



A novel approach to estimate rock deformation under uniaxial compression using a machine learning technique

Pradeep T.¹ · Divesh Ranjan kumar² · Manish Kumar³ · Pijush Samui⁴ · Danial Jahed Armaghani⁵

Received: 2 August 2023 / Accepted: 5 June 2024 / Published online: 14 June 2024
© The Author(s) 2024

Abstract

Understanding rock deformation is crucial for various engineering and geological applications, including mining, tunneling, and earthquake prediction. In this study, we propose a novel approach to estimate rock deformation under uniaxial compression using extreme gradient boosting (XGB), Extra trees regression (ETR), and K-Nearest Neighbours (KNN) algorithms. The proposed methodology involves three main steps. First, a comprehensive dataset of rock samples is collected, including various positions of the strain gauge, stress, and corresponding deformation measurements under uniaxial compression. These properties serve as input and output features for the machine learning models. Second, the XGB, ETR, and KNN algorithms are trained and tested using the collected dataset. These algorithms are known for their ability to handle complex relationships and nonlinearities, making them suitable for modeling the intricate behavior of rock deformation under compression. To ensure accurate predictions, a cross-validation technique is employed to optimize the hyperparameters of each algorithm. The trained models are then evaluated using various performance evaluations like performance parameters, Actual and predicted curves, Rank analysis, Sensitivity Analysis, Error matrix, and OBJ criteria. All models perform better (i.e., coefficient of determination greater than 0.9), however, XGB is a more robust model when compared to other models. Overall, this study presents a novel and promising approach to estimating rock deformation under uniaxial compression, offering a valuable tool for engineers and geologists working in the field of rock mechanics.

Keywords Rock strain · Strain gauge · Extreme gradient boosting · Extra trees regressor · K-Nearest neighbors

Introduction

Rocks, being solid materials, have the ability to undergo deformation when subjected to stress. Stress can be caused by various factors such as tectonic forces, gravitational forces, temperature changes, and human activities like mining and excavation. When stress is applied to a rock mass,

it results in rock strain, which is the measurable change in the shape or size of the rock. Measuring and analyzing rock strain is essential for understanding rock behavior and stability. Various techniques are used to quantify and monitor strain in rocks, including strain gauges, extensometers, and geological surveys (Petr et al. 2016; Xu et al. 2020). These measurements provide valuable data for assessing the

✉ Divesh Ranjan kumar
diveshk.ph21.ce@nitp.ac.in

Pradeep T.
tpradeep1125@gmail.com

Manish Kumar
mkumarod@gmail.com

Pijush Samui
pijush@nitp.ac.in

Danial Jahed Armaghani
danial.jahedarmaghani@uts.edu.au

² Research Unit in Data Science and Digital Transformation, Department of Civil Engineering, Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Pathumthani, Thailand

³ SRM Institute of Science and Technology (SRMIST), Deemed to Be University, Tiruchirappalli, Tamil Nadu, India

⁴ Department of Civil Engineering, National Institute of Technology, Patna, India

⁵ School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

¹ Department of Civil Engineering, National Institute of Technology Andhra Pradesh, Tadepalligudem, India

mechanical properties of rocks, predicting rock failure, and designing safe structures. Numerous studies on the identification of strain have been undertaken thus far as a result of the requirement for strain identification (Gundewar 2014; Laghaei et al. 2018; Isah et al. 2020; Li et al. 2020). In the study of rock, strain tensors for two main reasons. One is the assessment of intricate stress fields and the second is an estimation of complex strain fields through loading tests (Kikumoto and Togashi 2022). Monitoring of strain or displacement is frequently used to study the origin, development, and destruction of rock fractures. Rock failure could be forecast using the variation coefficient, according to a relation between the variation coefficient of rock surface displacement and the peak stress (Kong et al. 2023).

Rock strain prediction plays a crucial role in various fields such as geotechnical engineering, mining, and natural hazard assessment (Liu et al. 2013). Understanding the deformation behavior of rocks under different loading conditions is essential for ensuring the safety and stability of underground structures, optimizing mining operations, and assessing the potential risks associated with geological activities. Traditionally, rock strain prediction has relied on empirical and analytical models that require extensive data collection and manual analysis (Vergara et al. 2020). Conducting experiments in a lab can be expensive, requiring resources such as equipment, materials, and personnel. In some scientific experiments, obtaining representative samples or cores can be challenging, especially when dealing with large or inaccessible areas. Certain experiments may require long observation periods or involve time-sensitive processes. Analyzing large volumes of data generated in the lab can be a laborious and time-consuming task. However, advancements in machine learning techniques have opened up new possibilities for more accurate and efficient prediction models (Liu et al. 2017). Numerous academic institutions, research organizations, and industry companies have ongoing research projects and publications on this topic (Liu et al. 2017; Cioffi et al. 2020). Many researchers in the fields of geology, geotechnical engineering, and computational science are exploring the use of machine learning for predicting for different purposes as well as rock strain (Tariq et al. 2017; Cui et al. 2019; Kumar and Samui 2019; Sun et al. 2020; Asteris et al. 2021; Jahed Armaghani et al. 2021; Yu et al. 2021; Koopialipoor et al. 2022; Kumar et al. 2022c; Abdi et al. 2023a).

It's important to note that while machine learning can be a powerful tool, it is not a one-size-fits-all solution. The successful application of machine learning in the lab requires careful consideration of the specific problem, availability and quality of data, appropriate algorithm selection, and continuous validation and refinement of models to ensure accurate and reliable results (Indraratna et al. 2023; Medawela et al. 2023). Several researchers utilized the

machine learning algorithm to solve complex and non-linear problems in different fields, especially in engineering and science (Al-Jeznawi et al. 2023; Tran et al. 2024; Jitchaijaroen et al. 2024). The estimation of soil resilient modulus using several machine learning algorithms was proposed by various researchers in recent few years (Sadik 2023; Sadik et al. 2024). The single models require more computational resources, can be more complex to implement, and may have higher training times. Additionally, the effectiveness of ensemble methods depends on the diversity and quality of the individual models. Nevertheless, when appropriately applied, ensemble learning algorithms have demonstrated significant advantages in various domains and have become a valuable tool in machine learning (Dietterich 2000; Sagi and Rokach 2018; Kunapuli 2023). (Momeni et al. 2015) predict the uniaxial compressive strength of rock samples using hybrid artificial neural networks augmented by particle swarm optimization techniques. (Mohamad et al. 2015) predict the the unconfined compressive strength (UCS) of soft rocks based on ANN hybrid swarm based optimization algorithm. (Mohamad et al. 2018) proposed the back-propagation (BP) artificial neural network (ANN) to estimate the uniaxial compressive strength (UCS) of rocks with a quite accurately having the R values equal to 0.988 for training phase and 0.999 for testing phase. (Nazir et al. 2013a) proposed the a new, reliable correlation between UCS and brazilian tensile strengthBTS for limestone, achieving a strong 0.9 determination coefficient. Several machine learning technique have been utilized to predict the UCS of rock and prediction of elastic modulus of weak rock mass sample (Nazir et al. 2013b; Momeni et al. 2015; Armaghani et al. 2020; Abdi et al. 2023b; Dowlatshahi et al. 2023).

In this study, ensemble learning algorithms are used to predict rock strain. "XGB: A Scalable Tree Boosting System" by (Chen and Guestrin 2016), is the original paper introducing XGB, and provides an overview of the algorithm, as well as its advantages over other gradient boosting implementations. Its speed and scalability, as it can effectively handle large datasets with a high number of features. XGB is highly robust to noise and outliers in the data, which makes it suitable for real-world applications where data quality may be suboptimal. It provides a feature importance metric that allows users to identify the most important features in their dataset. Overall, XGB is a powerful machine-learning library that offers significant advantages over traditional single-model approaches, including improved accuracy, speed, scalability, and interpretability (Chen and Guestrin 2016; Ma et al. 2022; Su et al. 2022). Extra Trees Regression (ETR) is a powerful ensemble learning method that can be used to improve the accuracy of regression tasks. The algorithm is based on decision trees and uses multiple random subsets of the data and features to construct the trees. There have been

many research papers on Extra Trees Regression, and the results have shown that the algorithm can achieve high accuracy while requiring less training time than other methods (Geurts et al. 2006; Ahmad et al. 2018; Kumar et al. 2023c). KNN is a widely used and well-studied machine learning algorithm that has been applied to various domains. The literature on KNN includes both theoretical and practical aspects, and there have been many developments and improvements to the algorithm over the years. (Bhatia and Vandana 2010; Stevens et al. 2011). The limitations, Extra Trees can suffer from high variance, especially when the number of trees in the forest is large. This might lead to overfitting, especially on smaller datasets. While decision trees are generally interpretable, the ensemble nature of Extra Trees makes it harder to interpret the model's decision-making process compared to a single decision tree. tuning of hyperparameters like the number of trees, maximum depth, and minimum samples per leaf, which can be time-consuming and computationally expensive. XGBoost has several hyperparameters that need to be tuned, such as the learning rate, maximum depth, minimum child weight, etc. The performance of XGBoost can be sensitive to these parameters, and finding the optimal set of parameters might require extensive computational resources. Similar to other gradient-boosting algorithms, XGBoost can be prone to overfitting, especially when the dataset is noisy or when the number of trees (iterations) is too large. Prediction with KNN involves calculating distances between the query instance and all training instances, which can be computationally expensive, especially with large datasets or high-dimensional data. KNN typically requires storing the entire training dataset in memory, which can be memory-intensive for large datasets, limiting its scalability.

Predicting rock strain using machine learning is a promising area of research that has the potential to improve our understanding of rock deformation behavior and aid in the exploration and exploitation of geological resources. In this study, ensemble models are applied to estimate the strain in the rock sample by data collected from the laboratory experiment. The data consists of pairs of input and output data, where the input data is used to make a prediction and the output data represents the correct or desired output for that input. The input features are the Height of the strain gauge, the Angle of strain gauge, and Stress. There will be development, training, and testing of the model. Overall, model result analysis is a crucial step in the machine learning process, as it helps to evaluate the performance of a model. The results of all models must be compared in order to select the most robust model using the performance parameter, Actual and predicted curve, Rank analysis, Sensitivity Analysis, Error matrix, and OBJ criteria.

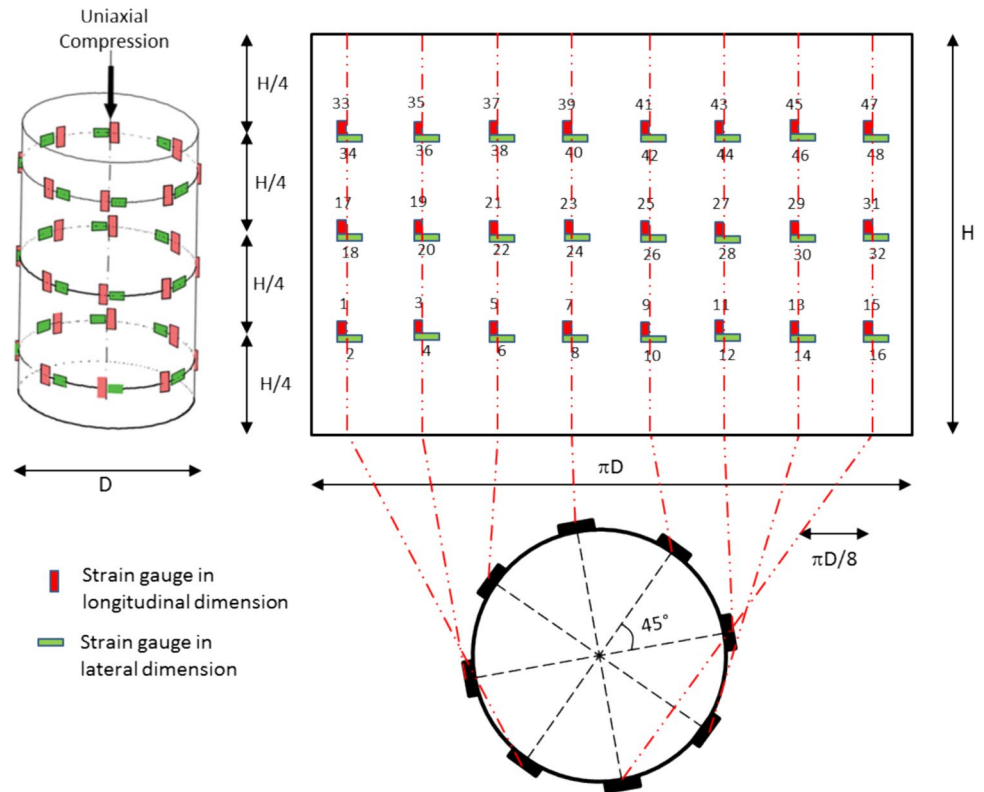
Research significance

Many regions are prone to natural hazards such as earthquakes, landslides, and rock falls. Climate change is causing significant shifts in environmental conditions, which can impact rock behavior and stability. By studying rock strain, researchers can gain insights into the behavior of rocks under high-stress conditions, which contributes to the understanding and prediction of earthquakes, landslides, and other geological hazards. Also affects the stability and performance of civil infrastructure, such as dams, bridges, and underground structures. This knowledge is crucial for developing strategies to mitigate the risks associated with such events. Predicting rock strain using ensemble models can aid in assessing the potential for these hazards and developing effective mitigation strategies. This research can contribute to enhancing the resilience of communities globally, minimizing the loss of lives and infrastructure, and improving disaster preparedness and response.

Details of data

In this study, the cylindrical granite rock sample having fixed diameter and height of 40 mm and 108 mm respectively was subjected to uniaxial compression loading to measure the load and deformation. Data from the lab was acquired during experimental testing to predict strain in a granite rock material. Measurements of the load and deformation on the longitudinal axis of the circumference of the rock sample were made using a uniaxial compression load cell and several strain gauge transducers. Figure 1 depicts the specifics of the strain gauge's setup. The outer perimeter was equipped with a total of 48 electronic strain gauges in both lateral direction and longitudinal directions. The readings of these loads and deformation were gathered in a data-gathering system. The collected dataset from the experiments was utilized to calculate stress and strain based on the rock's dimensions. The average Poisson ratio of this granite rock is 0.27. To develop the machine learning models total of 3000 datasets of granite rock samples were acquired which include the variables namely, height of rock samples, angles, stress of the strain gauge, and strain in lateral and longitudinal directions. Measuring strain at different heights within the rock mass offers insights into stress distribution vertically and identifies potential zones of weakness or concentration. Rocks often exhibit anisotropic behavior, signifying variation in mechanical properties with direction. Orienting strain gauges at specific angles aids in understanding

Fig. 1 Illustration of loading and strain gauges under uniaxial load



stress distribution along different axes, characterizing rock behavior under varying loading conditions. Therefore, to train and test the machine learning models the height, angle, and stress of the strain gauge are considered input variables. Whereas, the lateral and longitudinal strains are considered as the output variable. Soft computing techniques were used to predict the strain in a rock sample using these input and output data. The sample datasets are shown in Table 1.

After the input and output selection, the whole dataset is normalized in the range between [0,1] to standardize the formatting of the dataset. The main aim of normalization is to reduce the dimensional effect of different variables on the output. The following mathematical Eq. 1 has been utilized to normalize the input and output variables (Singh and Singh 2020). Out of the 3000 datasets, 2100 datasets were utilized for model training (TR) (about 70% of the total), while 900 datasets were used for model testing (TS) (30%). To ensure

Table 1 Sample dataset

Height (mm)	Angle ϕ (degrees)	Stress σ (MPa)	ϵ_x ($\cdot 10^{-3}$)	ϵ_y ($\cdot 10^{-3}$)
0	1	-0.90778	-0.68685	0.907419
1	-1	-0.87465	-0.68685	0.875976
0	-0.1428	-0.86087	-0.58059	0.875976
-1	-1	-0.81509	-0.58059	0.844533
0	0.7142	-0.79744	-0.68685	0.844533
0	-1	-0.73326	-0.58059	0.781648
0	-1	-0.69919	-0.58059	0.750205
0	-1	-0.67778	-0.58059	0.68732
0	-1	-0.66547	-0.58059	0.68732
0	-1	-0.60549	-0.58059	0.655877
0	-1	-0.57974	-0.58059	0.655877
0	-1	-0.5529	-0.58059	0.592992
0	-1	-0.48033	-0.58059	0.561549
0	-1	-0.45013	-0.58059	0.498663

the proper learning and validation of the models, the authors have selected the widely popular and successful criteria from the literature review.

$$N = \frac{N_a - N_{min}}{N_{max} - N_{min}} \tag{1}$$

where N is the Normalized value, N_a is actual value, N_{min} is Minimum value and N_{max} is Maximum value.

Statistical description of the dataset

Statistical analysis is a method of analyzing numerical data using statistical techniques to draw conclusions from the data Compared to many previous research types, a more comprehensive range of the database has been considered in this study. Table 2 summarizes a dataset of 3000 data points with five variables. The provided statistics offer insights into the central tendency (mean, median), spread (standard deviation), and distribution (quartiles) of the data for each variable.

Performance indices

Several performance indices along with the coefficient of determination (R^2), variance account factor (VAF), performance index (PI), Mean absolute error (MAE) (Chai and Draxler 2014), and Root mean square error (RMSE) were evaluated to check the performance of proposed models (Kumar et al. 2022a, 2023a, b). The mathematical expression of these performance indices along with their ideal values are presented in Table 3.

Table 3 Performance evaluation parameter with the ideal value

Parameters	Mathematical Expression	Ideal Value
R^2	$R^2 = \frac{\sum_{i=1}^n (A_i - A_{avg})^2 - \sum_{i=1}^n (A_i - P_i)^2}{\sum_{i=1}^n (A_i - A_{avg})^2}$	1
RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - P_i)^2}$	0
VAF	$VAF = \left(1 - \frac{var(A_i - P_i)}{var(A_d_i)}\right) \times 100$	100
PI	$PI = adj.R^2 + (0.01 \times VAF) - RMSE$	2
MAE	$MAE = \frac{1}{n} \sum_{i=1}^n (P_i - A_i) $	0

Where P indicates the predicted value of the target variable A indicates the experimental value of the target variable and n indicates the total number of datasets used in this study.

Machine learning algorithms

Extreme gradient boosting (XGB)

Extreme Gradient Boosting (XG-Boost) is a powerful and widely used machine learning algorithm for supervised learning problems, particularly in the fields of predictive modeling and data analysis. XG-Boost is based on the concept of boosting, where weak learners are combined to create a stronger, more accurate model. XG-Boost is an extension of the gradient boosting method, which iteratively adds decision trees to a model to improve its performance (Chen

Table 2 Statistical view of the dataset

Variable	H	ϕ	σ	ϵ_x	ϵ_y
Type	Input	Input	Input	Output	Output
Units	mm	degree	MPa	*10 ⁻³ mm	*10 ⁻³ mm
Number of data	3000	3000	3000	3000	3000
Percent	100	100	100	100	100
Mean	54	157.43	7.98	0.02	-0.09
Standard Deviation	22.05	103.07	4.54	0.01	0.06
Variance	486.23	10,624.54	20.65	0	0
Sum	162,000	472,500	523,717.7	1880.8	-6843.87
Maximum	81	315	16	0.06	0.01
Minimum	27	0	0	-0.01	-0.23
Q1	27	78.75	4.16	0	-0.13
Median	54	157.5	8	0.02	-0.08
Q3	81	225	11.74	0.03	-0.03
Mode	54	0	2.34	0	-2.14805
Skewness	0	0	-0.02	0.45	-0.31
Kurtosis	1.5	1.76	1.82	2.39	2.04

and Guestrin 2016; Bentéjac et al. 2021). The complexity of the trees is regulated by a variation of the loss function, as shown by Eq. (2), where T represents the number of leaves in the tree and w represents the output scores of the leaves. XG-Boost improves upon traditional gradient boosting by incorporating regularization to prevent overfitting, and by optimizing the objective function using second-order gradients, which leads to faster convergence and better accuracy.

$$L_{xgb} = \sum_{i=1}^N L(y_i, F(x_i)) + \sum_{m=1}^M \Omega(h_m) \quad (2)$$

$$\Omega(h) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

The split criterion of decision trees in XG-Boost can incorporate a user-defined loss function, which allows the algorithm to optimize for a specific metric that is relevant to the problem at hand. This can be used as a pre-pruning strategy by setting a threshold on the gain in the loss function that must be achieved by a potential split for it to be considered. If the gain is below the threshold, the split is not made, and the tree is pruned.

The loss function used in XG-Boost is typically a sum of two terms: a term that measures the goodness of fit of the model to the training data, and a regularization term that penalizes complex models to prevent overfitting. The regularization term includes a parameter γ , which controls the complexity of the tree. Trees with higher values of γ are simpler, as they are penalized more heavily for having many splits or a large number of leaf nodes.

In XG-Boost, the optimal value of γ can be selected using cross-validation or other methods, and this can help prevent overfitting and improve the generalization performance of the model. By combining regularization with the use of a pre-pruning strategy based on a loss function, XG-Boost can produce accurate and interpretable models that are robust to noise and outliers in the data (Kumar et al. 2022b).

Extra Tree Regressor (ETR)

The Extra Trees Regression (ETR) technique is a tree-based ensemble machine learning approach designed to mitigate the overfitting issue inherent in the original random forest algorithm (Geurts et al. 2006). The mathematical formulation used in this approach is the same as that of the decision tree regression (DTR) algorithm proposed by (Ahmad et al. 2018). Instead of utilizing the bagging approach to build the training subset for each tree, all datasets are used in this strategy to train all trees in an ensemble. By merging a certain number of estimators, ensemble algorithms aim to lower the model's variance relative to the variance of a single tree. The result is a model with greater stability and

generalization potential for the geotechnical practitioner. Such models predict a final output that is simply the mean of all individual trees' outputs that are causally in the same family. This algorithm's model includes a measure of unpredictability, which is one of its distinguishing features. The process of randomness is incorporated in two ways: (i) either a random subset of features is selected from the whole collection of features to be used in producing a split, (ii) or the thresholds of the selected features are chosen at random. On the basis of above stated two arbitrary choices are used in the following formula to get the optimal split. One benefit of utilizing such ensembled models is that, in most cases, no considerable hyperparameter adjustment is required to create a good-quality model, despite the fact that such models cannot be regarded as DTR models. However, the most crucial factor that should not be overlooked while creating ETR models is the selection of the number of estimators to be utilized.

K-nearest neighbor

Fix and Hodges (Fix and Hodges Jr 1952) and Altman (Altman 1992) created the k-nearest Neighbors algorithm (k-NN) in 1951 as a non-parametric classification technique. It belongs to the family of instance-based learning methods, where the algorithm doesn't explicitly learn a model but instead uses the training instances themselves to make predictions on new instances. The object's property value is the outcome of the k-NN regression. As k-NN predictions are based on the intuitive premise that neighboring objects may be similar, it makes sense to discriminate between the K nearest neighbors when making predictions. Like lazy learning, neighbors-based regression just stores examples from the training data rather than attempting to build a generic internal model (Roh et al. 2020). The choice of k can have a significant impact on the performance of the k-NN algorithm. A small value of k can lead to overfitting, where the algorithm captures the noise in the data and produces overly complex decision boundaries. On the other hand, a large value of k can lead to underfitting, where the algorithm fails to capture the local structure of the data and produces overly simplistic decision boundaries. One of the advantages of the k-NN algorithm is that it doesn't require a training phase to determine the model parameters, unlike many other machine learning models. Instead, the training dataset serves as a reference dataset to determine the number of nearest neighbors (k) to consider when making predictions. Calculating the average of the output strain in the lateral and longitudinal direction of the K nearest neighbors forms the foundation of the straightforward mathematical application of KNN regression. Generally, a straightforward KNN regression algorithm used at least three different distance functions.

(1) Euclidean distance function (E_d), (2) Manhattan distance function (M_d), and (3) Minkowski distance function (M_{id}), mathematical expression presented in Eq. (4), (5) and (6) respectively.

$$E_d = \sqrt{\sum_{i=1}^K (x_i - y_i)^2} \tag{4}$$

$$M_d = \sum_{i=1}^K |x_i - y_i| \tag{5}$$

$$M_{id} = \left\{ \sum_{i=1}^K (|x_i - y_i|)^q \right\}^{1/q} \tag{6}$$

Result and discussion

Modeling parameters of proposed models

In this study, three advanced machine learning models i.e., ETR, XGB, and KNN were proposed to predict the strain of lateral and longitudinal direction. It is important to note the optimum models were developed by following a trial-and-error approach used for selecting hyperparameters of proposed models in Python using the Scikit-optimizer package. In Table 4, the optimum hyperparameters of the XGBoost model, such as the number of estimators, number of iterations, learning rate, and number of boosting rounds, are listed along with their corresponding values.

The same dataset used for the construction of the ERT and KNN model and optimum hyperparameters with their corresponding values for the ETR model are mentioned in Table 5. These parameters include the number of estimators, maximum depth, maximum number of features, minimum number of leaf nodes, minimum number of samples to split a node, and the number of jobs to run in parallel. In the application of the KNN model the neighbor value considered one (i.e., $K = 1$) in this study, and other hyperparameters are taken as the default in the original study of the KNN algorithm.

Performance evaluation

To evaluate and compare the performance of proposed models, ten performance indices namely, R-squared (R^2), Adjusted R-squared ($AdjR^2$), Weighted Mean Absolute Percentage Error (WMAPE), Nash–Sutcliffe Efficiency (NS), Root Mean Square Error (RMSE), Variance Accounted For (VAF), Prediction Interval (PI), RMSE-observations Standard Deviation Ratio (RSR), Willmott’s Index of Agreement (WI), and Mean Absolute Error (MAE), were determined for both lateral and longitudinal dimensions presented in Table 6 and 7 respectively. The results presented in Tables 5 and 6 provide the proposed model’s performance in quantitative manners for both lateral (x) and longitudinal (y) dimensions. For an ideal model performance indices value should be equal to their respective ideal value mentioned in Table 2. Generally, the models that achieved the higher accuracy parameters and lower error parameter values are considered the

Table 4 Result of optimal hyperparameters value for XGB model

XGB hyperparameter	Description	Search Range	Optimal value
n_estimators	Number of estimators	[50–500]	200
Learning rate	Learning rate	[0.001–0.1]	0.03
Max_depth	Maximum depth	[3–15]	10
Objective	Type of objective	['squared_error', 'pseudo_huber']	squared_error
nround	Total number of boosting rounds	[5–20]	12

Table 5 Result of optimal hyperparameters value for ETR model

ETR hyperparameters	Description	Search Range	Optimal value
n_estimators	Number of estimators	[50–500]	100
max_depth	Maximum depth	[3–15]	12
Objective	Type of objective	['squared_error', 'pseudo_huber']	squared_error
min_samples_leaf	Minimum number of leaf node	[0–5]	1
subsample	Sample Split	[0.7–1.0]	0.7

Table 6 performance indices in lateral (x) dimension

Models Parameters	ETR		XGB		KNN	
	TR	TS	TR	TS	TR	TS
R ²	0.888	0.878	0.991	0.963	0.974	0.961
RMSE	0.142	0.150	0.042	0.077	0.065	0.082
VAF	87.62	86.47	99.04	96.32	97.36	96.04
PI	1.622	1.592	1.939	1.849	1.882	1.839
MAE	0.088	0.097	0.031	0.062	0.052	0.064

Table 7 Performance indices in longitudinal (y) dimension

Models Parameters	ETR		XGB		KNN	
	TR	TS	TR	TS	TR	TS
R ²	0.900	0.905	0.999	0.997	0.998	0.996
RMSE	0.172	0.172	0.019	0.029	0.024	0.035
VAF	88.983	89.69	99.87	99.68	99.77	99.54
PI	1.617	1.629	1.979	1.965	1.972	1.956
MAE	0.102	0.104	0.012	0.023	0.019	0.027

best-performing models. The proposed XGB models attained the maximum coefficient of determination and RMSE value ($R_{TR}^2 = 0.991, RMSE_{TR} = 0.042, R_{TS}^2 = 0.963, RMSE_{TS} = 0.077$) followed by KNN ($R_{TR}^2 = 0.974, RMSE_{TR} = 0.065, R_{TS}^2 = 0.961, RMSE_{TS} = 0.082$) and ETR ($R_{TR}^2 = 0.888, RMSE_{TR} = 0.142, R_{TS}^2 = 0.878, RMSE_{TS} = 0.150$) model during the both training and testing phases in the lateral (x) dimension. Similarly, the XGB model outperformed the KNN and ETR in the longitudinal (y) dimension. Finally, from the presented performance indices value for proposed machine learning models in both lateral (x) and longitudinal (y) dimensions presented in Tables 1 and 2, it can be concluded that the XGB model outperformed the KNN and ETR models.

Actual and predicted curve

The "actual vs predicted curve" in regression analysis pertains to a visual depiction that contrasts the observed or factual values of the dependent variable against the predicted values produced by a regression model. Visual assessment is a technique utilized to evaluate the degree of alignment between the predictions of a model and the actual data points. The Actual vs Predicted Curve displays a comparison between the observed values of the dependent variable in both the training and testing sets and the corresponding predicted values generated by the regression model. Each data depicted on the graph corresponds to an individual instance of observation. The comparison between the actual and predicted curve holds significance in regression analysis owing to its ability to offer a graphical evaluation of the model's efficacy. This enables one to comprehend the degree of precision with which the model captures the fundamental association between the independent and dependent variables.

Through a visual examination of the curve, it is possible to recognize discernible patterns, trends, or outliers, which may offer valuable insights into the efficacy and constraints of the regression model.

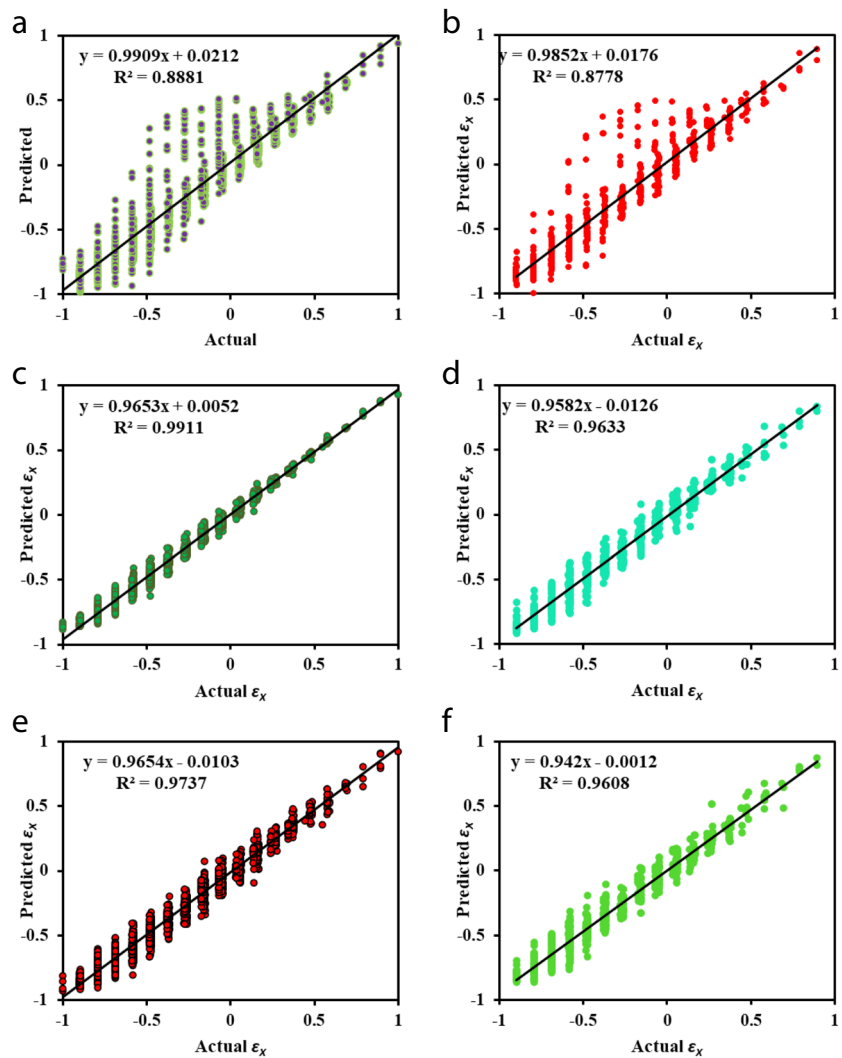
Moreover, the curve facilitates the assessment of the model's predictive precision. When the predicted values exhibit a high degree of similarity with the actual values, it suggests that the model is performing effectively. However, major outliers or patterns in the curve may point to places where the model should be strengthened or where more research is required. In general, the comparison between the actual and predicted curve is a valuable visual aid for evaluating a model, offering a rapid and intuitive means of gauging the efficacy and prognostic potential of a regression model.

In Fig. 2 and 3 (c and d), XGB model in both the cases and in both training and testing phases, the XGB model exhibits a curve that closely follows the diagonal line, indicating a strong agreement between the predicted and actual values. The performance of the KNN model in the Y-direction dataset is on par with the XGB, which is the best-performing model overall. The curve for the ETR model in the Y direction for both the training and testing phases exhibits deviations from the diagonal line and suggests discrepancies or errors in the model's predictions. In the X direction too, the ETR model's scatter plot doesn't indicate satisfactory performance. Thus, the performance of XGB and KNN is satisfactory in both cases, out of which XGB is the most robust model. The performance of ETR is not satisfactory in both datasets.

Rank analysis

The evaluation of soft computing models' performance is facilitated by score analysis or rank analysis. Through the

Fig. 2 Actual vs. predicted; Lateral dimension **a** ETR (Training) **b** ETR (Testing) **c** XGB (Training) **d** XGB (Testing) **e** KNN (Training) **f** KNN (Testing)



analysis of the scores produced by the model, it is possible to evaluate the model's capacity to effectively classify or forecast occurrences. Soft computing models have the potential to offer valuable insights into the level of confidence or uncertainty associated with their predictions, as inferred from the generated scores. Higher scores are indicative of a more robust correlation or probability of inclusion within a specific category, while reduced scores may suggest equivocation or indeterminacy. The analysis of scores is a useful tool for comprehending the degree of assurance exhibited by a model. This information is pertinent in various contexts, such as decision-making, risk evaluation, and the detection of cases that necessitate additional examination. On an individual basis, the technique assigned the highest rank to the model with the best value for each index and the lowest rank to the model with the poorest value for each index, for both the training and

testing results. The summation of the individual ranks was utilized to determine their ultimate score. Finally, the ultimate score for each model is determined by the aggregate ranks obtained from both the training and testing phases (Fig. 4).

The results of the rank analysis for the study are presented in Table 8 and Fig. 4 for the X-direction dataset, where TR stands for training and TS stands for testing. The XGB model is the top scorer in both the training phase (30) and the well-testing phase (30), followed by KNN (20 in both training and testing). In overall score as well, XGB scored highest (60), followed by KNN and ETR having scores of 40 and 20 respectively. Thus, this section concludes that XGB is the first-ranked model, KNN is second-ranked, and ETR is the lowest third-ranked. The score analysis for the Y-direction dataset

Fig. 3 Actual vs. predicted; longitudinal dimension **a** ETR (Training) **b** ETR (Testing) **c** XGB (Training) **d** XGB (Testing) **e** KNN (Training) **f** KNN (Testing)

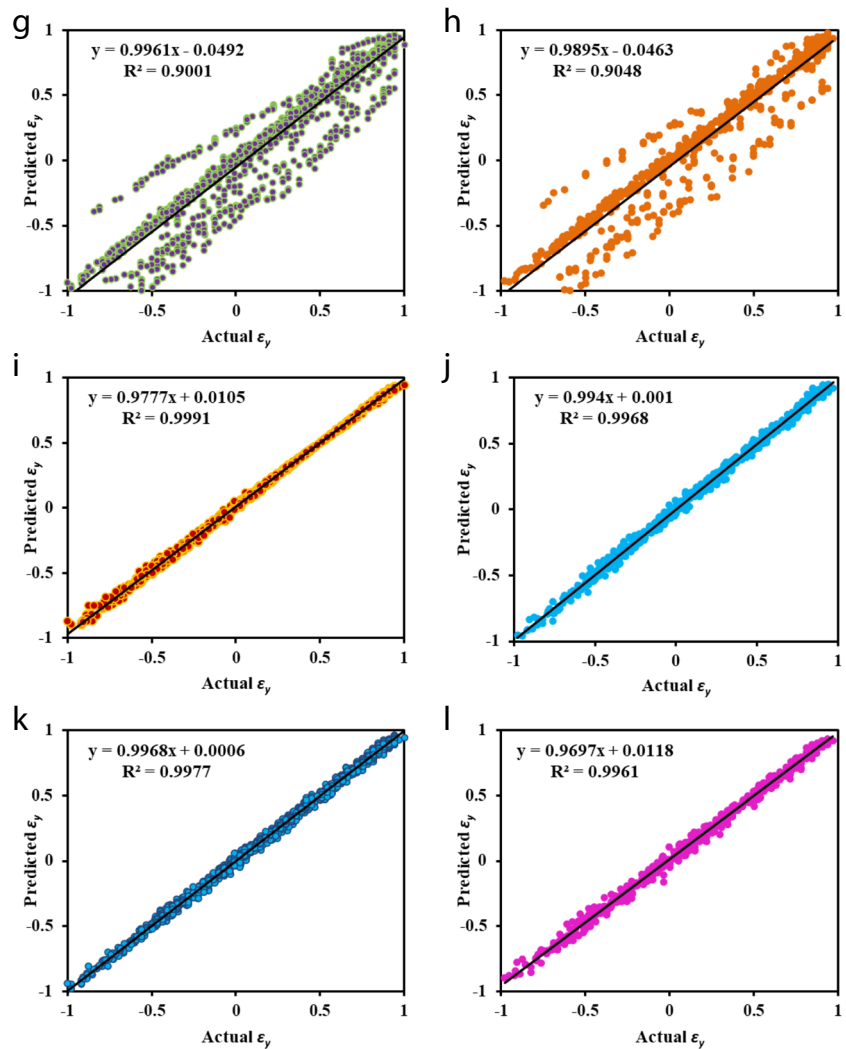


Table 8 Result of score analysis for lateral (x) and longitudinal (Y) dimension

Models	ETR		XGB		KNN	
	TR	TS	TR	TS	TR	TS
R^2	1	1	3	3	2	2
RMSE	1	1	3	3	2	2
VAF	1	1	3	3	2	2
PI	1	1	3	3	2	2
MAE	1	1	3	3	2	2
Sub Total	5	5	15	15	10	10
Total Score	10		30		20	
Rank	Third		First		Second	

is similar to that of the X-direction. Thus, this section concludes that XGB is the first-ranked model, KNN is second-ranked, and ETR is the lowest third-ranked. Overall, XGB is the best-performing model, followed by KNN and ETR respectively.

Sensitivity Analysis

Sensitivity analysis is a technique used to assess the impact of variations or changes in the input variables (i.e., Height, Angle, and stress) of a model on the output (i.e., lateral and longitudinal strain) of that model. It helps in understanding

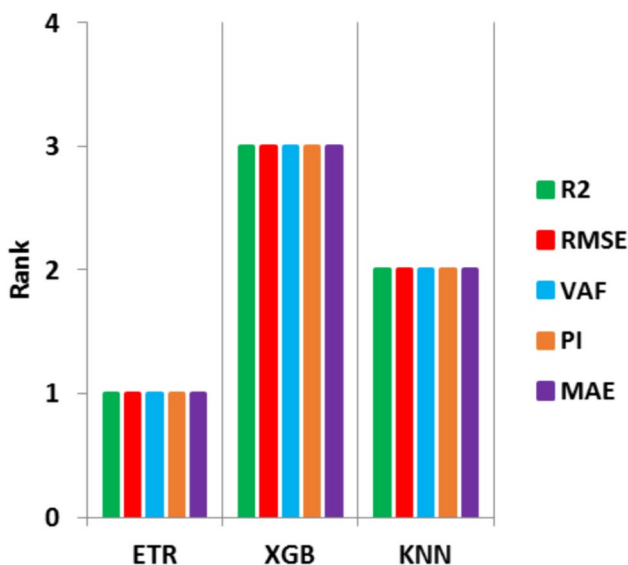


Fig. 4 Rank analysis for lateral (x) and longitudinal (Y) dimensions in both training and testing

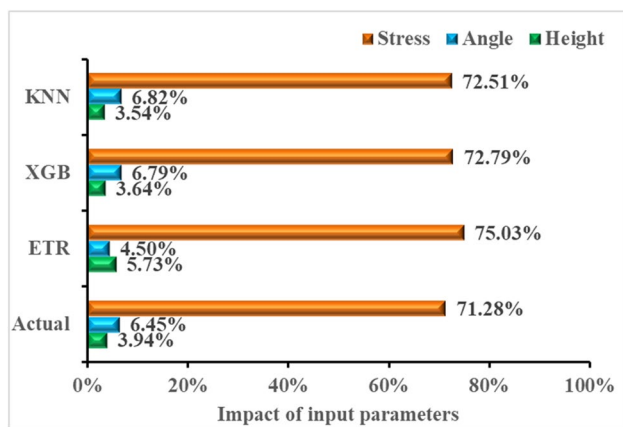


Fig. 5 Impact of input and output data with the predicted model (Lateral)

how sensitive the model is to different factors and allows for the identification of critical variables that have the most influence on the results. The purpose of sensitivity analysis is to gain insights into the relationships between the inputs and outputs of a system, evaluate the robustness of a model, and assess the risks and uncertainties associated with decision-making. By systematically varying the input variables within a defined range and observing the corresponding changes in the output, sensitivity analysis provides a quantitative measure of the impact of each input variable. In this study to determine the strength between input variables and output variables cosine amplitude method (CAM) has been used. The following mathematical expression is presented to calculate the impact strength

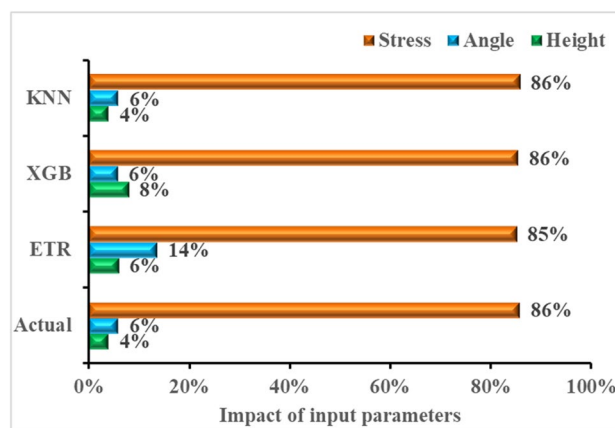


Fig. 6 Impact of input and output data with the predicted model (Longitudinal)

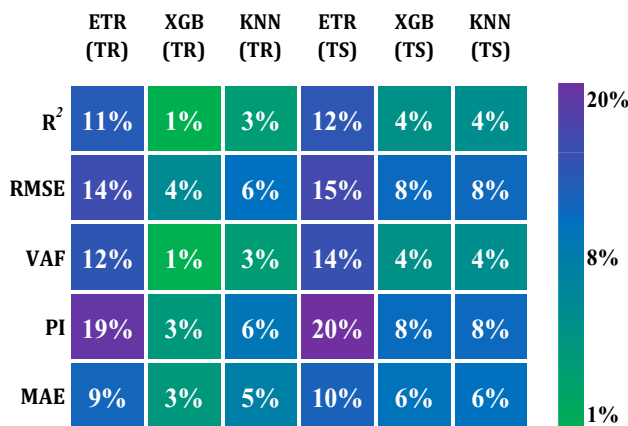


Fig. 7 Error matrix of lateral dimension

$$R_{ij} = \frac{\sum_{k=1}^n x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^n x_{ik}^2 \sum_{k=1}^n x_{jk}^2}} \times 100 \tag{7}$$

where the strength of correlation between the input data pair x_i and x_j and impact strength is denoted by R_{ij} . The value of R_{ij} closer to 100 means more impact of the corresponding input variable on the outstrain in both lateral (x) and longitudinal (y) direction. The impact of individual input parameters on output parameters for the actual dataset and all proposed models are presented in Figs. 5 and 6 for lateral and longitudinal direction respectively. From the presented result of sensitivity analysis, it can be concluded that the impact of stress in predicting the lateral and longitudinal strain is greater as compared to angle and height. In this study, stress has the highest influencing parameters in predicting the lateral and longitudinal strain followed by angle and height.

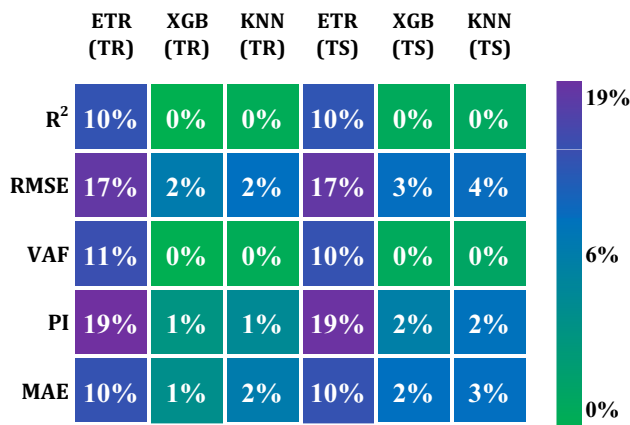


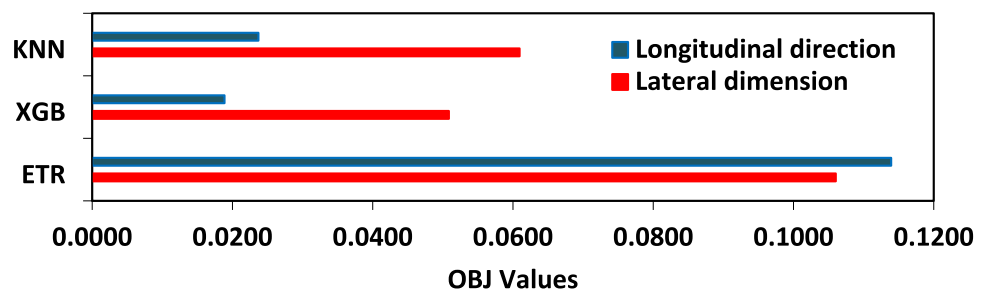
Fig. 8 Error matrix of longitudinal dimension

Error matrix

An error matrix is a table that summarizes the performance of a classification model. It provides a more detailed breakdown of the model's predictions by showing the number of correct and incorrect predictions for each class. Error matrix provides a detailed and intuitive representation of a regression model's performance, enabling deeper analysis, error identification, and informed decision-making. It is a valuable tool in evaluating, improving, and fine-tuning machine learning models.

The error matrix for both datasets is provided in Figs. 7 and 8. It can be observed that the XGB model in the training phase has an error of almost zero for the datasets of both, X and Y directions. In the testing phase as well, the errors exhibited by the XGB model are the lowest, compared to the other two models. Conversely, ETR exhibits errors as large as 10% and 10% for MAE criteria in training and testing, respectively. The errors for R^2 and RMSE are 12% and 15% respectively, for the testing phase. Thus, the matrix provides error recognition and comparison between the models, concluding the XGB and ETR as the best performing and the poorest model, respectively (Fig. 9).

Fig. 9 Illustration of OBJ value for lateral (x) and longitudinal (y) dimension



OBJ criteria

To determine the degree to which the generated solution is the genuine optimal solution (Hossein et al. 2013) suggested a set of objective (OBJ) criteria. The values of the objective function are used as a yardstick to determine the solution's quality. The mathematical equation used to characterize the OBJ value is presented as follows:

$$OBJ = \left(\frac{N_{TR} - N_{TS}}{N_{Total}} \right) \times \left(\frac{MAE_{TR}}{R^2_{TR}} \right) + \left(\frac{2N_{TS}}{N_{Total}} \right) \times \left(\frac{MAE_{TS}}{R^2_{TS}} \right) \tag{8}$$

where, N_{TR} and N_{TS} is the number of training data and testing data. R^2_{TR} = coefficient of determination for the training phase and R^2_{TS} = coefficient of determination for training phase. Similarly, MAE_{TR} = mean absolute error for the training phase and MAE_{TS} = mean absolute error for the training phase. For the best model, the OBJ value should be near zero, while for the ideal model, it should be zero. From Fig. 8 it can be concluded that the XGB is the best performing model. The values in the Y direction dataset is particularly are more encouraging, which indicates that the performance of the models varies as per the datasets and it's important to train and test the models before application to the field datasets. The performance of KNN is close to the XGB, however, that of ETR is not satisfactory.

Conclusion

In conclusion, this study successfully implemented three advanced machine learning models—Extra Trees Regressor (ETR), eXtreme Gradient Boosting (XGB), and K-Nearest Neighbors (KNN)—to predict the strain in lateral and longitudinal directions. The optimal models were developed using a trial and error approach to select hyperparameters via the Scikit-optimizer package in Python. The XGB model, in particular, demonstrated superior performance, with its optimal hyperparameters, such as the number of estimators and learning rate, carefully tuned to achieve the best results as listed. Similarly, the optimal parameters for

the ETR model are detailed in Table 4, and the KNN model was configured with the neighbor value set to one. The performance of these models was evaluated using ten different indices, including R^2 , RMSE, and MAE, for both lateral (x) and longitudinal (y) dimensions. The percentages used for model training and testing were, respectively, 70% and 30% of the main dataset. The predicted results were examined using the performance evaluation, rank analysis, the actual vs. predicted curve, the Error matrix, and the OBJ values. According to the experimental data, for each analysis, the XGB ($R^2=0.991$; 0.963 and 0.999; 0.997) was more accurate than the ETR and KNN models in training and testing of lateral and longitudinal dimensions. The "actual vs. predicted" curves further corroborated the superior performance of the XGB model, showing a strong alignment with the diagonal line, indicating accurate predictions. In contrast, the ETR model exhibited significant deviations, suggesting less accurate predictions. The rank analysis further confirmed the XGB model as the top performer, with the highest overall scores in both training and testing phases, followed by the KNN model, while the ETR model ranked lowest. Sensitivity analysis revealed that stress was the most influential parameter in predicting strain, followed by angle and height. Error matrix evaluations highlighted the XGB model's minimal errors compared to the other models, reinforcing its robustness and reliability. The OBJ criteria analysis also identified the XGB model as the closest to the ideal solution, particularly for the longitudinal (y) dimension dataset. Overall, the XGB model emerged as the most robust and reliable for predicting strain in both lateral and longitudinal directions, making it a valuable tool for practical applications in fields requiring precise strain predictions. The study underscores the importance of hyperparameter optimization and thorough model evaluation to achieve the best predictive performance. Further investigations are required to assess its performance under different loading conditions and with diverse rock types. The study's next directions might involve a full evaluation of the proposed ETR, XGB, and KNN models as well as hybrid models that combine deep learning, additional optimization methods, and neural networks.

Acknowledgements The authors are grateful to the Indian government's Department of Science and Technology for financing the project titled "Radical Decrease in natural catastrophic disasters". Ref no. ST/IMRCD/BRICS/Pilotcall2/RDRNCD/2018 (G) dated March 2019.

Data Availability The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflicts of interest The authors confirm that this research work content has no conflict of interest.

References

- Abdi Y, Momeni E, Armaghani DJ (2023a) Elastic modulus estimation of weak rock samples using random forest technique. *Bull Eng Geol Environ* 82:1–20
- Abdi Y, Momeni E, Armaghani DJ (2023b) Elastic modulus estimation of weak rock samples using random forest technique. *Bull Eng Geol Environ* 82:1–20. <https://doi.org/10.1007/s10064-023-03154-y>
- Ahmad MW, Reynolds J, Rezguy Y (2018) Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J Clean Prod* 203:810–821. <https://doi.org/10.1016/j.jclepro.2018.08.207>
- Al-Jeznawi D, Sadik L, Al-Janabi MAQ et al (2023) Developing Vs-NSPT Prediction Models Using Bayesian Framework. *Transp Infrastruct Geotechnol* 1–22. <https://doi.org/10.1007/s40515-023-00353-8>
- Altman NS (1992) An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am Stat* 46:175–185. <https://doi.org/10.1080/00031305.1992.10475879>
- Armaghani DJ, Momeni E, Asteris PG (2020) Application of group method of data handling technique in assessing deformation of rock mass. *Metaheuristic Comput Appl* 1:1–18
- Asteris PG, Mamou A, Hajihassani M et al (2021) Soft computing based closed form equations correlating L and N-type Schmidt hammer rebound numbers of rocks. *Transp Geotech* 29:100588. <https://doi.org/10.1016/j.trgeo.2021.100588>
- Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. Springer, Netherlands
- Bhatia N (2010) Vandana. *Survey of Nearest Neighbor Techniques* 8:302–305
- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geosci Model Dev* 7:1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. pp 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cioffi R, Travagliani M, Piscitelli G et al (2020) Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability* 12:492
- Cui L, Sheng Q, Zheng J, jie, et al (2019) Regression model for predicting tunnel strain in strain-softening rock mass for underground openings. *Int J Rock Mech Min Sci* 119:81–97. <https://doi.org/10.1016/j.ijrmm.2019.04.014>
- Dietterich TG (2000) Ensemble methods in machine learning. In: *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings 1*. Springer, pp 1–15
- Dowlatshahi MB, Hashemi A, Samaei M, Momeni E (2023) Feasibility of Artificial Intelligence Techniques in Rock Characterization. In: *Artificial Intelligence in Mechatronics and Civil Engineering: Bridging the Gap*. Springer, pp 93–110
- Fix E, Hodges Jr JL (1952) Discriminatory analysis-nonparametric discrimination: Small sample performance. *California Univ Berkeley Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn* 63:3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- C.S.Gundewar (2014) Government of India Ministry of Mines INDIAN BUREAU OF MINES Controller General Indian Bureau of Mines Application of Rock Mechanics in Surface and Underground Mining. Indian Bur Mines, Indira Bhavan, Civ Lines
- Gandomi AH, Yun GJ, Alavi AH (2013) An evolutionary approach for modeling of shear strength of RC deep beams. *Mater Struct* 46:2109–2119. <https://doi.org/10.1617/s11527-013-0039-z>
- Indraratna B, Armaghani DJ, Correia AG et al (2023) Prediction of resilient modulus of ballast under cyclic loading using machine learning techniques. *Transp Geotech* 38:100895

- Isah BW, Mohamad H, Ahmad NR, et al (2020) Uniaxial compression test of rocks: Review of strain measuring instruments. *IOP Conf Ser Earth Environ Sci* 476: <https://doi.org/10.1088/1755-1315/476/1/012039>
- Jahed Armaghani D, Kumar D, Samui P et al (2021) A novel approach for forecasting of ground vibrations resulting from blasting: modified particle swarm optimization coupled extreme learning machine. *Eng Comput* 37:3221–3235
- Jitchaijaroen W, Keawsawasvong S, Wipulanusat W, et al (2024) Machine learning approaches for stability prediction of rectangular tunnels in natural clays based on MLP and RBF neural networks. *Intell Syst with Appl* 200329. <https://doi.org/10.1016/j.iswa.2024.200329>
- Kikumoto M, Togashi Y (2022) Method for Measuring Three-Dimensional Strain Tensor of Rock Specimen Using Strain Gauges. *Rock Mech Rock Eng* 55:4093–4107. <https://doi.org/10.1007/s00603-022-02849-0>
- Kong X, Lu H, Liu C, Zhao B (2023) Experimental study on precursor characteristics of rock failure based on strain and temperature changes. *Case Stud Therm Eng* 41:102632. <https://doi.org/10.1016/j.csite.2022.102632>
- Koopialipoor M, Asteris PG, Mohammed AS et al (2022) Introducing stacking machine learning approaches for the prediction of rock deformation. *Transp Geotech* 34:100756
- Kumar M, Samui P (2019) Reliability Analysis of Pile Foundation Using ELM and MARS. *Geotech Geol Eng* 37:3447–3457. <https://doi.org/10.1007/s10706-018-00777-x>
- Kumar DR, Samui P, Burman A (2022a) Determination of Best Criteria for Evaluation of Liquefaction Potential of Soil. *Transp Infrastruct Geotechnol* 1–20. <https://doi.org/10.1007/s40515-022-00268-w>
- Kumar DR, Samui P, Burman A (2022b) Prediction of Probability of Liquefaction Using Soft Computing Techniques. *J Inst Eng Ser A* 103:1195–1208. <https://doi.org/10.1007/s40030-022-00683-9>
- Kumar M, Biswas R, Kumar DR et al (2022c) Metaheuristic Models for the Prediction of Bearing Capacity of Pile Foundation 2:129–147
- Kumar DR, Samui P, Wipulanusat W et al (2023a) Soft-computing techniques for predicting seismic bearing capacity of strip footings in slopes. *Build* 13:1371. <https://doi.org/10.3390/buildings13061371>
- Kumar R, Kumar A, Ranjan Kumar D (2023b) Buckling response of CNT based hybrid FG plates using finite element method and machine learning method. *Compos Struct* 319:117204. <https://doi.org/10.1016/j.compstruct.2023.117204>
- Kumar R, Rai B, Samui P (2023c) A comparative study of prediction of compressive strength of ultra-high performance concrete using soft computing technique. *Struct Concr* 1–18. <https://doi.org/10.1002/suco.202200850>
- Kunapuli G (2023) Ensemble Methods for Machine Learning. Simon and Schuster
- Laghaei M, Baghbanan A, Hashemolhosseini H, Dehghanipoodeh M (2018) Numerical determination of deformability and strength of 3D fractured rock mass. *Int J Rock Mech Min Sci* 110:246–256. <https://doi.org/10.1016/j.ijrmmms.2018.07.015>
- Li N, Wang X, Qiao R et al (2020) A prediction model of permanent strain of unbound gravel materials based on performance of single-size gravels under repeated loads. *Constr Build Mater* 246:118492. <https://doi.org/10.1016/j.conbuildmat.2020.118492>
- Liu Z, Shao J, Xu W, Meng Y (2013) Prediction of rock burst classification using the technique of cloud models with attribution weight. *Nat Hazards* 68:549–568. <https://doi.org/10.1007/s11069-013-0635-9>
- Liu Y, Zhao T, Ju W, Shi S (2017) Materials discovery and design using machine learning. *J Mater* 3:159–177
- Ma L, Zhou C, Lee D, Zhang J (2022) Prediction of axial compressive capacity of CFRP-confined concrete-filled steel tubular short columns based on XGBoost algorithm. *Eng Struct* 260. <https://doi.org/10.1016/j.engstruct.2022.114239>
- Medawela S, Armaghani DJ, Indraratna B et al (2023) Development of an advanced machine learning model to predict the pH of groundwater in permeable reactive barriers (PRBs) located in acidic terrain. *Comput Geotech* 161:105557
- Mohamad ET, Jahed Armaghani D, Momeni E, Abad ANK, SV, (2015) Prediction of the unconfined compressive strength of soft rocks: a PSO-based ANN approach. *Bull Eng Geol Environ* 74:745–757. <https://doi.org/10.1007/s10064-014-0638-0>
- Mohamad ET, Armaghani DJ, Momeni E et al (2018) Rock strength estimation: a PSO-based BP approach. *Neural Comput Appl* 30:1635–1646. <https://doi.org/10.1007/s00521-016-2728-3>
- Momeni E, Jahed Armaghani D, Hajihassani M, Mohd Amin MF (2015) Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. *Meas J Int Meas Confed* 60:50–63. <https://doi.org/10.1016/j.measurement.2014.09.075>
- Nazir R, Momeni E, Armaghani DJ, Amin MFM (2013a) Correlation between unconfined compressive strength and indirect tensile strength of limestone rock samples. *Electron J Geotech Eng* 18 I:1737–1746
- Nazir R, Momeni E, Armaghani DJ, Amin MFM (2013b) Prediction of unconfined compressive strength of limestone rock samples using I-type schmidt hammer. *Electron J Geotech Eng* 18 I:1767–1775
- Petr W, Lubomir S, Jan N et al (2016) Determination of stress state in rock mass using strain gauge probes CCBO. *Procedia Eng* 149:544–552
- Roh S-B, Kim YS, Ahn T-C (2020) Lazy Learning for Nonparametric Locally Weighted Regression. *Int J Fuzzy Log Intell Syst* 20:145–155. <https://doi.org/10.5391/IJFIS.2020.20.2.145>
- Sadik L, Al-Jeznawi D, Alzabeebee S, et al (2024) An Explicit Model for Soil Resilient Modulus Incorporating Freezing–Thawing Cycles Through Offspring Selection Genetic Algorithm (OSGA). *Transp Infrastruct Geotechnol* 1–16. <https://doi.org/10.1007/s40515-024-00399-2>
- Sadik L (2023) Developing Prediction Equations for Soil Resilient Modulus Using Evolutionary Machine Learning. *Transp Infrastruct Geotechnol* 1–23. <https://doi.org/10.1007/s40515-023-00342-x>
- Sagi O, Rokach L (2018) Ensemble learning: A survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 8:e1249
- Singh D, Singh B (2020) Investigating the impact of data normalization on classification performance. *Appl Soft Comput* 97:105524. <https://doi.org/10.1016/j.asoc.2019.105524>
- Stevens KN, Cover TM, Hart PE (2011) Nearest Neighbor. SpringerReference I: https://doi.org/10.1007/springerreference_62518
- Su J, Wang Y, Niu X et al (2022) Prediction of ground surface settlement by shield tunneling using XGBoost and Bayesian Optimization. *Eng Appl Artif Intell* 114:105020. <https://doi.org/10.1016/j.engappai.2022.105020>
- Sun D, Lonbani M, Askarian B et al (2020) Investigating the applications of machine learning techniques to predict the rock brittleness index. *Appl Sci* 10:1–17. <https://doi.org/10.3390/app10051691>
- Tariq Z, Elkhatatny S, Mahmoud M et al (2017) A new technique to develop rock strength correlation using artificial intelligence tools. *Soc Pet Eng - SPE Reserv Characterisation Simul Conf Exhib RCSC 2017*:1340–1353. <https://doi.org/10.2118/186062-ms>
- Tran DT, Onjaipurn T, Kumar DR, et al (2024) An eXtreme Gradient Boosting prediction of uplift capacity factors for 3D rectangular anchors in natural clays. *Earth Sci Informatics* 1–15. <https://doi.org/10.1007/s12145-024-01269-8>
- Vergara MR, Arismendy A, Liberos A, Brzovic A (2020) Numerical investigation into strength and deformability of veined rock mass. *Int J Rock Mech Min Sci* 135. <https://doi.org/10.1016/j.ijrmmms.2020.104510>
- Xu S, Wang S, Zhang P et al (2020) Study on strain characterization and failure location of rock fracture process using distributed optical fiber under uniaxial compression. *Sensors* 20:3853
- Yu C, Koopialipoor M, Murlidhar BR et al (2021) Optimal ELM–Harris Hawks Optimization and ELM–Grasshopper Optimization Models to Forecast Peak Particle Velocity Resulting from Mine Blasting. *Nat Resour Res* 30:2647–2662. <https://doi.org/10.1007/S11053-021-09826-4/FIGURES/9>