Geomechanics and Engineering, *Vol. 40, No. 2 (2025) 123-137*
https://doi.org/10.12989/gae.2025.40.2.123

Forecasting mechanical properties of soilcrete enhanced with metakaolin employing diverse machine learning algorithms

Ala'a R. Al-Shamasneh*[1], Arsalan Mahmoodzadeh[2],
Nejib Ghazouani[3] and Mohamed Hechmi El Ouni[4,5]

[1]Department of Computer Science, College of Computer & Information Sciences, Prince Sultan University, Rafha Street, Riyadh 11586, Saudi Arabia
[2]IRO, Civil Engineering Department, University of Halabja, Halabja, 46018, Iraq
[3]Department of Civil Engineering, College of Engineering, Northern Border University, Arar 73222, Saudi Arabia
[4]Department of Civil Engineering, College of Engineering, King Khalid University, PO Box 394, Abha 61411, Kingdom of Saudi Arabia
[5]Center for Engineering and Technology Innovations, King Khalid University, Abha 61421, Saudi Arabia

(*Received October 22, 2024, Revised January 17, 2025, Accepted January 20, 2025*)

Abstract. Soil in combination with cement, more commonly referred to as soilcrete, has gained great popularity within the construction sector. To this end, its mechanical properties have to be determined quickly and accurately. Unfortunately, the conventional methods for determining them include lab tests which are rather expensive and error prone. A better solution however comes in the form of machine learning (ML) algorithms which have tremendous potential. Hence, this study sought to analyze how efficient the three algorithms in predicting the uniaxial compressive strength (UCS) of soilcrete. 400 samples of soilcrete were manufactured and analyzed, using two types of soil including clay and limestone along with metakaolin which served as a mineral additive. A total of 80% of the dataset was made use of for training while the remaining 20% served a testing purpose, in addition to the 37 datasets which were specifically designed for evaluation purposes. A Stepwise procedure was completed and a total of 8 parameters were identified including metakaolin and soil type, super plasticizer content, water to binder ratio, shrinkage, binder density and finally ultrasonic velocity. Most of the algorithms were able to achieve satisfactory results however Gaussian process regression (GPR), support vector regression (SVR) and null-space SVR (NuSVR) were able to stand out due to their potential performance. Focusing on the trained models and lab tests that were done, this research managed to establish the proper superplasticizer constitution (1%), water-to-binder ratio (0.4) and metakaolin content (12%) with the goal of achieving the highest UCS value in the provided soilcrete specimens. Furthermore, a graphical user interface (GUI) was created based on the trained ML models. For the civil engineers and researchers who need to estimate the UCS of the soilcrete specimens, this GUI greatly simplifies the process.

Keywords: graphical user interface; laboratory test, machine learning; soilcrete; uniaxial compressive strength

1. Introduction

Soilcrete, a composite material made of cement and soil, is widely used today, most especially in geotechnical engineering and construction work, which is why it has garnered huge interest in the past few years. When using or constructing a structure with soilcrete, its mechanical properties such as compressive strength, modulus of elasticity, and shear strength, are greatly needed in the evaluation and designing of the structures load bearing capacity as well as stability. Determining these mechanically properties with accurate measurements of different structures ensures safety and construction longevity. Sadly, standard testing methods that laboratories utilize involve too many cost resources, time, and sample preparation. Likewise the materials to be tested, soil and cement, display high variability within their characteristics, meaning there could potentially be test result discrepancies that could lead to building and designing errors.

The obstacles faced by geotechnical engineering are now being solved fortuitously due to the emergence of machine learning (ML) techniques (Abdelmawla *et al.* 2023, Lawal *et al.* 2023, Mahmoodzadeh *et al.* 2022a, b, c). ML algorithms allow data mining and uncover the relationship between the set of input variables and the desired outcome (Gupta *et al.* 2021). From varied input parameters that include soil composition, cement content, curing duration, and moisture content, these algorithms are able to retrieve the forecasted mechanical characteristics of soilcrete by being trained on large data sets of vetted soilcrete samples (Asteris *et al.* 2017).

Employing ML techniques in assessing the soilcrete strengthening should have considerable benefits for several reasons. To begin with, the costs and time associated with laboratory testing are considerably reduced when utilizing ML algorithms (Naugler and Church 2019). Engineers whom require complex estimating models for soilcrete's mechanical properties, are able to forgo procedures which are long and costly through the use of efficient estimators.

*Corresponding author, Assistant Professor
E-mail: ashamasneh@psu.edu.sa

http://www.techno-press.org/?journal=gae&subpage=7
ISSN: 2005-307X (Print), 2092-6219 (Online)

This so, enables the prompt decision making within the construction phases enabling both cost savings and the improvement in timelines (Kakasor Ismael Jaf *et al.* 2023). Second, ML methodologies work out well in this case as they enable one to account for change in the soil and cement characteristics by deploying multiple input parameters. When many elements are analyzed at once, ML models establish complex interactions between them, which explains why mechanical properties can be forecasted more accurately. This adds to the reliability and forecast of soilcrete behaviour which gives engineers more assurance in designing structures. Finally, ML methods offer more capabilities in monitoring soilcrete at anytime and if so required, estimate soilcrete properties in an real time scenario. ML models can receive sensor data from construction sites or already erected buildings and can continuously estimate and improve the predictions available, reducing the need of maintenance and keeping the strength and safety of the structure intact. It has to be noted that the use and development of ML approaches in predicting soilcrete properties is still rather limited. In particular, Asteris *et al.* (2017) researched the use of artificial neural networks to predict the mechanical features of soilcrete materials from 134 data inputs and confirmed in a broad sense that the ANN method was quite precise. The progression of this field has been rather sluggish with the only significant proximal area to explore being the effectiveness of varying ML models in predicting the mechanical nature of soilcrete.

This article intends to examine the ability of ML methods to estimate the uniaxial compressive strength (UCS) of soilcrete materials in a professional and detailed manner. The ML techniques examined in this study encompass a range of algorithms, including ANN, Gaussian process regression (GPR), support vector regression (SVR), decision tree regressor (DTR), extra tree regressor (ETR), gradient boosting regressor (GBR), hist GBR (HGBR), null space SVR (NuSVR), extreme gradient boosting (XGBoost), random forest (RF), voting regressor (VR), and multilayer perceptron regression (MLPR). Because of their capacity to grasp intricate relationships and patterns in data, these techniques are largely used in predictive modeling. A major goal is to conduct an analysis for these techniques and determine the best possible way for estimating UCS of soilcrete. For this purpose, several soilcrete samples were prepared in a way that would enable them to be tested for UCS thereby producing a large amount of data. The two soil materials employed in the making of soilcrete specimens are Natural Clay Ground Soil (CGS), and Crushed Quarry Limestone Sand (CQLS) with metakaolin being the mineral additive. Also, unseen datasets are used to estimate the performance of ML models in generalization. The parameters which have significant influence on the UCS of soilcrete are determined by Stepwise method. A graphical user interface (GUI) is also implemented based on the trained ML models. This GUI can be an important asset especially for civil engineers and researchers since it will allow them to easily estimate the UCS of the soilcrete specimens.

The contributions of this study are as follows:

- In order to increase the accuracy of UCS testing, cooling mediums, two different types of soils and minerals components are employed this all results in a more diverse data set which results in a highly optimized ML model.
- Using the Stepwise method, the researcher attempts to clarify and define the different factors influencing UCS for soilcrete. Such information can aid and assist civil engineers and researchers within the field.
- For this analysis, an unseen dataset is used for estimating the generalization ability of the ML models. This analysis gives some information on the level of accuracy of the trained ML models to ascertain the UCS of soilcrete specimens that were not part of the training data set.
- Laboratory results and the trained ML models are used to determine the optimum values for factor such as metakaolin (MK) content as well as the water to binder (W/B) ratio used for the soilcrete cement.
- For simulating the ML models that has been trained the research created a GUI which can be used by civil engineers and also researchers. This interface has a greater ease of use for determining the UCS of soilcrete specimens resulting in ML technologies being more applicable in practice.

The whole process of this study is shown in Fig. 1.

2. Applied ML algorithms

The ML methods studied comprise twelve algorithms including SVR, NuSVR, GPR, ANN, MLPR, XGBoost, DTR, RF, GBR, HGBR, VR and ETR. A brief discussion of such techniques and their advantages over other ML techniques follows this.

GPR is a non parametric Bayesian approach, which describes a method for implementing a probabilistic model of the underlying relations between input data and output data targets. This approach handles all datasets of all sizes efficiently (Rasmussen 2004). SVR, as it is commonly known is a supervised learning concept which uses support vector machines to solve regression problems. It is well suited in the performance of tasks that involve high dimensional data (Cortes and Vapnik 1995). The method NuSVR on the other hand expands on the concepts of SVR and uses an altered loss function to robustly address outlier and noise control problems and overcome non linearity (Prasad and Jaganathan 2019).

DTR classifies data into branches and its leaf nodes determine which class will be selected based on majority voting. It is a clear model and also works with non-linear models quite well (Quinlan 1986). ANN is a simulation-based algorithm that works as a collection of nodes or interconnected neurons. This type of algorithm performs well with intricate relationships and has the ability to learn from large datasets and infer from them without any issue (Gad 2018). ETR ia an ensemble based algorithm that combines several decision trees and then takes an average

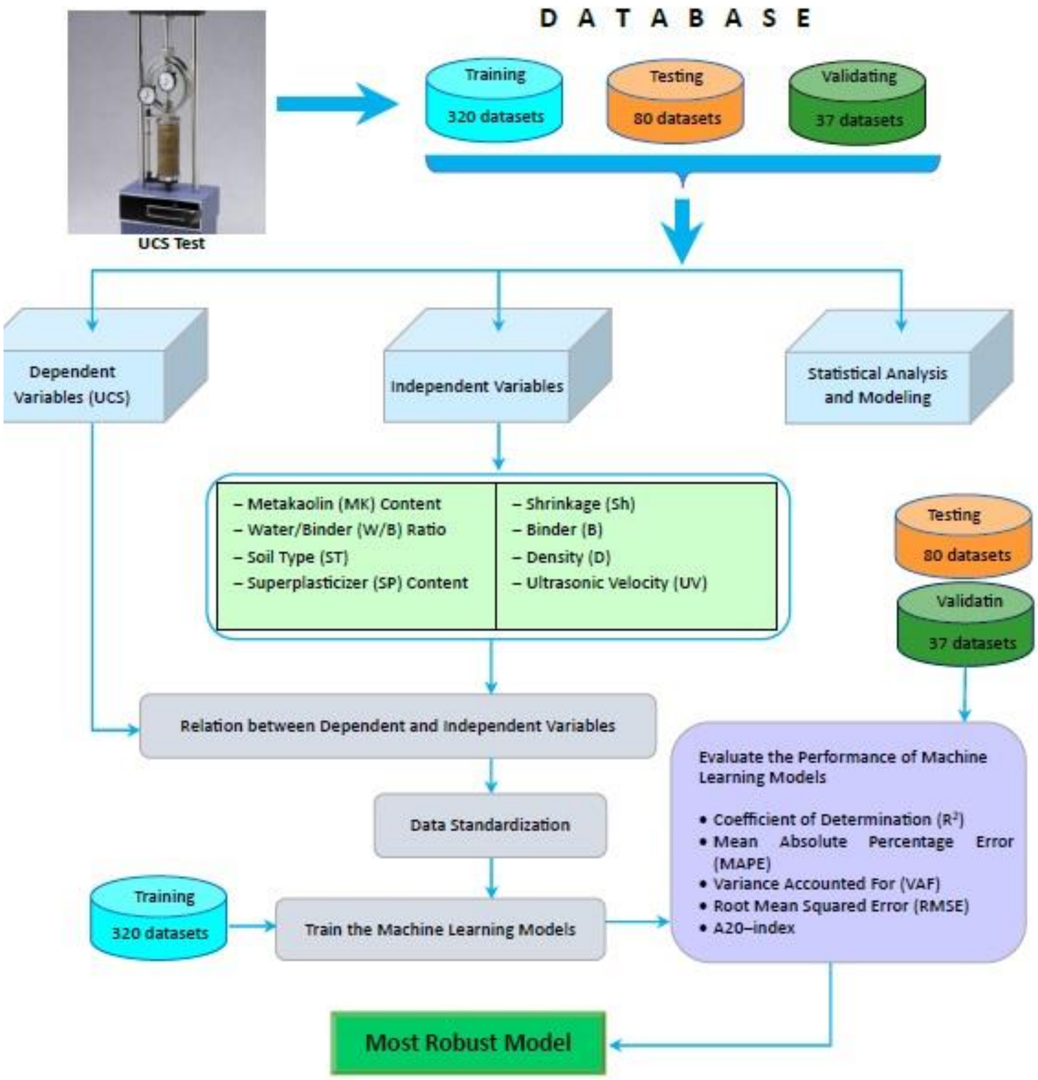

Fig. 1 Flowchart of the current study

or a majority of these trees to use the decision. It decreases the chances of overfitting and makes predictions with a high degree of accuracy that are robust (Chen and Guestrin 2016). GBR is becoming known for focusing on producing high accuracy, prediction, by using multiple weak learners, in this case decision trees, in an ensemble sequentially (Chen and Guestrin 2016, Jin and Agrawal 2003). HGBR however aims to improve the continuous to discrete feature transformations by using a histogrambased algorithm with the goal of enhancing how fast the models can be trained and the accuracy at which it predicts (Gayathri *et al*. 2022).

The XGBoost algorithm combines various models to create a predictive model that is accurate and fast that for large datasets (Chen and Guestrin 2016). RF is also ensemble learning based where several decision trees or average, and also uses majority vote for decision prediction. It antagonizes overfitting and makes dependable derisions (Ho 1998). MLPR normally consists of an input layer, hidden layers and an output layer. There are weighted connections between the neurons, and an activation function is used on the weighted sum of the inputs. The MLPR changes these weights during the training so that the difference between the predicted and actual values is smallest (Murtagh 1991). VR integrates forecast values from a number of regression models into one final forecast. This approach also entails ensemble boosting which integrates weight averages of various forecasts for improved results (Phyo *et al*. 2022).

We used the Jupyter Notebook environment as the working environment for the implementation of the ML models, particularly the 3.7 version available in Anaconda Navigator. Anaconda is a scientific open source software platform that is freely available and utilizes Python as its base. The underlying purpose however is to ease the use of packages in terms of management and execution. The operations were performed on a laptop with 32 Gigabytes of Random Access Memory and a 2.60 GHz Intel Core i7-10750H Computer Processor. All models were adjusted

Table 1 Optimized hyperparameters for the ML models

Model	Hyperparameter	Typical Range/Values
SVR	C (Regularization)	100
	epsilon (Tube size)	0.01
	kernel	rbf
NuSVR	nu	0.25
	C	100
	kernel	rbf
GPR	kernel	rbf
	alpha (Noise level)	1e-5
ANN	hidden_layer_sizes	(100,50)
	activation	relu
	learning_rate_init	0.01
MLPR	hidden_layer_sizes	(100,50)
	activation	relu
	learning_rate_init	0.01
XGBoost	n_estimators	500
	learning_rate	0.05
	max_depth	6
DTR	max_depth	10
	min_samples_split	5
RF	n_estimators	500
	max_depth	20
	max_features	sqrt
GBR	n_estimators	500
	learning_rate	0.05
	max_depth	6
HGBR	max_iter	500
	learning_rate	0.05
	max_leaf_nodes	63
VR	weights	distance
	n_neighbors	5
ETR	n_estimators	500
	max_depth	20

according to the working sets in order to obtain the best possible outcomes. The process of optimization was in this case the task of finding an adequate combination of values for the parameters. In this study, part of this optimization was performed with the use of trial and error method. The trial and error method is one of the most popular techniques used in exploring such difficult questions. The agent seeks all approaches until they complete their task or give up. The problem will be attempted multiple times with varying approaches until an appropriate solution is found. Table 1 shows the optimized hyperparameters for the ML models used in this study.

3. Data preparation and analysis

400 cylindrical soilcrete specimens were tested, with a height-to-diameter ratio of 2. These soilcrete samples comprised varying amounts of natural clay ground soil (CGS) and crushed quarry limestone sand (CQLS), which replaced the aggregate phase. A mineral additive called metakaolin (MK) was added to the ordinary Portland cement-based binder and water mixture at different water/binder (W/B) ratio values. The study focused on measuring the 28-day uniaxial compressive strength (UCS) of the soilcrete samples.

The study focused on the different categories of binders (B) that were made with varying percentages of CEM I 42.5 N Portland cement (PC) and MK. To ensure homogeneity within the binder categories, MK and PC were blended together with a laboratory swing mill for a duration of one hour. Soil materials (CGS and CQLS) were utilized to create batches of samples. The clay soilcrete was developed by combining CGS with the binder at varying W/B ratio values, while the sand based soilcrete was developed using CQLS with the same binder using different W/B ratio values.

Overall 400 datasets were created that included eight predictors which included ST soil type, W/B ratio, MK, superplasticizer (SP), shrinkage (Sh), density (D), and ultrasonic velocity (UV), and a single target variable (UCS). Of the entire dataset, 80% which consisted of 320 data points were used to train the model, and the other 20% were held back for model testing. The remaining 80 data points were also randomly selected and used as test data. Tables 2 and 3 show the general characteristics of the training and testing datasets, along with numerical and nominal parameters as well.

Feature engineering is an essential task in ML because it entails the choice, transformation, and fabrication of features from the raw data in order to improve the model's performance. The Stepwise approach which is commonly employed in ML and statistical models is adopted to select features in a given dataset with a minimum user interface. This process consists of stepping methods where a factor is incorporated into the model or removed from it, depending its significance in correlation orifiability. For this purpose, we have three criteria the Akaike Information Criterion (AIC) and a modified AIC known as AICc and BIC , etc. which assist in estimating the error of prediction of the model. The lower the value of these metrics, the better the fit of the model since the model inflicts more self punishment for adding parameters. AICc is sensitive to limited sample sizes, while BIC is even more punitive on the ever present problem of AIC adding extraneous variables. Our investigation reveals that with regard to both sample size and the number of variables, the intelligent model advocated in this thesis has the best AIC value.

To employ the best model, StepAIC from the MASS package is used and then three choices can be chosen from – "both" which is stepwise regression that utilizes merging selection with both forward and backward selection, and "backward" which is stepwise regression using only forward selection within the R softeare. According to data mining, our objective with the tools and mining is to make sure that the value for StepAIC is as minimum and as achievable as possible. The results shown in Table 4 give a rather significant result which is a 3 star p-value. The small p-value allows for the null hypothesis to be rejected which

Table 2 General specification of numerical datasets.

		W/B	MK % w/w in the dry mix	B % w/w in the dry mix	SP % w/w of the cementitious materials	Sh %	D kg/m³	UV m/s	UCS Mpa
Training data	count	320	320	320	320	320	320	320	320
	mean	0.83	5.54	38.66	1.11	8.90	1913.69	3224.22	38.60
	std	0.28	4.49	10.79	1.11	7.24	170.92	631.47	10.51
	min	0.30	0.00	20.00	0.00	0.00	1528.70	1635.78	17.49
	25%	0.70	2.00	30.00	0.00	3.00	1759.53	2983.33	30.59
	50%	0.90	5.00	35.00	1.00	7.50	1917.10	3376.67	36.27
	75%	1.05	10.00	50.00	2.00	14.13	2078.25	3585.39	46.84
	max	1.20	15.00	55.00	3.00	26.00	2186.20	4100.00	68.36
Testing data	count	80	80	80	80	80	80	80	80
	mean	0.52	7.50	37.50	1.50	11.41	2003.08	3410.74	47.63
	std	0.12	4.64	11.53	1.13	6.60	188.20	815.15	7.96
	min	0.30	0.00	20.00	0.00	0.00	1528.70	1635.78	33.17
	25%	0.45	3.75	28.75	0.75	6.00	1958.28	2780.54	41.68
	50%	0.53	7.50	37.50	1.50	11.00	2096.60	3831.67	46.91
	75%	0.61	11.25	46.25	2.25	16.00	2129.60	4004.17	53.54
	max	0.70	15.00	55.00	3.00	25.00	2186.20	4100.00	69.79

Table 3 General specification of datasets regarding the nominal parameters

Parameter		Count	Class	Vector	Frequency [%]
ST	Training data	320	CGS	[1 0]	35
			CQLS	[0 1]	65
	Test data	80	CGS	[1 0]	50
			CQLS	[0 1]	50

portrays that there is high possible relationship or correlation with the predictor and the target variables. The accepted p-value threshold is 0.05. In the stepwise method's first step, the use of all eight parameters yields in a p-value higher than 0.05 which as for example between 0.1 and 0.3 means all the parameters for UCS models are significant in the statistical sense.

A matrix illustrating three bonds between parameters is presented in the form of Fig. 2. This matrix reflects a low dependency between the input parameters and the output parameter (UCS). Consequently, no strong linear relationship exist between these parameters, which implies that linear methods are not effective in patterns recognition. It is quite clear that the connection between input and output is non-linear and complex and that is why advanced non-linear ML algorithms are needed to comprehend this relationship.

4. Data standardization

Within ML algorithms, data standardization plays a paramount role by making certain that all characteristics are in a common unit. This is especially true for the aforementioned algorithms that rely on distances, as standardization of data guarantees a level playing field for all features; otherwise, features with a greater scale may dominate and skew the results. Hence, standardizing the data removes this bias and allows the algorithm to be trained using all of the features equally.

StandardScaler method is one of the most commonly used methods for data standardization. It is the last stage in data pre-processing that modifies the data by calculation of the mean and dividing it by the standard deviation. This ensures that the distribution of each feature retains a mean of 0 and standard deviation of 1. The standardization is mathematically expressed as Eq. (1).

$$X_{scaled} = \frac{X-\mu}{\sigma} \tag{1}$$

where X is the original feature value, μ is the mean of the feature, and σ is the standard deviation.

In ML pipelines, the StandardScaler method is regularly used to avoid complications whereby ML models are trained on non standardized data.

Last but not least, it goes without saying that the non graphical input variables are normalized whereas the numerical input variables usually are standardized. As it was mentioned earlier, in the case of non numerical parameters, all the vector components corresponding to parameter information binary values composed of 0 and 1.

5. Performance evaluation of the ML models

The following evaluation criteria have been used to evaluate the ML models' performances to estimate the UCS of soilcrete in this study. These evaluation criteria are of utmost importance in assessing the performance of ML

Table 4 Feature engineering using the Stepwise method

	Estimate	Std. Error	t-value	p-value	Code
Intercept	7.777e+00	2.351e+00	3.307	0.00103	**
ST	1.831e+00	3.321e-01	5.514	6.38e-08	***
W/B	-2.436e+01	6.469e-01	-37.661	< 2e-16	***
MK	3.174e-01	3.330e-02	9.533	< 2e-16	***
SP	-3.933e+00	1.445e-01	-27.226	< 2e-16	***
Sh	-1.227e-01	2.300e-02	-5.334	1.63e-07	***
D	2.616e-02	1.164e-03	22.479	< 2e-16	***
UV	5.715e-04	2.760e-04	2.071	0.03904	*

Note: Significant codes: 0–0.001 (***); 0.001–0.01 (**); 0.01–0.05 (*); 0.05–0.1 (…); 0.1–1 (..); >1 (.)

models in estimating the UCS of soilcrete. Each criterion offers unique insights into the accuracy and efficacy of the models.

- Coefficient of determination (R^2): The models are evaluated by R^2 value which indicates the fraction of variance in the target variable in relation to what ML models would do. A high value of R^2 is indicative of higher correlation between the model and the data collected (Eq. (2)).
- Mean absolute percentage error (MAPE): MAPE determines the common difference in percentage average between the forecasted value and the logged of UCS. A low MAPE figure suggests that a soft computing or other model has an accurate estimation of the values (Eq. (3)).
- Root mean squared error (RMSE): RMSE represents the value of squared function of the mean of the differences between the forecasted values and the actual values. It gives the magnitude in form of a single figure of the total deviation between what the models have predicted and the actual values, the smaller the RMSE the better the performance of the model predictions (Eq. (4)).
- Variance accounted for (VAF): VAF refers to the ratio between the variance of the dependent variables in relation to the variance which has been estimated by means of the ML models. The larger the VAF the greater the role of the models to the variability in the values of the target parameter explains (Eq. (5)).
- a20−index: The model in this context is expected to be evaluated more accurately with this a20−index. In this way it judges the set of values which are predicted to be a certain range of plus or minus 20 percent of the target parameters. The higher the a20−index, the better the models will be able to generate precise estimates within this range (Eq. (6)).

$$\mathrm{R}^2 = \left(\frac{\sum_{i=1}^{n} (f(x_i) - \bar{f}(x))(f^*(x_i) - \overline{f^*}(x))}{\sqrt{\sum_{i=1}^{n} (f(x_i) - \bar{f}(x))^2 \sum_{i=1}^{n} (f^*(x_i) - \overline{f^*}(x))^2}} \right)^2 \tag{2}$$

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{f(x_i) - f^*(x_i)}{f(x_i)} \right| \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} (f(x_i) - f^*(x_i))^2} \tag{4}$$

$$\mathrm{VAF} = 1 - \left[\frac{var(f(x_i) - f^*(x_i))}{var(f(x_i))} \right] \times 100\% \tag{5}$$

$$a20 - inxex = \frac{n - no.of\ points\ between\ x = 1.20y\ and\ x = 0.80y}{n} \tag{6}$$

Where $f(x_i)$ and $f^*(x_i)$ are the measured and estimated values of parameter x for i^{th} dataset, respectively. n is the total number of test datasets.

To evaluate the ML algorithms' performance, each algorithm is assessed and awarded points based on predetermined criteria. Afterward, total attained points for each of the specified evaluation criteria are found out for each of the algorithms. In this situation, the algorithm that receives the highest total points is considered and recommended as the best and the most precise UCS estimator for soilcrete. Therefore, such a procedure makes it possible to conduct a multi-dimensional evaluation of the ML models' performance. This makes it possible to pinpoint the algorithm that is most precise and appropriate for estimating the UCS of soilcrete.

6. Results analysis and comparison

Utilizing graphs that employ the a20-index metric, we analyze the individually computed UCS values of the various algorithms in Fig. 3. As it can be seen, most of the points have their predicted output within the x=1.20y and x=0.80y lines, which indicates that most of the ML algorithms used in this study have an a20-index value greater than one. This means that all ML algorithms tested for the estimation of UCS for soilcrete in this study are reliable and valid.

Given that at least five algorithms have an a20-index score of one, it is unfeasible to differentiate between these algorithms with respect to this criterion on its own. Hence, other methods of evaluation have to be calculated as presented in Table 5. The score of the ML algorithm will be based on the achievement of the evaluation requirement. The computed values for each evaluation criteria show the methods' reliable accuracy, the sum of results for each algorithm is presented in the last column of Table 5. The GPR algorithm has a higher ranking score of 50, while the DTR algorithm has a lower ranking score of 8.

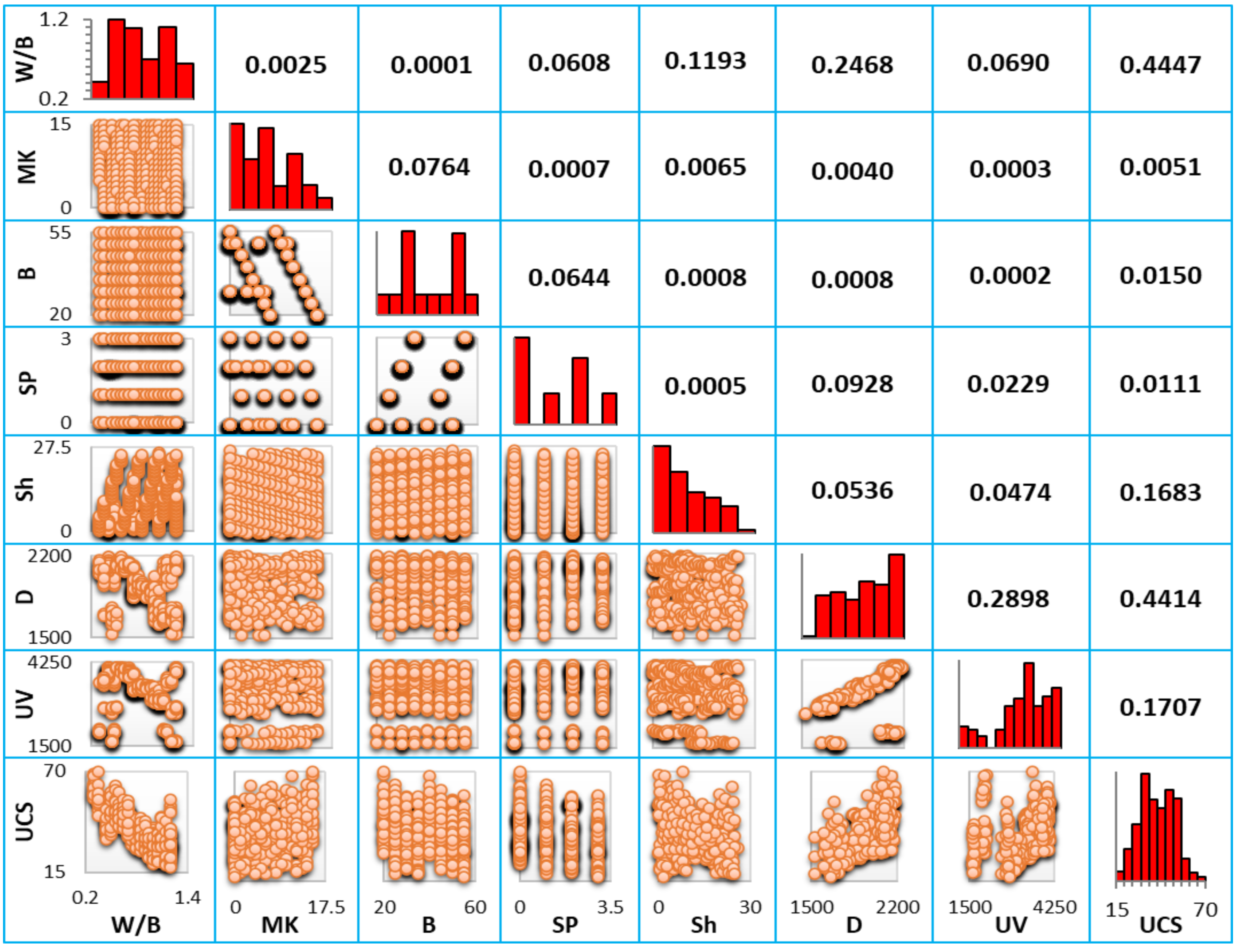

Fig. 2 Correlation matrix between the predictors and UCS

Table 5 Ranking score of the ML models based on the statistical metrics results

	R^2	Score	MAPE	Score	RMSE	Score	VAF [%]	Score	a20–index	Score	Ranking score
DTR	0.47	2	0.10	1	5.75	2	70.5	2	0.88	1	8
SVR	0.90	10	0.03	7	2.39	10	95.8	9	1.00	7	43
NuSVR	0.94	11	0.02	8	1.84	11	97.3	10	1.00	7	47
GPR	0.95	12	0.02	8	1.61	12	98.1	11	1.00	7	50
XGBoost	0.71	6	0.07	4	4.20	6	86.8	5	0.98	5	26
RF	0.49	3	0.09	2	5.63	3	70.2	1	0.90	2	11
ETR	0.63	4	0.07	4	4.80	4	81.6	4	0.94	4	20
GBR	0.66	5	0.08	3	4.57	5	81.6	4	0.94	4	21
HGBR	0.80	7	0.06	5	3.50	7	89.8	6	1.00	7	32
ANN	0.82	8	0.05	6	3.33	8	90.7	7	0.99	6	35
VR	0.39	1	0.09	2	6.15	1	78.40	3	0.91	3	10
MLPR	0.87	9	0.05	6	2.76	9	93.70	8	1.00	7	39

In this case, the GPR algorithm seems to outperform the other algorithms, and this is because these overall results provide an in-depth analysis of what is currently regarding Fig. 4. Algorithms seem to have been more crafted towards drawing final scores than caring about the final outcomes. These claims need further testing on additional datasets that have not been tested on.

Ultimately, this section assigns conclusions based on the previously shown graphs. However, these scores were derived from sample methods that were not restricted. As for the other algorithms at hand, without being tested on other algorithms it is difficult to portray an estimate of what this algorithm's performance might actually be when used solely as a GPR algorithm.

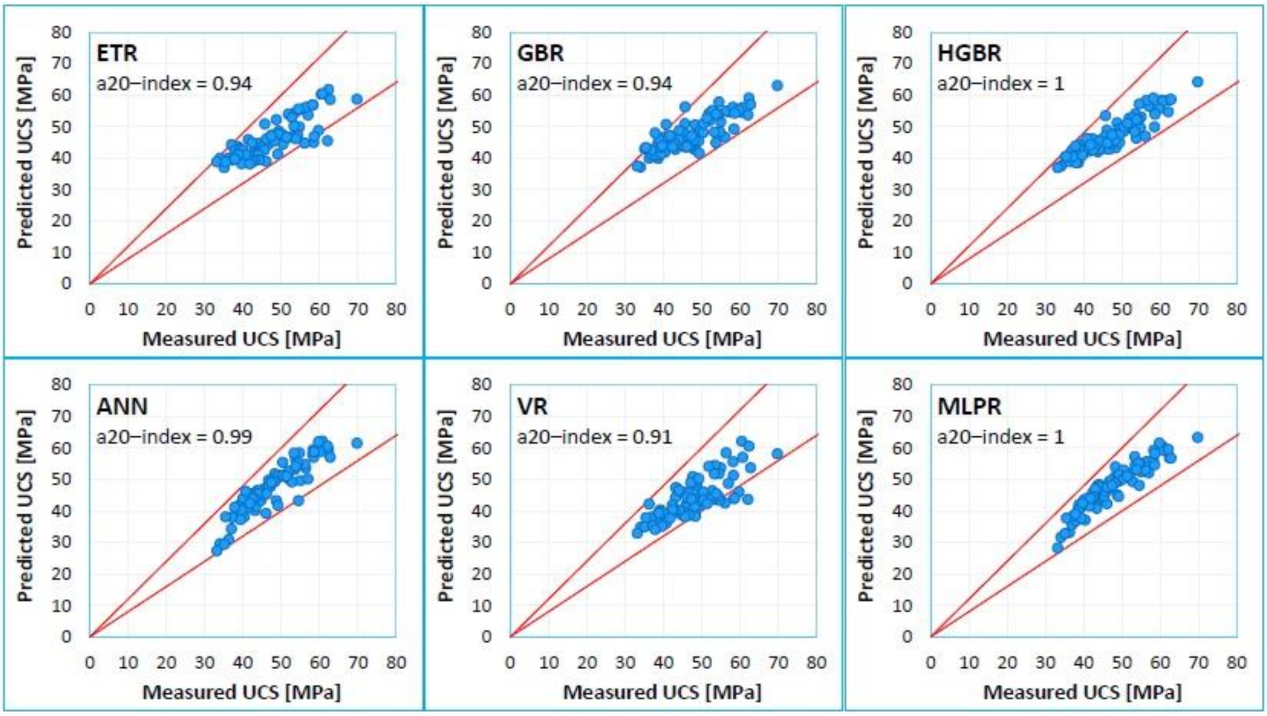

Fig. 3 Measuring the accuracy of the ML methods in estimating the UCS of soilcrte through the a20–index

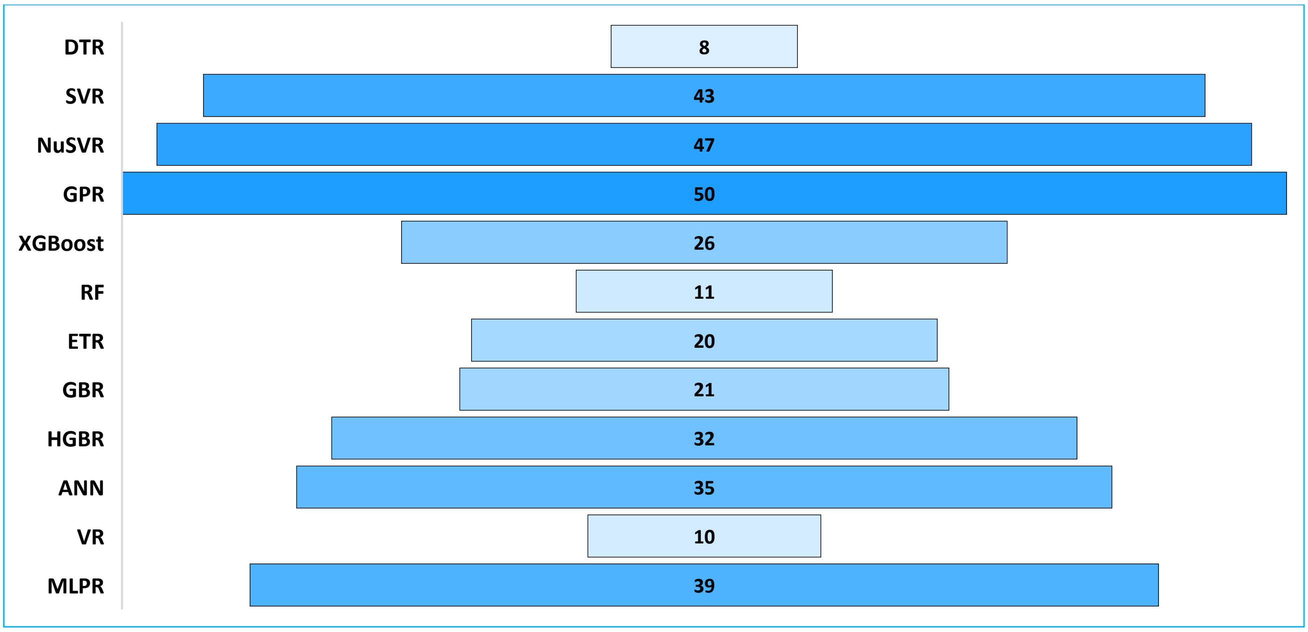

Fig. 4 Comparison between the ML models' performances through the ranking scores obtained for the statistical metrics

Next, we explore how effortless it is for trained algorithms to cope with new datasets. This time, we will try to change the behavior, or even the type, of one of the parameters in the system while leaving unchanged all the other input parameters. This is done with the aim of understanding how those algorithms predict the UCS of soilcrete with constant parameter but altered value. Busy algorithms must perform parameter tuning here as well. There is always the expectation that these algorithms are sensitive to parameter settings.

Initially, the MK parameter is varied from 1% to 15%, and other parameters are hold fixed. For lab testing, fifteen soilcrete specimens were created with the configurations noted in Table 6. The results obtained from these laboratory tests are available in Table 6. These 15 sets of data serve as the core basis for gauging the performance of the ML models. The output of each ML model for this new dataset is examined against the experimentally obtained data in Fig. 5. From the laboratory test results, it is noted that when MK content in soilcrete specimens is increased from 1% to 12%,

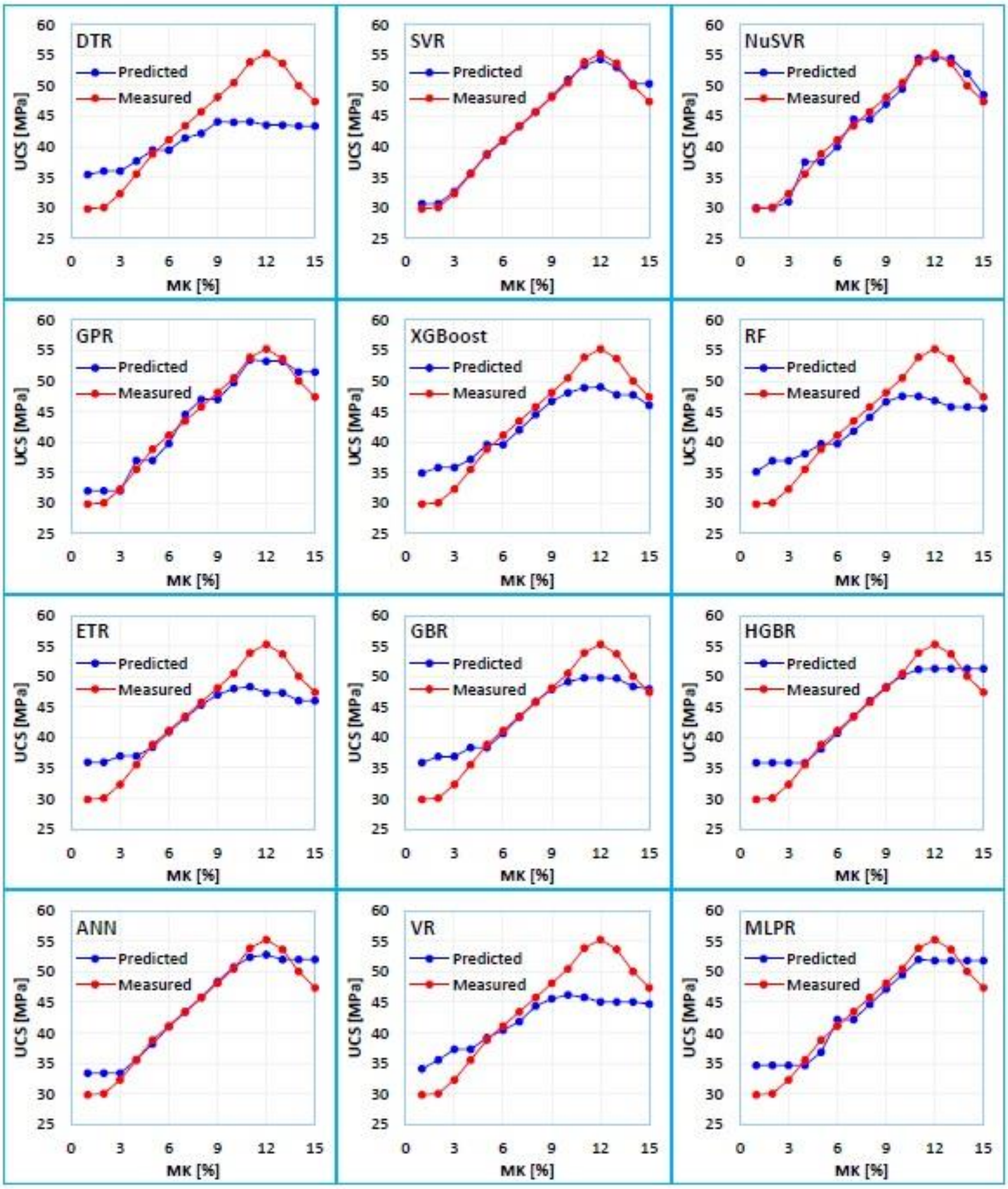

Fig. 5 Comparing the performance of the ML models for soilcrete's UCS estimation through changing the value of MK parameter and keeping other parameters constant

the UCS values tend to rise from 29.84 MPa to 55.28 MPa. When the MK content is increased from 12% to 15%, the UCS value, however, drops to 47.39 MPa. Thus, it can be inferred that the optimal MK content in soilcrete specimens for maximum UCS is 12%. In addition, the other ML models, SVR, NuSVR, and GPR, were in close correlation and agreement with the laboratory measurements.

The rest of ML models, however, still showed significant deviation from the laboratory results as the estimated UCS value remained erratic but showed an increase with the increase in MK content from 1% to 12%.

Along similar lines, the ANN and MLPR models were on par with the lab results.

It can, however, be summarized that, in accuracy and performance with new dataset, SVR, NuSVR, and GPR models achieved the best results of all the ML models. But among these three models, the most accurate UCS values of soilcrete were obtained using SVR model.

At this step, our target is to vary the W/B ratio between the preset limits of 0.3 and 1.2 without altering the other parameters. For laboratory testing purposes, 19 soilcrete specimens were carefully prepared, all of which were

Table 6 An evaluation dataset to check the performance of the trained ML models for soilcrete's UCS estimation by changing the value of the MK parameter and keeping other parameters constant

ST	W/B	MK [% w/w in the dry mix]	B [% w/w in the dry mix]	SP [% w/w of cementitious materials]	Sh [%]	D [kg/m^3]	UV [m/s]	UCS [MPa]
CGS	0.8	1	38	1.11	8.9	1910	3220	29.84
CGS	0.8	2	38	1.11	8.9	1910	3220	30.06
CGS	0.8	3	38	1.11	8.9	1910	3220	32.29
CGS	0.8	4	38	1.11	8.9	1910	3220	35.54
CGS	0.8	5	38	1.11	8.9	1910	3220	38.82
CGS	0.8	6	38	1.11	8.9	1910	3220	41.12
CGS	0.8	7	38	1.11	8.9	1910	3220	43.44
CGS	0.8	8	38	1.11	8.9	1910	3220	45.78
CGS	0.8	9	38	1.11	8.9	1910	3220	48.13
CGS	0.8	10	38	1.11	8.9	1910	3220	50.51
CGS	0.8	11	38	1.11	8.9	1910	3220	53.89
CGS	0.8	12	38	1.11	8.9	1910	3220	55.28
CGS	0.8	13	38	1.11	8.9	1910	3220	53.66
CGS	0.8	14	38	1.11	8.9	1910	3220	50.04
CGS	0.8	15	38	1.11	8.9	1910	3220	47.39

Table 7 An evaluation dataset to check the performance of the trained ML models for soilcrete's UCS estimation by changing the value of W/B ration parameter and keeping other parameters constant

ST	W/B	MK [% w/w in the dry mix]	B [% w/w in the dry mix]	SP [% w/w of cementitious materials]	Sh [%]	D [kg/m^3]	UV [m/s]	UCS [MPa]
CGS	0.3	5	38	1.11	8.9	1910	3220	51.88
CGS	0.35	5	38	1.11	8.9	1910	3220	54.46
CGS	0.4	5	38	1.11	8.9	1910	3220	58.30
CGS	0.45	5	38	1.11	8.9	1910	3220	57.45
CGS	0.5	5	38	1.11	8.9	1910	3220	53.51
CGS	0.55	5	38	1.11	8.9	1910	3220	51.28
CGS	0.6	5	38	1.11	8.9	1910	3220	49.43
CGS	0.65	5	38	1.11	8.9	1910	3220	47.66
CGS	0.7	5	38	1.11	8.9	1910	3220	44.72
CGS	0.75	5	38	1.11	8.9	1910	3220	42.33
CGS	0.8	5	38	1.11	8.9	1910	3220	40.60
CGS	0.85	5	38	1.11	8.9	1910	3220	38.14
CGS	0.9	5	38	1.11	8.9	1910	3220	35.90
CGS	0.95	5	38	1.11	8.9	1910	3220	33.53
CGS	1	5	38	1.11	8.9	1910	3220	29.64
CGS	1.05	5	38	1.11	8.9	1910	3220	25.96
CGS	1.1	5	38	1.11	8.9	1910	3220	23.19
CGS	1.15	5	38	1.11	8.9	1910	3220	21.06
CGS	1.2	5	38	1.11	8.9	1910	3220	19.70

exactly as specified in Table 7. The complete results of these experiments are captured in the respective table.

These 19 datasets formed the basis for testing the results of the ML models. The trained ML models were analyzed to predict the UCS of soilcrete on this dataset. The results improvement obtained from each model was thoroughly analyzed as snapshot undertaken for various models, which are presented in Fig. 6.

It was established at the laboratory that as the W/B ratio of soilcrete specimens was increased from 0.3 to 0.4, the UCS value increased from 51.88 MPa to 58.30 MPa. Further increases in W/B ratio from 0.4 to 1.2 resulted in

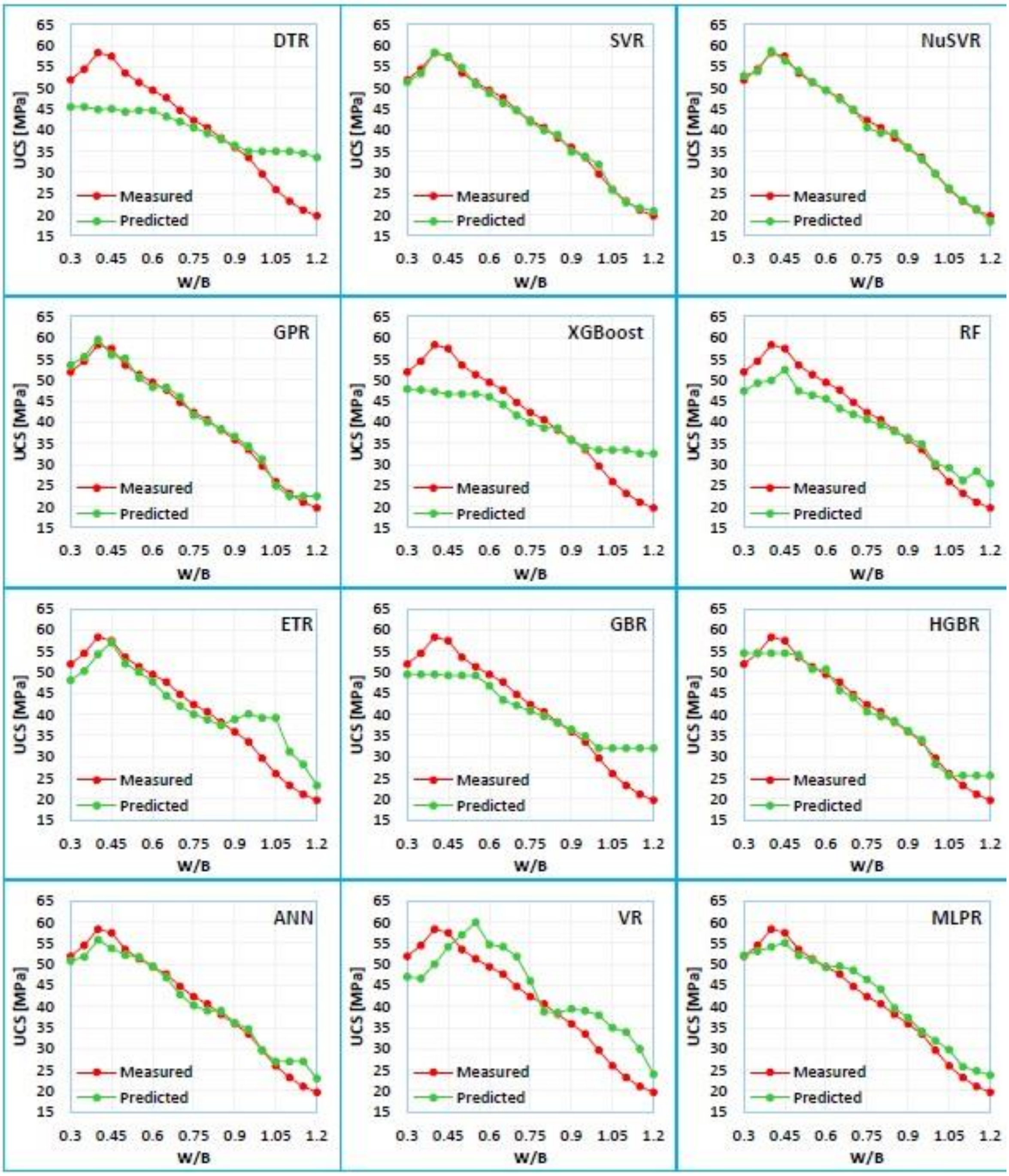

Fig. 6 Comparing the performance of the ML models for soilcrete's UCS estimation through changing the ratio of W/B and keeping other parameters constant

decrease of UCS value to 19.70 MPa. The results show that the optimal W/B ratio for getting the maximum UCS in soilcrete specimens is 0.4.

Following the analysis and presentation of the data in Fig. 6, it can be noted that some ML models like SVR, NuSVR, GPR, ANN, MLPR, RF, and HGBR have performed exceptionally well by demonstrating a high level of accuracy in this step. Their estimations coincided with the laboratory measurements, indicating their level of accuracy. These models have proven to be reliable tools in predicting the UCS values for the new dataset. It is, however, clear that the DTR, GBR, ETR, XGBoost, and VR models may require further refinement as they were noted for their blatantly inaccurate measurements when compared to the laboratory results. Their estimations and predictions were not close to the numbers, revealing some weakness or inconsistency in their application.

Again, three SVR, NuSVSR, and GPR models are the reference models, and that references yields higher accuracy regardless of what algorithm is chosen when applied on the new dataset, reinforcing their credibility and reliability.

Table 8 An evaluation dataset to check the performance of the trained ML models for soilcrete's UCS estimation by changing the value of SP parameter and keeping other parameters constant

ST	W/B	MK [% w/w in the dry mix]	B [% w/w in the dry mix]	SP [% w/w of cementitious materials]	Sh [%]	D [kg/m^3]	UV [m/s]	UCS [MPa]
CGS	0.8	5	38	0	8.9	1910	3220	36.83
CGS	0.8	5	38	1	8.9	1910	3220	42.34
CGS	0.8	5	38	2	8.9	1910	3220	38.2

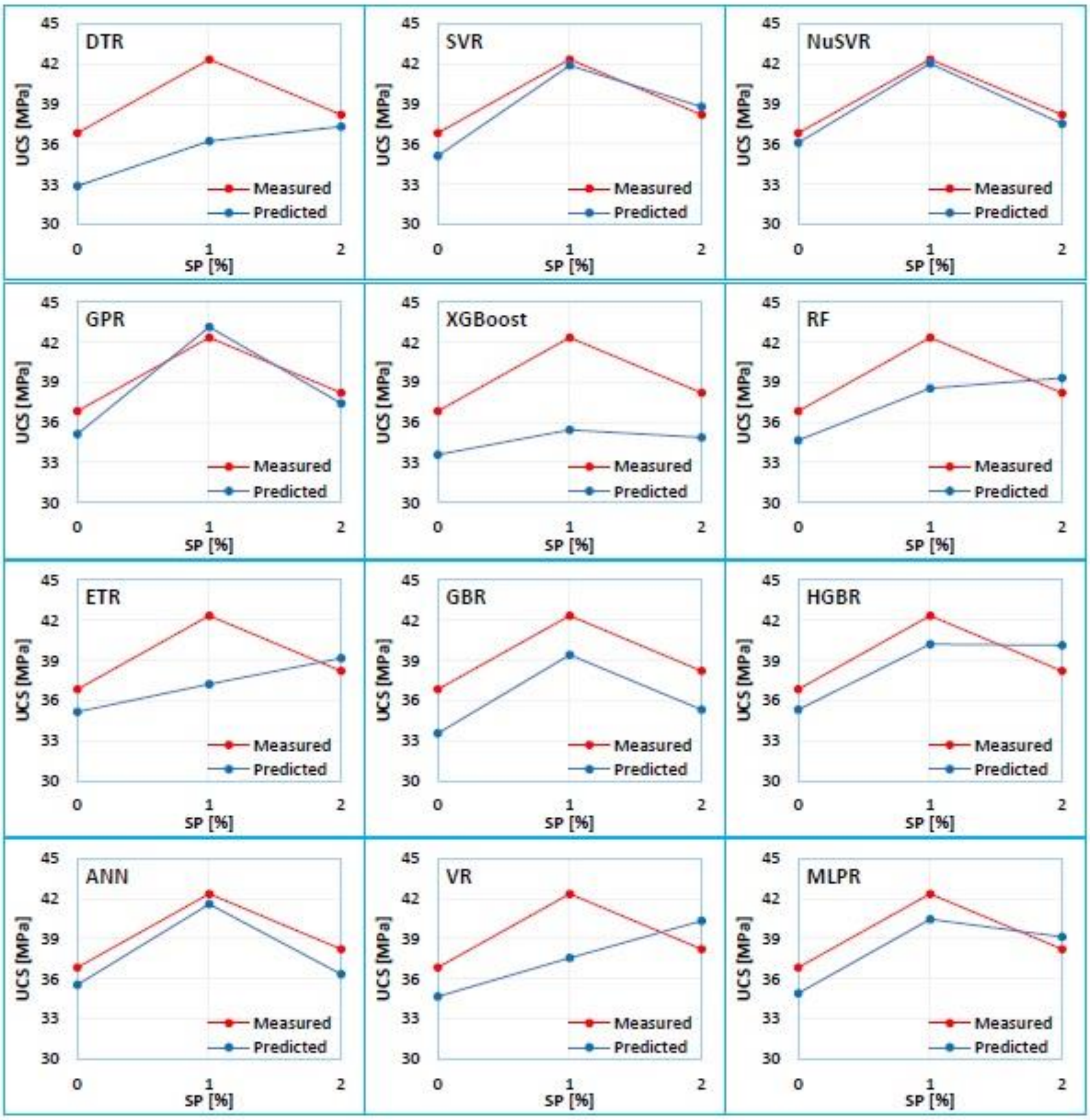

Fig. 7 Comparing the performance of the ML models for soilcrete's UCS estimation through changing the value of SP parameter and keeping other parameters constant

These models provided estimations that were closer to laboratory results, and proved to be some of the more trustworthy models concerning their application in the field of ML after building the model.

We are currently adjusting the SP parameter, and for this set of experiments, we maintain all other parameters constant while changing the SP parameter within the set range of between 0 to 2%. To allow for thorough laboratory testing, we manufactured 3 soilcrete specimens for the purpose of all of them meeting the requirements in Table 8. The details of these analyses can be found in the respective tables.

These three datasets serve as the basis for assessment of performance of the trained ML models. The trained ML models were used to test their capability to estimate the UCS of soilcrete on this particular set. The estimates given by each model were analyzed together with the experimental results, and the details were summarized in a more convenient shape in Fig. 7.

Reviewing the laboratory data, one pattern stood out. As the SP increases from zero to one percent, the UCS of the soilcrete specimens increases from 36.83 MPa to 42.34 MPa. As was expected, as the SP content at the soilcrete was increased further to 2%, UCS value dropped to 38.2 MPa. This clearly demonstrates that 1% SP is optimal internally in soilcrete specimens to achieve maximum UCS.

Peering Through the interpretation, SVR, NuSVR, GPR and ANN models were observed to be very accurate during this step, as shown in Fig. 7. Their estimations greatly correlated with the lab measurements, thus having great accuracy. These models have proven their competency in estimating the UCS values for the new dataset. But it begs to be said that the rest of the ML models were rather off in their estimation when compared with the laboratory measurements. Their estimates did not positively correlate with the actual values and hence, it can be deduced that there are some lacking factors in those models or their are certain issues within the model.

As in the previous evaluation results, models SVR, NuSVR, and GPR, outperformed as always boosting their accuracy and performance which bore results for the new dataset. It assists in proving them to be credible and reliable strong ML models. Such models are endorsed since they have proven time and again to provide estimates that are favorable to the lab results, thus earning them as dependable for this purpose.

Keeping in mind the datasets employed in this research, this comprehensive analysis of the results indicates the usage of SVR, NuSVR and GPR models for predicting UCS of soilcrete specimens.

There are several reasons why the SVR, NuSVR, and GPR models have demonstrated the most robust results compared to other ML models:

- SVR, NuSVR, and GPR models are based on well-established and robust algorithms extensively studied and refined in ML. These algorithms are designed to handle complex relationships and patterns in data, making them suitable for accurately predicting the UCS values in soilcrete specimens.
- These models capture non-linear relationships between input variables and the target variable. They can model intricate patterns and variations in the data, which is crucial for accurately estimating the UCS values.
- They are known for their robustness to outliers and noise in the data. They are less sensitive to individual data points that may deviate from the overall trend, allowing them to provide more reliable and consistent predictions.
- These models can effectively generalize from small samples, making them suitable for limited data availability.
- The SVR, NuSVR, and GPR models are based on solid theoretical principles. They have a strong mathematical foundation, allowing a deeper understanding of their behavior and performance. This theoretical grounding contributes to their robustness and reliability in predicting the UCS values.

7. Development of a Graphical user interface (GUI)

Without a doubt, learning ML algorithms and their code, particularly for a civil engineer not well versed with programming and algorithms can be a daunting task. Also, experimental tests to determine the mechanical properties of materials can require a lot of time and money. To help these issues, an adequate user GUI has been designed in this study, which is shown in Fig. 8. Thanks to this GUI, all that has to be done to estimate the UCS of soilcrete with any ML algorithm used in this study is plug and play. Just the set the needed parameter values or types and everything else will be done by the GUI. This software has proved to be of great importance during early steps of the engineering structure designs because estimating UCS is a difficult task. Moreover, people searching for the new datasets based on UCS of soilcrete would find this GUI very helpful. Because of the SVR, NuSVR, and GPR algorithms, a lot of data can be generated and used for future research with this GUI.

8. Key limitations and future work

There are several shortcomings within this research that should be noted, and there do exists several useful points which could be usefully researched in the future so as to work around these shortcomings:

- One of the main constraints of this study is the relatively small dataset size. An extensive dataset would bolster the strength and depends of ML models by encompassing an extensive variation of soilcrete characteristics translating to better generalization. Future work could focus on collecting a larger dataset to enhance the robustness and reliability of the models.
- The study focused on a relatively closed set of input parameters for predicting soilcrete UCS. To enhance the models, it is recommended that additional parameters or different combinations of parameters be included. This will improve the accuracy and applicability of the models.
- The study rested primarily on assessing the performance of twelve ML models which makes the work narrow. Future work may broaden the scope of the research for UCS prediction of soilcrete by employing wider range of ML models.

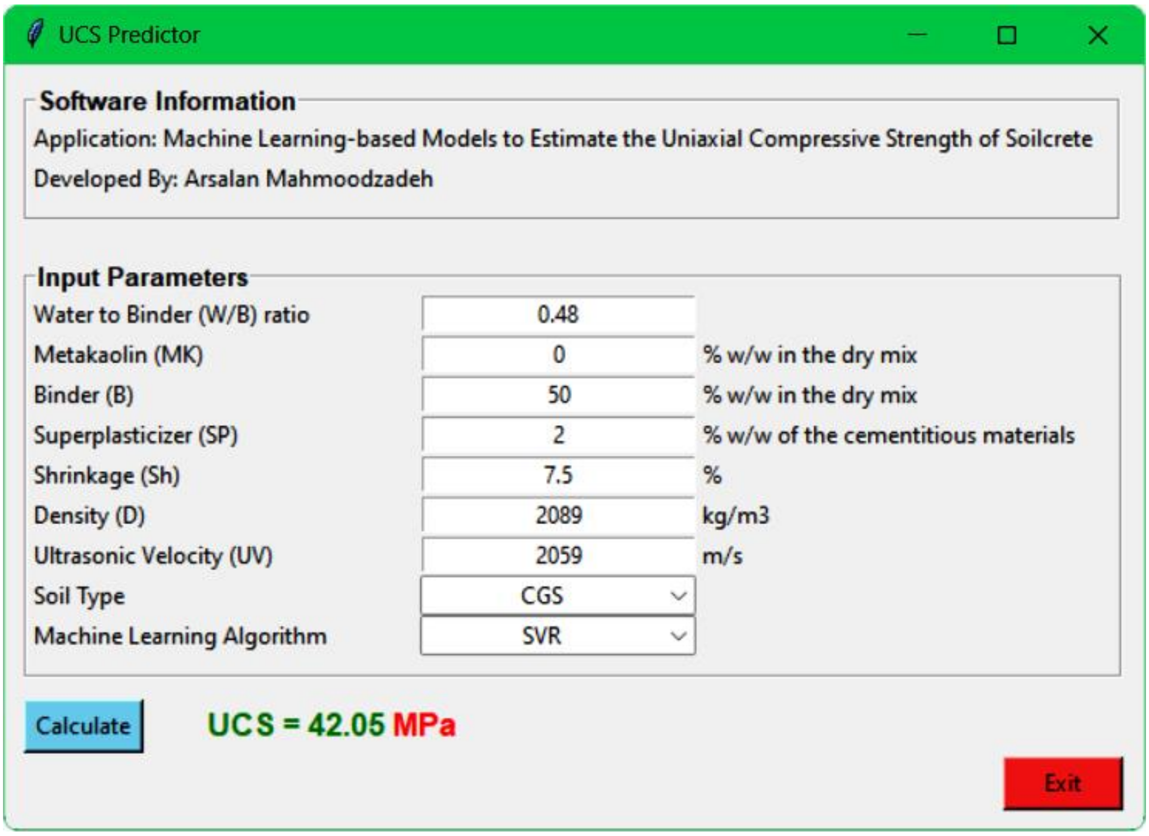

Fig. 8 A GUI developed in this study to estimate the UCS of soilcrete

- The work was devoid of uncertainty analysis of ML predictions of UCS which was a limitation of the study scope. Future work could alleviate this issue by implementing uncertainty estimation techniques such as bootstrapping or Bayesian methods.
- The study did not focus on the interpretability of the ML models. Future research could explore techniques that enhance the interpretability of the models, enabling engineers to gain insights into the underlying factors that influence UCS predictions.

By addressing these limitations and incorporating the suggested future work, the study can further advance the accuracy, reliability, and practicality of the ML models in predicting the UCS of soilcrete.

9. Conclusions

The main purpose of the study was to evaluate twelve ML algorithms predicting UCS of soilcrete. It was necessary to build a database of 400 data points for laboratory testing in order to accomplish this. Using the Stepwise method, eight parameters were found to have an effect on the UCS of soilcrete. These data points were divided in a way that 80% of data went to training algorithms and 20% was kept for testing. In total, 37 data points were used to evaluate the algorithms. Subsequently each algorithm underwent meticulous analysis after training on the test and evaluation datasets. Various evaluation criteria such as RMSE, VAF, R^2, MAPE, and a20–index were employed to compare the performance of each model to the measured results and to each other. Subsequently, a GUI software based on the trained ML algorithms was developed to facilitate the estimation of soilcrete's UCS.

Key findings from this study include the following:

- When assessing ML models, it is wise to not focus only on a singular evaluation criteria. This is because each criterion has a distinct approach in estimating algorithm accuracy. Therefore, results estimated by each algorithm must be evaluated using multiple criteria.
- The performance of each algorithm in estimating the UCS of soilcrete on the test dataset, ranked from highest to lowest accuracy, is as follows: GPR -> NuSVR -> SVR -> MLPR -> ANN -> HGBR -> XGBoost -> GBR -> ETR -> RF -> VR -> DTR.
- The algorithms' performance on unseen datasets considerably differed from their performance on the test datasets. SVR, NuSVR, and GPR algorithms consistently provided accurate and satisfactory performance across all stages, unlike other algorithms. The success of this study's SVR, NuSVR, and GPR models can be attributed to their ability to effectively solve problems with a limited dataset.
- Based on the findings from the laboratory experiments, the optimal MK content, W/B ratio, and SP content for achieving maximum UCS of soilcrete specimens were identified by about 12%, 0.4, and 1%, respectively. The SVR, NuSVR, and GPR models also obtained these results.

- It is essential not to rely solely on the performance of ML algorithms on training and test data, as they may exhibit high accuracy due to issues like overfitting. The capability of algorithms trained on new and unseen data should always be assessed to determine if they can respond to new data effectively.
- The capability of such algorithms to be presented as GUI programs can be very useful considering that civil engineering specialists and even some researchers may have difficulties with the ML algorithms and the programming languages related to them. Thus, a GUI was made for this research that integrated various trained ML algorithms, which allows predicting the UCS of soilcrete specimens.

In this research, the SVR, NuSVR, and GPR algorithms performed the best, so their use is recommended further in the developed GUI software for soilcrete's UCS estimation. Other than that, this software also permits the incorporation of information obtained from laboratory tests which increases the confidence in estimating the UCS of soilcrete.

Acknowledgments

The authors would like to thank Prince Sultan University for their support. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/176/45. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFMRA-2025-2105-01".

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.

Asteris, P.G., Alexandridis, A., Kolovos, K.G., Anesti, E.G., Douvika, M.G., Karamani, C., Kassolis, M.G. and Skentou, A. (2017), "Prediction of mechanical characteristics of soilcrete materials using artificial neural networks", *Proceedings of the 7th International Conference on Mechanics and Materials in Design*.

Chen, T. and Guestrin, C. (2016), "XGBoost", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785.

Cortes, C. and Vapnik, V. (1995), "Support-vector networks", *Machine Learning*, **20**(3), 273-297. https://doi.org/10.1007/BF00994018.

Gad, A.F. (2018), "Artificial neural networks", *Practical Computer Vision Applications Using Deep Learning with CNNs*, 45-106. Apress. https://doi.org/10.1007/978-1-4842-4167-7_2.

Gayathri, R., Rani, S.U., Čepová, L., Rajesh, M. and Kalita, K. (2022), "A comparative analysis of machine learning models in prediction of mortar compressive strength", *Processes*, **10**(7), 1387. https://doi.org/10.3390/pr10071387.

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R.K. and Kumar, P. (2021), "Artificial intelligence to deep learning: machine intelligence approach for drug discovery", *Molecular Diversity*, **25**(3), 1315-1360. https://doi.org/10.1007/s11030-021-10217-3.

Ho, T.K. (1998), "The random subspace method for constructing decision forests", *IEEE T. Pattern Anal.*, **20**(8), 832-844. https://doi.org/10.1109/34.709601

Jin, R. and Agrawal, G. (2003), "Communication and memory efficient parallel decision tree construction", *Proceedings of the 2003 SIAM International Conference on Data Mining*, https://doi.org/10.1137/1.9781611972733.11.

Kakasor Ismael Jaf, D., Ismael Abdulrahman, P., Salih Mohammed, A., Kurda, R., Qaidi, S.M.A. and Asteris, P.G. (2023), "Machine learning techniques and multi-scale models to evaluate the impact of silicon dioxide (SiO2) and calcium oxide (CaO) in fly ash on the compressive strength of green concrete", *Constr. Build. Mater.*, **400**, 132604. https://doi.org/10.1016/j.conbuildmat.2023.132604.

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.

Mahmoodzadeh, A., Mohammadi, M., Abdulhamid, S., Hama Ali, H., Ibrahim, H. and Rashidi, S. (2022a), "Forecasting tunnel path geology using Gaussian process regression", *Geomech. Eng.*, **28**(4), 359-374. https://doi.org/10.12989/gae.2022.28.4.359.

Mahmoodzadeh, A., Nejati, H.R., Mohammadi, M., Ibrahim, H., Mohammed, A. and Rashidi, S. (2022b), "Assessment of wall convergence for tunnels using machine learning techniques", *Geomech. Eng.*, **31**(3), 265-279. https://doi.org/10.12989/gae.2022.31.3.265.

Mahmoodzadeh, A., Nejati, H.R., Mohammadi, M., Ibrahim, H., Rashidi, S. and Mohammed, A. (2022c), "Meta-heuristic optimization algorithms for prediction of fly-rock in the blasting operation of open-pit mines", *Geomech. Eng.*, **30**(6), 489-502. https://doi.org/10.12989/gae.2022.30.6.489.

Murtagh, F. (1991), "Multilayer perceptrons for classification and regression", *Neurocomput.*, **2**(5-6), 183-197. https://doi.org/10.1016/0925-2312(91)90023-5.

Naugler, C. and Church, D.L. (2019), "Automation and artificial intelligence in the clinical laboratory", *Critical Reviews in Clinical Laboratory Sciences*, **56**(2), 98-110. https://doi.org/10.1080/10408363.2018.1561640.

Phyo, P.P., Byun, Y.C. and Park, N. (2022), "Short-term energy forecasting using machine-learning-based ensemble voting regression", *Symmetry*, **14**(1), 160. https://doi.org/10.3390/sym14010160.

Prasad, D.V.V. and Jaganathan, S. (2019), "Null-space based facial classifier using linear regression and discriminant analysis method", *Cluster Comput.*, **22**(4), 9397-9406. https://doi.org/10.1007/s10586-018-2178-z.

Quinlan, J.R. (1986), "Induction of decision trees", *Machine Learning*, **1**(1), 81-106. https://doi.org/10.1007/BF00116251.

Rasmussen, C.E. (2004), *Gaussian Processes in Machine Learning*, 63-71. https://doi.org/10.1007/978-3-540-28650-9_4.

IC