delineated below. Six ML algorithms, including linear regression (LR), random forest regressor (RFR), support vector regression (SVR), Gaussian process regression (GPR), extra trees regressor (ETR), and K-nearest neighbors (KNN), are utilized as the base models. Each model is trained on the training dataset and evaluated on the test dataset. The predicted outputs (RFTs) generated by the ML models on the test data are recorded, after which the meta-model is trained on these predictions. In this study, the GPR model served as the meta-model. The final outputs predicted by the meta-model (GPR) are then compared with the original test outputs to assess the prediction performance of the stacking regressor model. The entire stacking regressor modeling process utilized in this study is illustrated in Fig. 2.

```
#Python Code (Jupyter Notebook environment provided in the Anaconda Navigator version 3.7 is used)
#Stacking Regressor
from sklearn.ensemble import StackingRegressor
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import ExtraTreeRegressor
from sklearn.svm import SVR
from sklearn.model_selection import train_test_split
# Sample dataset
```

Table 1 A comprehensive overview of the database

		UTS	R	B	α_B	α_0	α_1	K_{IC}
Training set	count	320	320	320	320	320	320	320
	mean	15.72	40.52	34.57	0.71	0.26	0.69	1.72
	std	8.00	9.34	8.94	0.08	0.04	0.04	0.13
	min	5.20	21.34	13.55	0.48	0.17	0.57	1.36
	25%	12.30	34.00	26.72	0.63	0.22	0.66	1.65
	50%	15.30	38.00	33.00	0.75	0.25	0.70	1.72
	75%	17.20	48.00	41.00	0.77	0.27	0.73	1.81
	max	43.68	78.30	65.00	0.92	0.44	0.70	2.02
Testing set	count	80	80	80	80	80	80	80
	mean	15.84	40.65	34.65	0.70	0.26	0.69	1.72
	std	8.33	10.02	9.04	0.09	0.04	0.04	0.14
	min	4.60	24.90	17.95	0.55	0.22	0.62	1.36
	25%	12.10	34.00	28.28	0.63	0.22	0.66	1.64
	50%	15.30	37.00	34.70	0.74	0.25	0.70	1.73
	75%	17.20	48.53	41.00	0.77	0.27	0.73	1.82
	max	37.83	75.25	60.00	0.83	0.39	0.76	2.02

```
X, y = ... # Your features and target variable
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split (X, y, test_size =
0.2, random_state = 42)
# Define base models
base_models = [
    ('gpr', GaussianProcessRegressor(normalize_y = True)),
    ('svr', SVR()),
    ('lr', LinearRegression()),
    ('rf', RandomForestRegressor()),
    ('knn', KNeighborsRegressor()),
    ('ETR', ExtraTreeRegressor()),
    # Add more regressors as needed]
# Define meta-model
meta_model = GaussianProcessRegressor ()
# Create stacking regressor
stacking_regressor = StackingRegressor(estimators=base_models,
final_estimator=meta_model)
# Train stacking regressor
stacking_regressor.fit(X_train, y_train)
# Make predictions
    predictions = stacking_regressor.predict(X_test)
```

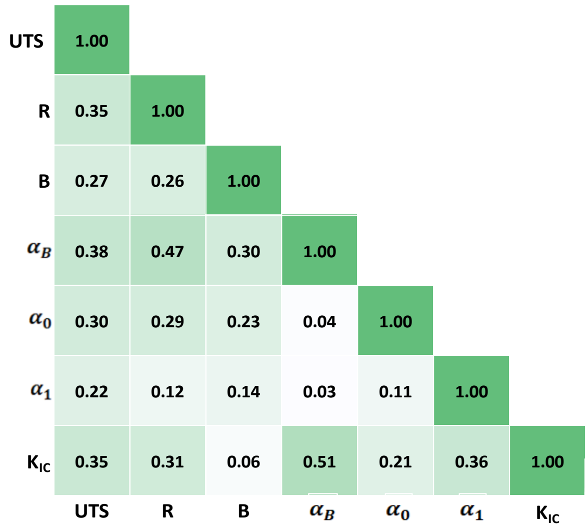

Fig. 3 Correlation matrix between features

3. Dataset preparation

ISRM recommends various test specimens for laboratory determination of K_{IC}. These specimens typically feature artificially created pre-existing fractures, categorized into Chevron cracks and straight-through cracks. Chevron cracks are more commonly employed in ISRM-recommended test specimens for measuring K_{IC} due to several advantages: (I) they mitigate issues associated with sharp crack points, (II) they reduce the complexity of laboratory equipment, (III) they offer extensive geometric flexibility, (IV) they simplify loading constants, and (V) they are easier to fabricate. Therefore, CCNBD specimens are favored for quantifying K_{IC}.

The primary notations for the CCNBD testing specimens with pre-existing Chevron cracks include: thickness of disc specimen (B), radius of disc specimen (R), length of initial crack (a_0), length of crack (a), length of final crack (a_1), and the radius of the Chevron crack incision made using a diamond saw blade (R_S). Compressive loads (P) must also be applied to measure the CCNBD speci. Other pertinent notations include: $\alpha_0 = a_0/R$, $\alpha_1 = a_1/R$, $\alpha_B = B/R$, and $\alpha_s = R_s/R$. The CCNBD specimens' dimensions must adhere to the ranges specified by the ISRM: $\alpha_1 \leq 0.8$, $\alpha_1 \geq 0.4$, $\alpha_1 \geq \alpha_B/2$, $\alpha_B \geq 1.1729\,.(\alpha_1)^{5/3}$, $\alpha_B \leq 1.04$, and $\alpha_B \geq 0.44$.

In this study, to develop the database for ML methods, 400 Granite rock specimens were prepared, all adhering to the dimensional criteria of the CCNBD testing specimens. Ultimately, 400 datasets were obtained, encompassing six input parameters: uniaxial tensile strength (UTS), B, R, α_0, α_1, α_B, and the output parameter K_{IC}. The dataset was partitioned, with 80% designated for training and 20%

reserved for testing. A comprehensive overview of the database is provided in Table 1.

To elucidate further on the database, the correlation matrix between the input variables, as well as between the input variables and the output (K_{IC}), is depicted in Fig. 3. This matrix provides a comprehensive overview of the relationships between the various parameters involved in the study. Notably, the highest correlation, with a 0.47, is observed between the parameters R and α_B . This significant correlation suggests a moderate relationship between the R and the α_B. The correlations among the other input parameters are lower than this value, indicating weaker relationships. These lower correlation values imply that the input parameters exhibit a degree of independence. Consequently, to accurately estimate the K_{IC} parameter, it is imperative to consider all the input parameters.

4. Data normalization

Data normalization is a crucial preprocessing step in ML and statistical modeling that aims to standardize the range of independent variables or features of data. The primary objective of normalization is to transform the data into a standard scale without distorting differences in the ranges of values. The key objectives of data normalization are:

1. Standardization of data range: Normalization scales data to a specific range, typically [0, 1] or [-1, 1], ensuring that no feature dominates due to its scale.
2. Enhancement of model performance: Normalization speeds up the convergence of gradient descent and improves the accuracy of ML models by standardizing the data.
3. Reduction of redundancy: It mitigates the effect of outliers and reduces redundancy, facilitating more meaningful comparisons across different features.

Different data normalization methods exist, each with unique advantages and applications. Among these, the Min-Max scaling method is utilized in this study data due to its effectiveness in handling the range and scale of different features. By applying Min-Max scaling, all input variables are transformed to a standard scale, facilitating the integration and comparison of diverse data points. This normalization process is crucial for enhancing the performance of the ML models, ensuring that each feature is equally represented and contributed appropriately to the predictions.

The Min-Max scaling method transforms the data into a specific range, typically between 0 and 1. The transformation formula is given by Eq. (1).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x' is the normalized value, x is the original value, $x_{\min}$ is the minimum value of the feature, and $x_{\max}$ is the maximum value of the feature.

The advantages of the Min-Max scaling method are:

- Simplicity and interpretability: The Min-Max scaling method is straightforward to interpret. It linearly transforms the data without altering the relationships between the values.
- Preservation of data distribution: This method preserves the relationships between the data points, maintaining the values' original distribution and relative spacing.
- Improved convergence in algorithms: Algorithms that are sensitive to the scale of data benefit significantly from Min-Max scaling, leading to faster convergence and more stable solutions.

5. Statistical evaluation indices

Several statistical evaluation indices were employed to evaluate the accuracy of the forecasting models, including the coefficient of determination (R^2), root mean square error (RMSE), mean absolute percentage error (MAPE), and variance accounted for (VAF).

R^2 metric quantifies the proportion of the variance in the dependent variable that can be predicted from the independent variables, reflecting the model's goodness of fit. It provides a clear measure of how well observed outcomes are replicated by the model.

RMSE is the square root of the mean of the squared deviations between the predicted and observed values, offering a robust measure of the model's prediction error. RMSE provides a direct interpretation of prediction errors in the same units as the response variable, making it intuitive to understand.

MAPE expresses forecasting accuracy as a percentage calculated by averaging the absolute errors divided by the actual values. Thus, it provides an intuitive measure of prediction accuracy. This metric facilitates comparison between models with different scales, as it provides a percentage error.

VAF indicates the proportion of the total variance explained by the model, thereby offering insight into the model's explanatory power and effectiveness. VAF gives insight into the effectiveness of the model in capturing the variability of the data.

Using these indices together offers a comprehensive evaluation, balancing different aspects of model performance and robustness. They provide a multifaceted view that can lead to more informed decisions regarding model selection and improvement. However, reliance on these metrics alone without considering the specific context and data characteristics could lead to biased or incomplete conclusions.

6. Results and discussion

The initial phase in implementing the stacking regressor method involves meticulously training the base models. This foundational step is crucial as it sets the stage for the subsequent integration of predictions from multiple models to enhance overall predictive accuracy.

This study selected six sophisticated algorithms as base models: GPR, SVR, RF, LR, KNN, and ETR. These algorithms were chosen for their diverse methodological approaches and proven efficacy in various regression tasks.

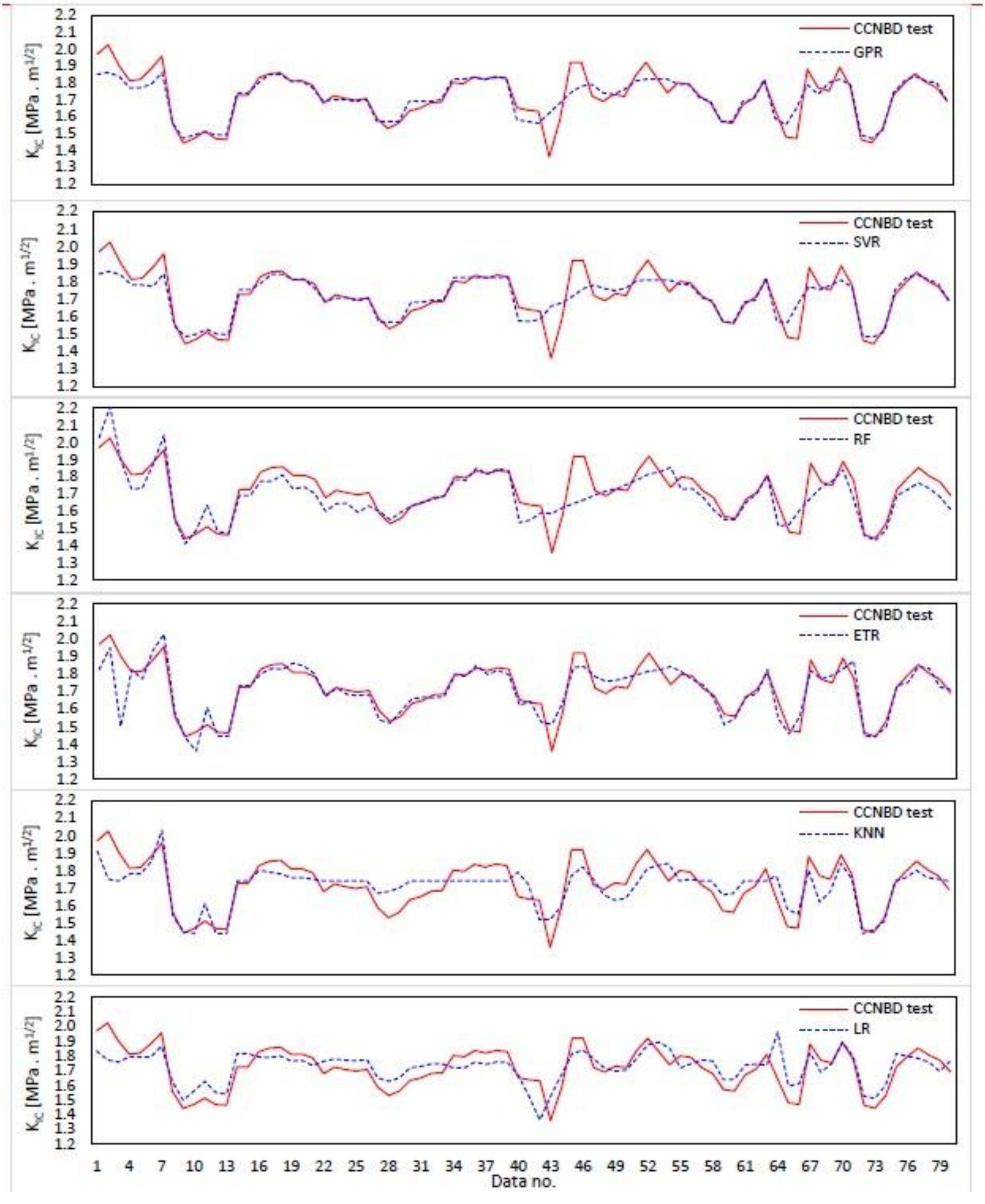

Fig. 4 Comparison of the base models' predictions with the CCNBD tests

Each model was rigorously trained on 80% of the dataset, ensuring that the models had ample data to learn the underlying patterns and relationships. Following the training phase, the models were evaluated on the remaining 20% of the data, which served as the test set to assess the models' generalization capabilities.

The predicted output parameter values for the test data generated by each base model were then meticulously compared with the actual measured values. This comparison is essential for understanding how well each model can predict unseen data and identifying potential overfitting or underfitting issues.

In Fig. 4, the predicted K_{IC} values by each base model are juxtaposed with the measured values, providing a visual representation of the model's performance. This figure reveals that the GPR and SVR models yielded predictions

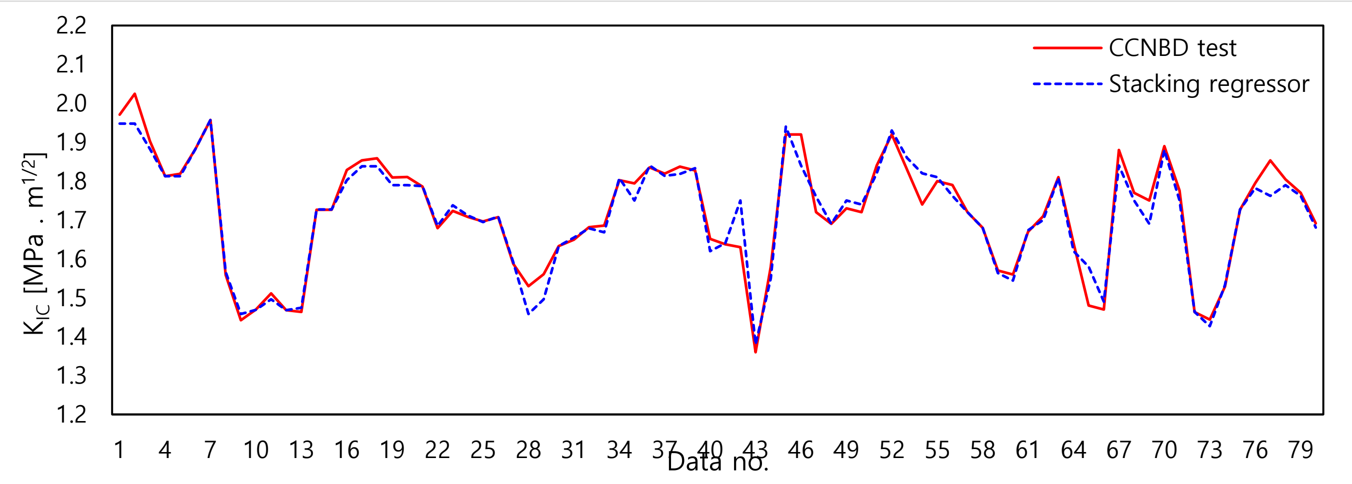

Fig. 5 Comparison of the stacking regressor's predictions and the CCNBD tests

Table 2 Comparative analysis of the base models' performance via statistical indices

	GPR	SVR	RFR	ETR	KNN	LR
R^2	0.81	0.76	0.67	0.68	0.67	0.58
Score	5	4	2	3	2	1
MAPE	0.02	0.03	0.04	0.05	0.04	0.05
Score	4	3	2	1	2	1
RMSE	0.06	0.07	0.08	0.08	0.08	0.09
Score	4	3	2	2	2	1
VAF	0.90	0.87	0.84	0.83	0.82	0.76
Score	6	5	4	3	2	1
Ranking score	19	15	10	9	8	4

that were remarkably closer to the laboratory results than the other models, highlighting their superior predictive capabilities.

The performance of each base model was further scrutinized using various statistical indices, as summarized in Table 2. Based on the comprehensive analysis of these statistical indices, the GPR model delivered the most accurate and reliable performance among all the base models evaluated. This conclusion underscores the GPR model's efficacy in capturing the complex relationships inherent in the data and its robustness in making precise predictions.

Upon completing the training phase of the base models, the next critical step is the formulation of the meta-model. The meta-model is an integrative framework that consolidates the predictions from the individual base models into a unified predictive output. While any ML algorithm can theoretically function as a meta-model, the choice in this study gravitates towards the GPR algorithm due to its demonstrated capability in handling complex relationships and uncertainties inherent in predictive tasks.

In this context, the inputs fed into the meta-model consist exclusively of the K_{IC} values forecasted by each base model. This configuration results in the meta-model having six distinct input parameters, each corresponding to the outputs generated by the respective base model. Notably, each base model contributes 80 data points to this meta-modeling phase, reflecting the robustness of the ensemble approach in leveraging diverse predictive capabilities.

The primary objective of the meta-model is to produce a refined prediction of the original K_{IC} value derived from the test data. This final output parameter serves as a culmination of the integrated insights gleaned from the ensemble of base models, thereby enhancing the overall accuracy and reliability of the predictive framework.

The results produced by the stacking regressor model are juxtaposed with the results obtained from the CCNBD tests in Fig. 5. It is evident from the comparison that the differences between them are minimal, underscoring the remarkable accuracy of predictions achieved through the stacking regression method. The outcomes projected by the stacking regressor model are compared with the predictions generated by each base model across a range of statistical indices meticulously outlined in Table 3. Through this detailed analysis, the superior predictive performance of the stacking regressor becomes apparent, showcasing its ability to consistently outperform each of the base models in terms of accuracy and reliability in predicting the K_{IC}. This comparative evaluation underscores the efficacy of integrating multiple predictive models within the stacking regressor framework, enhancing the overall robustness and precision of the predictive capabilities employed in rock mechanics, geotechnics, and civil engineering applications.

In Fig. 6, the predictive performance of both the base models and the stacking regressor model for forecasting the K_{IC} parameter is meticulously evaluated using a Taylor diagram. This diagram provides a comprehensive analysis by considering key statistical measures such as the correlation coefficient and standard deviation. This sophisticated visual representation vividly illustrates the remarkable accuracy and reliability of the stacking regressor method. The diagram demonstrates that the stacking regressor method significantly outperforms the base models in predicting the K_{IC} parameter.

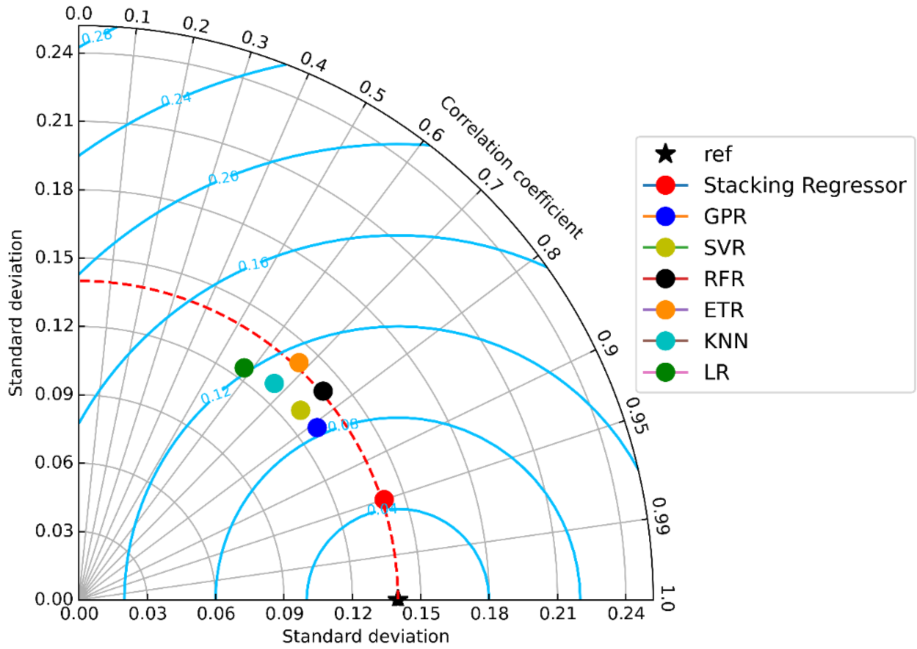

Fig. 6 Performance comparison of the predictive models via Taylor diagram

Table 3 Comparative analysis of the stacking regressor model and base models via statistical indices

	Meta-model	GPR	SVR	RFR	ETR	KNN	LR
R^2	0.95	0.81	0.76	0.67	0.68	0.67	0.58
Score	6	5	4	2	3	2	1
MAPE	0.01	0.02	0.03	0.04	0.05	0.04	0.05
Score	5	4	3	2	1	2	1
RMSE	0.03	0.06	0.07	0.08	0.08	0.08	0.09
Score	5	4	3	2	2	2	1
VAF	0.97	0.90	0.87	0.84	0.83	0.82	0.76
Score	7	6	5	4	3	2	1
Ranking score	23	19	15	10	9	8	4

The cross-validation technique offers crucial insights into the model's generalizability by mitigating any training bias that could arise from a single data split. Various cross-validation techniques are employed for model selection, including the hold-out method, K-fold cross-validation, leave-one-out cross-validation, and the bootstrap method. In this analysis, K-fold cross-validation has been chosen to assess the predictive performance. In this phase, the K-fold cross-validation technique is employed to more rigorously assess the predictive efficacy of the models. This method involves partitioning the dataset into K equal-sized subsets or folds. The model is then trained K times, using K-1 folds as training data and the remaining fold as validation data. This iterative process ensures that each fold is used exactly once as validation data.

K-fold cross-validation enhances the reliability of the model's performance assessment by reducing the variance that may result from a single train-test split. Utilizing different subsets of the data for training and validation across multiple iterations also provides a more accurate estimate of the model's ability to generalize to unseen data.

In this study, a K-value of 5 is employed, and the results of the statistical evaluation criteria are presented in Table 4. The cumulative scores from all folds for each prediction model are presented in Fig. 7. The results indicate that all algorithms exhibit remarkable accuracy in forecasting KIC. However, the stacking regressor model delivers the most accurate predictions. These findings substantiate the robustness and generalizability of the stacking regressor model, reinforcing confidence in its predictive capabilities

7. Sensitivity analysis

The mutual information test (MIT) is employed to conduct a sensitivity analysis of the input parameters on the K_{IC}, explicitly focusing on rock fracture toughness (KIC). The MIT measures the dependency between input and

Table 4 Comparison among the 5-fold cross-validation results produced by the predictive models

Method		R^2	Score	MAPE	Score	RMSE	Score	VAF	Score
Stacking regressor	Fold 1	0.96	7	0.01	5	0.02	5	0.97	6
	Fold 2	0.94	5	0.02	5	0.03	5	0.96	7
	Fold 3	0.95	6	0.01	5	0.02	5	0.97	7
	Fold 4	0.95	7	0.01	6	0.02	6	0.96	7
	Fold 5	0.94	6	0.02	6	0.02	5	0.96	7
GPR	Fold 1	0.82	6	0.02	4	0.04	4	0.92	5
	Fold 2	0.81	4	0.03	4	0.03	5	0.90	6
	Fold 3	0.82	5	0.02	4	0.03	4	0.91	6
	Fold 4	0.81	6	0.02	5	0.03	5	0.91	6
	Fold 5	0.81	5	0.03	5	0.04	4	0.90	6
SVR	Fold 1	0.78	5	0.03	3	0.06	3	0.89	4
	Fold 2	0.76	3	0.04	3	0.06	4	0.87	5
	Fold 3	0.76	4	0.04	3	0.07	3	0.87	5
	Fold 4	0.77	5	0.03	4	0.06	4	0.86	5
	Fold 5	0.74	4	0.05	4	0.08	3	0.85	5
RFR	Fold 1	0.70	4	0.04	2	0.06	3	0.86	3
	Fold 2	0.67	2	0.04	3	0.08	3	0.84	4
	Fold 3	0.68	3	0.05	2	0.07	3	0.83	3
	Fold 4	0.67	3	0.04	3	0.07	3	0.84	4
	Fold 5	0.65	3	0.06	3	0.08	3	0.82	4
ETR	Fold 1	0.69	3	0.05	1	0.08	2	0.84	2
	Fold 2	0.67	2	0.05	2	0.09	2	0.83	3
	Fold 3	0.67	2	0.04	3	0.08	2	0.84	4
	Fold 4	0.68	4	0.05	2	0.07	3	0.83	3
	Fold 5	0.64	2	0.07	2	0.09	2	0.81	3
KNN	Fold 1	0.68	2	0.05	1	0.08	2	0.84	2
	Fold 2	0.67	2	0.06	1	0.09	2	0.82	2
	Fold 3	0.67	2	0.05	2	0.08	2	0.80	2
	Fold 4	0.66	2	0.06	1	0.08	2	0.81	2
	Fold 5	0.64	2	0.07	2	0.09	2	0.79	2
LR	Fold 1	0.60	1	0.05	1	0.09	1	0.78	1
	Fold 2	0.58	1	0.05	2	0.10	1	0.76	1
	Fold 3	0.57	1	0.06	1	0.11	1	0.75	1
	Fold 4	0.58	1	0.06	1	0.10	1	0.75	1
	Fold 5	0.55	1	0.08	1	0.13	1	0.72	1

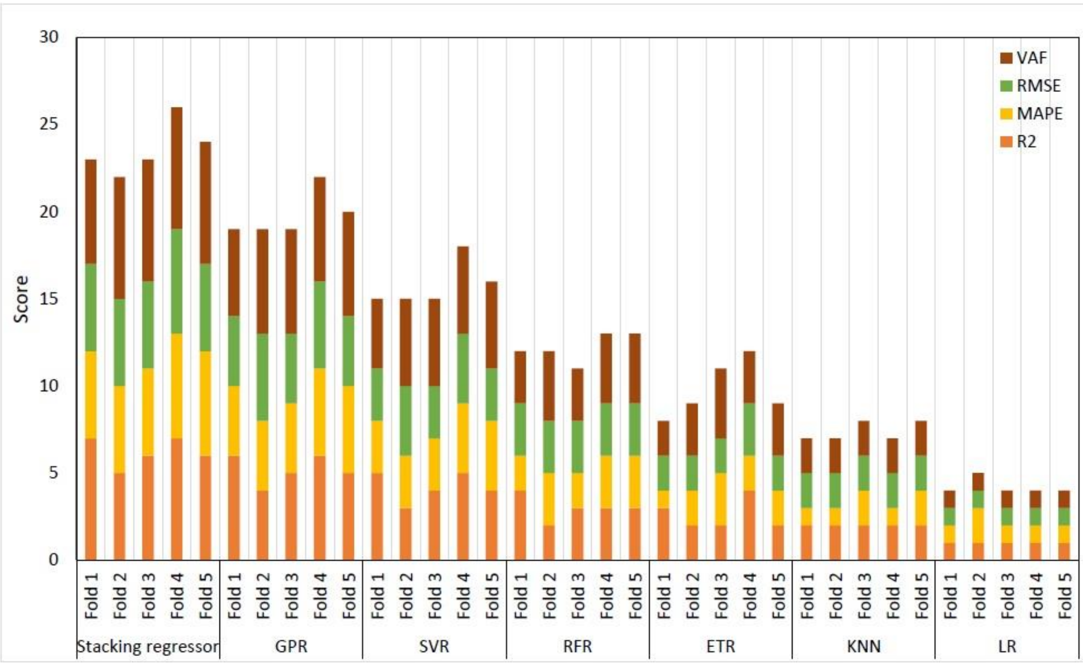

Fig. 7 5-fold cross-validation results

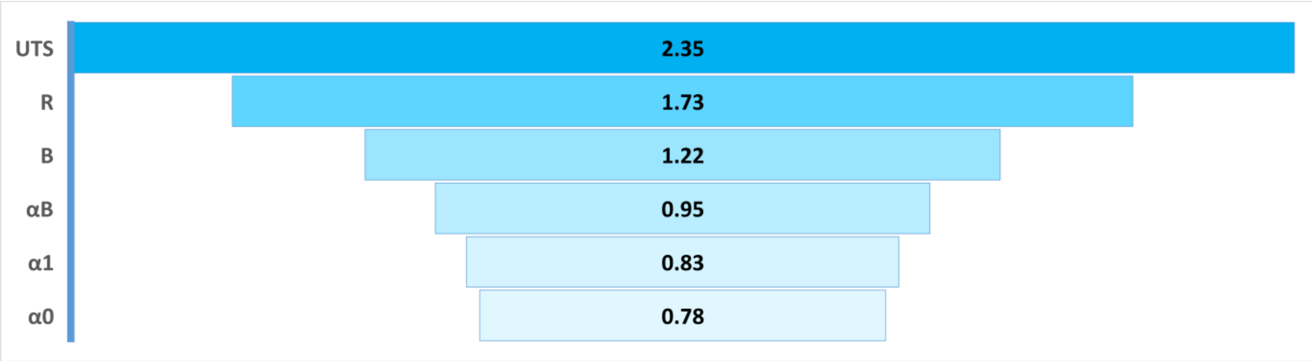

Fig. 8 MIT score between the input parameters and K_{IC}

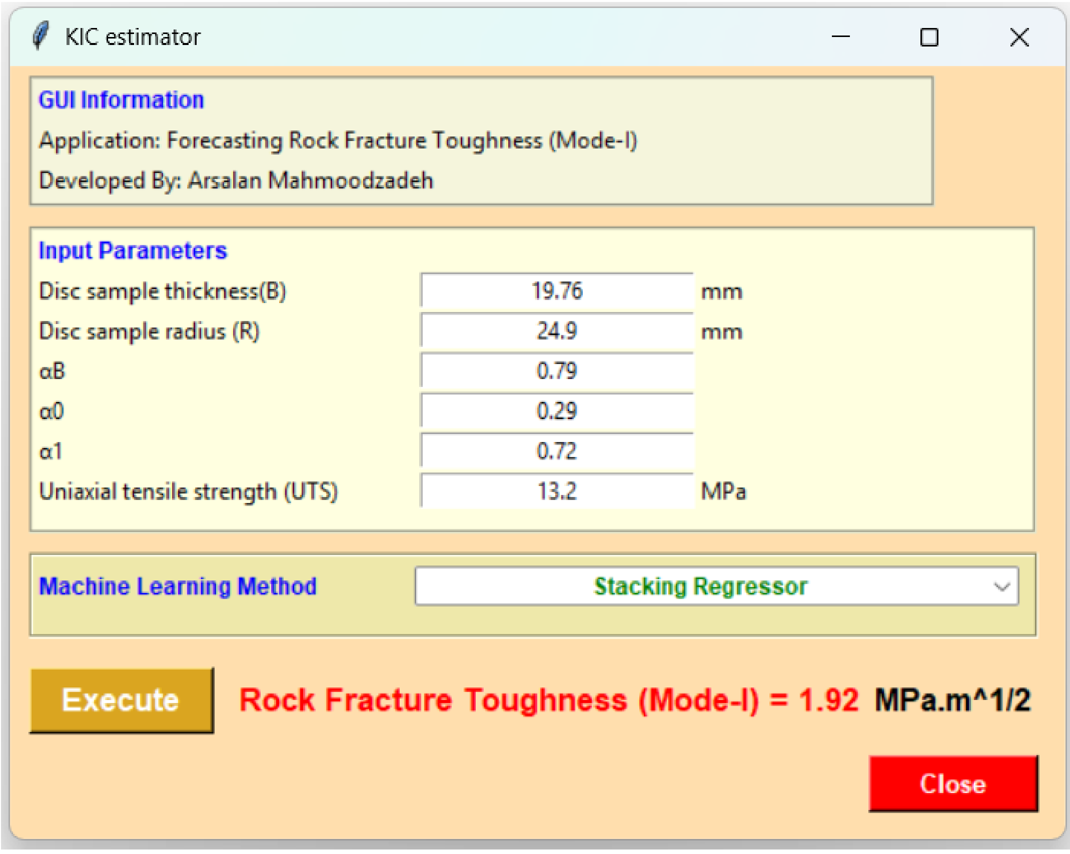

Fig. 9 The developed ML-based GUI to estimate K_{IC}

output parameters, quantifying the amount of information obtained about the output by observing each input. Higher mutual information values indicate a stronger relationship between input and output parameters. Eq. (2) evaluates the mutual information values between each input parameter and the output (K_{IC}).

$$MI = \sum_{x \in Input\ Parameters} \sum_{y=K_{IC}} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \quad (2)$$

where, x and y are input and output parameters, respectively.

By examining the mutual information values, it was established that the UTS parameter exerts the most significant influence on the K_{IC}, with an MIT score of 2.35 (Fig. 8). This high score indicates a strong dependency between the UTS and the K_{IC}, highlighting the critical role that tensile strength plays in determining the material's resistance to fracture. Understanding this relationship is essential for developing more accurate predictive models and informing the design and implementation of engineering projects where KIC is crucial.

8. Graphical user interface

This research also develops an ML-based graphical user interface (GUI) to predict the K_{IC} parameter. This GUI is designed to be user-friendly, allowing engineers and researchers to input the six relevant parameters (UTS, B, R, α_0, α_1, α_B) without requiring extensive expertise in ML or programming. The interface facilitates the seamless integration of data input, model computation, and results visualization. Users can easily upload their dataset, train the model, and obtain predictions with minimal effort. Additionally, the GUI includes options for selecting different ML models, enabling users to compare the performance of various ML algorithms.

The development of this GUI represents a significant step towards practical application, bridging the gap between advanced ML techniques and their usability in real-world engineering scenarios. By streamlining the prediction process and making it more accessible, the GUI contributes to more efficient and accurate assessments of RFT, ultimately supporting better-informed decision-making in geotechnical and civil engineering projects.

9. Key limitations and suggestions

9.1 Key limitations

- Limited dataset: The study was conducted using 400 data points from granite specimens, which may not fully represent the variability in RFT across different rock types and geological conditions.
- Generalization to other rock types: The model was trained explicitly on granite samples, and its performance on other rock types remains to be determined. This limits the findings' generalizability to other geological contexts.
- Simplified assumptions: The study assumes that the six input parameters (UTS, R, B, α_B, α_0, α_1) are sufficient to predict K_{IC} accurately. However, other factors, such as the rock's environmental conditions and microstructural properties, might also influence fracture toughness.

9.2 Suggestions for future research

- Expand the dataset: Future studies should include a larger and more diverse dataset encompassing various rock types and geological settings to enhance the model's robustness and generalizability.
- Incorporate additional parameters: Consider including other relevant parameters, such as mineralogical composition, porosity, and environmental factors, to capture a more comprehensive range of influences on rock fracture toughness.
- Uncertainty analysis: Implement uncertainty quantification methods to understand the confidence intervals of the predictions better and communicate the model outputs' reliability to stakeholders.
- Cross-disciplinary approaches: Collaborate with geologists, material scientists, and field engineers to integrate interdisciplinary knowledge and improve the model's accuracy and applicability.

By addressing these limitations and incorporating these suggestions, future research can build on this study's findings, leading to more accurate, reliable, and practical models for predicting RFT in various engineering and geological contexts.

10. Conclusions

The conclusions drawn from this research highlight the efficacy and advantages of using a stacking regressor model to predict the K_{IC} of rocks. Key findings include:

- As various statistical indices demonstrated, the stacking regressor model significantly outperformed individual base models (GPR, SVR, RF, LR, KNN, and ETR) in terms of prediction accuracy for the KIC parameter.
- Combining the strengths of multiple base models, the stacking regressor method leverages ensemble learning to achieve more reliable and precise predictions than single-model approaches.
- The meta-model, constructed using the GPR model, demonstrated robustness and generalizability, providing accurate predictions across different subsets of data as confirmed through K-fold cross-validation.
- The development of an ML-based GUI enhanced the model's practical applicability, making it accessible to engineers and researchers without extensive expertise in ML.
- Sensitivity analysis using the MIT method revealed that the UTS parameter exerts the most significant influence on the K_{IC}, underscoring its importance in predictive modeling.
- The study suggested that further refinement of the model and incorporation of additional data could enhance predictive accuracy even further. It also encouraged exploring other advanced ML techniques to address the inherent variability in rock properties.

Acknowledgments

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/01/29720).

References

Abdelmawla, A., Ma, S., Yang, J.J. and Kim, S.S. (2023), “Subsurface anomaly detection utilizing synthetic GPR images and deep learning model”, *Geomech. Eng.*, **33**(2), 203-209. https://doi.org/10.12989/gae.2023.33.2.203.

Chandra, D. and Vishal, V. (2021), “A critical review on pore to continuum scale imaging techniques for enhanced shale gas recovery”, *Earth-Sci. Rev.*, **217**, 103638. https://doi.org/10.1016/j.earscirev.2021.103638.

Chang, S.H., Lee, C.I. and Jeon, S. (2002), “Measurement of rock fracture toughness under modes I and II and mixed-mode conditions by using disc-type specimens”, *Eng. Geol.*, **66**(1-2), 79-97. https://doi.org/10.1016/S0013-7952(02)00033-9.

Gudmundsson, A. (2014), "Elastic energy release in great earthquakes and eruptions", *Front. Earth Sci.*, 2. https://doi.org/10.3389/feart.2014.00010.

Guha Roy, D., Singh, T.N. and Kodikara, J. (2018), "Predicting mode-I fracture toughness of rocks using soft computing and multiple regression", *Measurement*, **126**, 231-241. https://doi.org/10.1016/j.measurement.2018.05.069.

Guha Roy, D., Singh, T.N., Kodikara, J. and Talukdar, M. (2017), "Correlating the mechanical and physical properties with mode-I fracture toughness of rocks", *Rock Mech. Rock Eng.*, **50**(7), 1941-1946. https://doi.org/10.1007/s00603-017-1196-0.

Hashempour, S., Boostani, R., Mohammadi, M. and Sanei, S. (2022), "Continuous scoring of depression from EEG signals via a hybrid of convolutional neural networks", *IEEE T. Neural Syst. Rehab. Eng.*, **30,** 176-183. https://doi.org/10.1109/TNSRE.2022.3143162.

Kamran, M., Shahani, N.M. and Armaghani, D.J. (2022), "Decision support system for underground coal pillar stability using unsupervised and supervised machine learning approaches", *Geomech. Eng.*, **30**(2), 107-121. https://doi.org/10.12989/gae.2022.30.2.107.

Khan, N.A., Mohammadi, M. and Djurović, I. (2019), "A modified viterbi algorithm-based IF estimation algorithm for adaptive directional time–frequency distributions", *Circui. Syst. Signal Pr.*, **38**(5), 2227-2244. https://doi.org/10.1007/s00034-018-0960-z.

Lawal, A.I., Kim, M. and Kwon, S. (2023), "Soft computing based mathematical models for improved prediction of rock brittleness index", *Geomech. Eng.*, **33**(3), 279-289. https://doi.org/10.12989/gae.2023.33.3.279.

Liu, Y. and Dai, F. (2021), "A review of experimental and theoretical research on the deformation and failure behavior of rocks subjected to cyclic loading", *J. Rock Mech. Geotech. Eng.*, **13**(5), 1203-1230. https://doi.org/10.1016/j.jrmge.2021.03.012.

Mahmoodzadeh, A., Mohammadi, M., Daraei, A., Faraj, R.H., Mohammed Dler Omer, R. and H. Sherwani, A.F. (2020), "Decision-making in tunneling using artificial intelligence tools", *Tunn. Undergr. Sp. Tech.*, **103**, 103514. https://doi.org/10.1016/j.tust.2020.103514.

Mahmoodzadeh, A., Mohammadi, M., Hashim Ibrahim, H., Gharrib Noori, K.M., Nariman Abdulhamid, S. and Farid Hama Ali, H. (2021), "Forecasting sidewall displacement of underground caverns using machine learning techniques", *Automat. Constr.*, **123**, 103530. https://doi.org/10.1016/j.autcon.2020.103530.

Mahmoodzadeh, A., Nejati, H.R., Mohammadi, M., Hashim Ibrahim, H., Khishe, M., Rashidi, S. and Farid Hama Ali, H. (2022), "Prediction of mode-I rock fracture toughness using support vector regression with metaheuristic optimization algorithms", *Eng. Fract. Mech.*, **264**, 108334. https://doi.org/10.1016/j.engfracmech.2022.108334.

Wang, Y.T., Zhang, X. and Liu, X.S. (2021), "Machine learning approaches to rock fracture mechanics problems: Mode-I fracture toughness determination", *Eng. Fract. Mech.*, **253**, 107890. https://doi.org/10.1016/j.engfracmech.2021.107890.

IC