

# Machine Learning Methods to Predict and Analyse Unconfined Compressive Strength of Stabilised Soft Soil with Polypropylene Columns

Md. Ikramul Hoque, Muzamir Hasan, Md Shofiqul Islam, Moustafa Houda, Mirvat Abdallah & Md. Habibur Rahman Sobuz

To cite this article: Md. Ikramul Hoque, Muzamir Hasan, Md Shofiqul Islam, Moustafa Houda, Mirvat Abdallah & Md. Habibur Rahman Sobuz (2023) Machine Learning Methods to Predict and Analyse Unconfined Compressive Strength of Stabilised Soft Soil with Polypropylene Columns, Cogent Engineering, 10:1, 2220492, DOI: [10.1080/23311916.2023.2220492](https://doi.org/10.1080/23311916.2023.2220492)

To link to this article: <https://doi.org/10.1080/23311916.2023.2220492>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 14 Jun 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Received: 25 January 2023  
Accepted: 29 May 2023

\*Corresponding author: Md. Ikramul Hoque, Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, Gambang, Malaysia  
E-mail: [ikramul3300@becm.kuet.ac.bd](mailto:ikramul3300@becm.kuet.ac.bd)

Reviewing editor:  
Sanjay Kumar Shukla, School of Engineering, Edith Cowan University, AUSTRALIA

Additional information is available at the end of the article

## CIVIL ENGINEERING | RESEARCH ARTICLE

# Machine Learning Methods to Predict and Analyse Unconfined Compressive Strength of Stabilised Soft Soil with Polypropylene Columns

Md. Ikramul Hoque<sup>1\*</sup>, Muzamir Hasan<sup>1</sup>, Md Shofiqul Islam<sup>2</sup>, Moustafa Houda<sup>3</sup>, Mirvat Abdallah<sup>3</sup> and Md. Habibur Rahman Sobuz<sup>4</sup>

**Abstract:** In this study, several machine learning approaches are used for the prediction of the unconfined compressive strength (UCS) of polypropylene-stabilised soft soil. This research work generates new data and applies several machine learning algorithms for the analysis of UCS. Fifty-two samples are in our generated data. In our generated data, five input features are used: Column Reinforcement Type, Column Diameter, Area replacement ratio, Column Penetration Ratio and Max\_Deviator Stress. On the other hand, the output consists of three target stress class. Our experimental result shows that Random Forest (RF) provides good prediction result of unconfined compressive test (UCT) and that is satisfied. RF model gets result of mean absolute error of 0.0625, mean square root error of 0.0625, root mean sqrt error of 0.2500,  $r^2$  value of 0.8942 and accuracy of 0.9375. In addition, the sequential model got training loss of 0.2535, training accuracy of 0.9024, validation loss of 0.4056 and validation accuracy: 0.9091. The results showed that the suggested RF and sequential model performs excellently in predicting the UCS of stabilised soft soil with polypropylene. Our technique is more practical and time-



Md. Ikramul Hoque

### ABOUT THE AUTHOR

Md Ikramul Hoque is working as an Associate Professor at Khulna University of Engineering and Technology (KUET), Now he is studying as a Ph.D. student at Civil Engineering, at University Malaysia Pahang (UMP). He is an active researcher in the field of Civil Engineering. He has Published 26 articles in different high-impact facto journals. His research focus is in Geotechnical Engineering, Construction engineering and management.

### PUBLIC INTEREST STATEMENT

This research work generates new data and applies several machine learning algorithms for the prediction of the unconfined compressive strength (UCS) of polypropylene-stabilised soft soil. Fifty-two samples are in our generated data. In our data, five input features are used: 'Type', 'Column Diameter', 'Column Penetration Ratio' and 'Max\_Deviator Stress'. On the other hand, the output consists of target stress class. The results showed that the suggested RF and sequential model performs excellently in predicting the UCS. This method is more practical and time-consuming than arduous laboratory work. Our method will be applied in real world for the following purpose: (1) increase the bearing capacity of soft soil; (2) improve the stability of soft soil; (3) minimise the settlement and lateral deformation of soft soil; (4) solve the problem of building infrastructure on soft soil and (V) reduce the expense of soft soil.

consuming than arduous laboratory work. In the future, we will do the experiment with various soft soil characteristics to develop high-performing machine and deep learning models.

**Subjects:** Artificial Intelligence; Foundations and Piling; Geomechanics

**Keywords:** Soft soil; UCT; Machine learning; Unconfined compressive strength; Prediction; ANN

## 1. Introduction

Unconfined compressive strength (UCS) of materials, such as soils and industrial waste products, is crucial for foundation design and construction, as well as slope stability analysis and long-term structural stability. A crucial factor in the design and functionality of the pavement is the stabilised material's UCS measurement. The physicochemical qualities of the materials, the cementitious admixture used to stabilise the material and the curing time all affect the UCS of stabilised materials (Suthar, 2020). Laboratory tests are used to explore a material's UCS, and accurate machining equipment is needed to measure UCS. For the laboratory approach to accurately determine UCS, the specimen's height and diameter must be known. In addition to being difficult, extensive and time-consuming, laboratory tests also make it extremely difficult to collect representative samples. Therefore, researchers have been obliged to examine alternative methodologies for prediction of UCS of stabilised soft soil with polypropylene due to economic considerations and practical issues.

Artificial neural networks (ANNs), a computational intelligence technique, have been utilised to predict the UCS of soil and rock material during the past few decades (Majdi & Rezaei, 2013; Sathyapriya et al., 2017), and these results support the effectiveness of this approach. The ideal number of neurons in the hidden layer, the number of hidden layers and various tuning parameters (such as learning rate, activation, optimisers) must all be set up before using a back-propagation neural network-based modelling approach. A neural network may overtrain if there are many training iterations, which could impair the model's ability to predict.

In a novel approach (Kim et al., 2021) a multi-layer perceptron (MLP) model was developed to spatially predict and assess the range of effective cohesion for residual soils, which are frequently linked to rainfall-induced slope failures in Singapore. As a consequence, by using the specified values of observed effective cohesion as learning data and four index soil parameters as input information, the suitable effective cohesions may be calculated using the MLP network.

Another recent study (Kim et al., 2021) looked at how well the soil parameters effective cohesion ( $c'$ ) as well as effective friction angle ( $\phi'$ ) might be used to predict the spatial variation of shear strength properties. Regression kriging (RK), Random Forest (RF), as well as ordinary kriging (OK) were contrasted and assessed for their abilities to predict the  $c'$  and  $\phi'$  of residual soils in Singapore. Additionally, this research demonstrated that RF and RK have been more responsive to variation in sample size than OK. The advantages of auxiliary variables for mapped shear strength characteristics are shown by these findings.

A new method with a soil database, including saturated and unsaturated hydraulic and mechanical soil parameters, was created (Li et al., 2022). The unidentified soil qualities were predicted using machine learning techniques. Using the standard kriging method, geographical distributions of various unsaturated as well as saturated soil properties have been generated depending on the projected soil parameters. The created database contains the average values of SWCC, saturated as well as unsaturated strength of shear parameters, including saturated permeability in the various zones. Alternatively assuming constant soil qualities, the proposed data set can be used

in regional GIS-based water balance and slope stability assessments to take into account geographical variability.

The utilisation of the algorithm named support vector machine (SVM), Gaussian processes (GPs), RFs, M5 model trees (Pal & Deswal, 2009; Solomatine & Dulal, 2003, 2003; Solomatine & Y, 2004, 2004) and regression-based technique for various civil engineering issues has been described in multiple publications over the past 10 years (Lamorski et al., 2008; Lee & Chern, 2013; Mahesh & Deswal, 2010; Sabat, 2015; Samui, 2008). Sometimes, the adoption of the M5 model tree approach is better than the neural network methods, according to many researchers (Bhattacharya and Solomatine & Dulal, 2003) (Lamorski et al., 2008; Mahesh & Deswal, 2010; Sabat, 2015; Solomatine & Dulal, 2003) Gaussian processes are conceptually and practically easier to build than back-propagation neural networks. (Ayyappa et al., 2020) GP models have a close relationship with strategies like SVM (Vapnik et al., 1995). A thorough examination of the literature finds (Suthar, 2020) that no studies have used these modelling techniques to forecast the UCS of stabilised soft soil with polypropylene to yet. In an effort to investigate its potential for forecasting the UCS of stabilised soft soil using laboratory data set, an attempt has been made to keep in mind the utility of these modelling methodologies in civil engineering applications. Depending on the Random Forest Regression Algorithm, Intelligent Model was used to predict compressibility of cement-stabilised dredged soil (Guo et al., 2021). Stress-strain modeling of sands used ANNs (Ellis et al., 1995).

In our proposed method, we generate data from our laboratory with 53 samples of soft soil pipe with polypropylene. Then we developed several machine learning algorithm for prediction of UCS of stabilised soft soil with polypropylene. The main contributions of this method are listed below:

- (a) Designed an unconfined compression testing (UCT) system that quickly estimates the compressive strengths of soft soils with adequate cohesion to allow testing in an unconfined condition.
- (b) Developed several machine learning algorithms for prediction of UCS of stabilised soft soil with polypropylene.
- (c) Checked the validation and robustness of the proposed method that is cost effective with less computational complexity.

## 2. Related study

Hassan et al. (2019) performed several experiments to investigate the increase in clay shear strength by embedding a single column of crushed coconut shells (CCS). Because of CCS, a waste material, the cost of soil improvement may be minimised. To determine the shear strength, the UCT was performed with four batch kaolin samples with control mechanism. Four samples were used in each batch to find the precise value. Test variables are the height of the compressed coconut shell columns of 100 mm, 80 mm and 60 mm, in which the individual penetration ratio of the column is 1.00, 0.80 and 0.60. A total of 16 UCTs with a height of 100 mm and a diameter of 50 mm were conducted on the kaolin specimen. For the 4% area-displacement ratio at column penetration ratio of 0.60, 0.80 and 1.00, the increase in shear strength by embedded with crushed coconut shell columns is 19.02%, 34.76% and 24.34%. Result-based relation among the increases in shear strength is affected by the column height. The highest strength is not created by the maximum column height, and this proves that the principle of “critical column length” is valid in this analysis.

Aslani et al. (2019) directed an exploratory program to examine the shear strength of the stone column-reinforced clay bed. A broad direct shear testing system with in-plane dimensions of 305 × 305 mm to model the undrained, short-term behaviour of the clay bed reinforced with a stone column. In this examination, the impact of key boundaries, including zone substitution proportion, stone section plan, typical weights worth and stone segment material, was tentatively assessed.

Three distinct replacement ratios, stone column arrangements (single, square and triangular) and normal pressures (35, 55 and 75 kPa), and two materials, including crushed gravel and fine-grained sand, were used in the tests for this purpose. The results demonstrated that the shear strength and the total rigidity of the clay bed improved in the presence of the stone column. The structure of the stone column, the replacement ratio of the area and the material of the stone column have also been shown to affect improving shear strength. Square column arrangements and single columns were connected to the most and least increment in shear strength and stiffness values, respectively. The variance of the stress concentration ratio of stone columns under shear loads was calculated using appropriate instruments in this analysis. The equivalent shear intensity and equivalent shear parameters estimated from tests were also compared with those predicted by analytical relationships at the value of stress concentration of 1 and stress fixation esteem acquired from tests.

An experimental analysis by Hasan (2018) determines the undrained shear strength of clay reinforced with group encapsulated lime bottom ash columns and their association with the different sizes of group encapsulated lime bottom ash columns. Small-scale modelling column samples with a diameter of 50 mm and a height of 100 mm were constructed utilising 10 mm and 16 mm diameters of the lime bottom ash column. Each diameter shall have three distinct column lengths of 60 mm, 80 mm and 100 mm with three samples of each category of length. As per experimental findings, the existence of group encapsulated columns of lime bottom ash improved the strength properties of the soft soil. Since the rise in shear strength relies not only on the critical column length and the penetration ratio of the community of encapsulated columns of lime bottom ash but also on the column diameter of lime bottom ash.

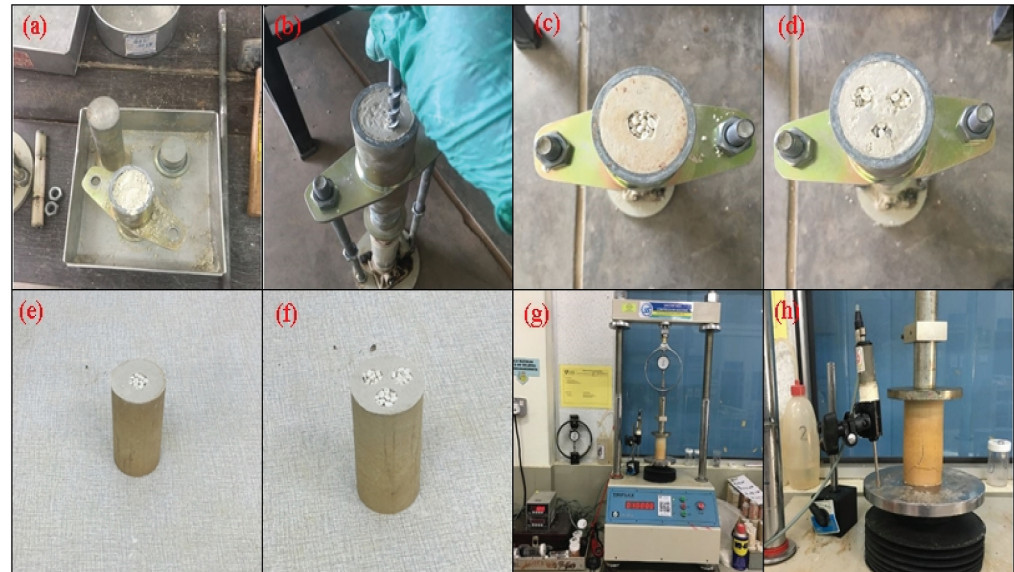
BLACK et al. (2011) studied the settlement behaviour of single and small groups of stone columns through a system analysis of a particularly evident triaxial cell. The effect of the slenderness and the area replacement ratios on the settlement's efficiency of the strengthened stone column was examined. The settlement improvement factor increases with an increase in the area replacement and slenderness ratio of the stone column up to 30–40% and 8–10, accordingly, above which these variables have a marginal impact.

The stability of stone columns in clay was investigated by Zahmatkesh and Choobbasti (2010). Finite element studies have been carried out using 15-noded triangular elements utilising Plaxis software to test the clay settle reinforced with stone columns. A drained analysis of clay, stones and sand was conducted using the Mohr–Coulomb criterion. Interface elements were used in the interface between the stone column and clay. For the measurement of the stresses due to soil compaction, the column installation was simulated. The lateral earth pressure coefficient after the installation of the stone column and the soil's Settlement reduction ratio (SRR) were determined from numerical results. Based on this study, the variation of stress in soft soil after column distance from column installation is substantially reduced.

### 3. Materials

We use our own data set. Here we give an overall description of data to present overall information about our data. In this study, several approaches were applied experimentally for prediction of UCS of stabilised soft soil with polypropylene. We implemented each model on our data set with 52 samples. Therefore, in 52 data sets, 80% in training and the remaining 20% in testing the model. The data set consists of Column Reinforcement Type, Column Diameter (mm), Column Penetration Ratio and Max Deviator Ratio as UCS (kPa). The overall details of the data are given in Figure 1. The data set is categorised into two groups as input parameters and output parameters. For predicting the UCS values, five input parameters consisting of Column Reinforcement Type, Column\_Diameter, Column\_Penetration\_Ratio and Max\_Deviator\_Stress value were used to get UCS value at different time. Evaluation outcome of different model on the basis of different statistical outcomes of machine learning, that is, mean absolute error (MAE), mean square error (MSE), root mean sqrt error (RMSE),  $r^2$  value, accuracy, precision, recall and F1 score. These

**Figure 1. (a)–(h) Sample preparation and testing in laboratory.**



machines learning statistical parameters were computed to evaluate the model fitness. The lower value of the loss function and greater accuracy value suggest that the model is more fit. The user-defined parameter's ideal value for evaluating the UCS data set is shown in Figure 1. The choice of user-defined parameters affects how well each model performs. After performing numerous trials, the best values (constants) for various user-defined parameters utilised in various models have been determined. The performance of the model in both training and testing was taken into account while deciding the constant values. The performance of models is most frequently impacted by variations in constant value in testing mode, while this impact is minimal in training mode. In order to prevent overfitting, the constants were chosen so that models perform optimally in both training and testing modes. In Figures 1(a-h), it shows the sample preparation tools and material used in the laboratory. In our method, soft soil is just kaolin (kaolin is the powder form of China clay) mixed with water. The kaolin will be air-dried and then combined with 20% water, which is the optimum moisture content of the kaolin acquired from the standard compaction test. It will be poured into the customised steel mould after uniform mixing of the soil and compacted into three layers so that no air voids are left in the soil. With five free fall blows of a 3.1 kg personalised steel hammer, each layer will be compacted. Because of the addition to that amount of water, the kaolin clay is free to establish its structure by adding pressure on it. The custom mould will be constructed to compress the amount of clay used within a sample of 50 mm in diameter and 100 mm in height. Then the specimens will be extruded from the mold and placed in a special case, and left for at least 24 h to maintain the pore pressure inside the specimen. Then the samples were tested with the UCT machine.

In Table 1, target class is target stress class. Target stress class is classified into three categories: Low Stress, Mid Stress and Standard Stress. These target classes are labelled based on the value of Max deviator stress. If the Max\_Deviator\_Stress is below 15 kPa it is labelled to low stress; if Max\_Deviator\_Stress is above 15 kPa and below or equal to 18 kPa it is labelled as mid stress and if Max\_Deviator\_Stress is above 18 kPa it is labelled as standard stress.

Here we added Tables 2 and 3 to get deeper information of our data set. In Table 1, the name of the data column and its data type is shown on the right side. Here 52 indicates the total number of samples. In Table 3, statistical information is clearly presented. Count, mean, standard deviation, minimum value, percentile value and maximum value are given in Table 3.

**Table 1. Overview of our data set with 10 samples**

Sl. no.	Type	Column diameter (mm)	Area replacement ratio	Column penetration ratio	Max deviator stress (kpa)	Target stress class
1	0	0	0	0	12.56	Low_Stress
2	0	0	0	0	12.80	Low_Stress
3	0	0	0	0	12.40	Low_Stress
4	0	0	0	0	12.84	Low_Stress
5	1	10	4	0.6	14.44	Low_Stress
6	1	10	4	0.6	17.42	Mid_Stress
7	1	10	4	0.6	20.68	Standard_Stress
8	1	10	4	0.6	20.38	Standard_Stress
9	1	10	4	0.8	14.40	Low_Stress
10	1	10	4	0.8	16.18	Mid_Stress

**Table 2. Overview of our data set and its types**

Sl. no.	Column	Non-null count.	D type
0	No.	52 non-null	int64
1	Type	52 non-null	int64
2	Column diameter	52 non-null	int64
3	Area replacement ratio	52 non-null	float64
4	Column penetration ratio	52 non-null	float64
5	Max. deviator stress	52 non-null	float64
6	Target stress class	52 non-null	object

Here we have added Figure 2 that gives overall amount of data of each target class. From Figure 2, it is clearly identified that Mid Stress and Standard Stress class are almost similar in amount of 20 among 52 samples. This indicates our data output gives good result of UCS stress.

We have so made a correlation matrix among the input features, as shown in Figure 3. In this figure, value 1 indicates that it has full similarity and other values less than 1 indicate that it has less similarity. From the analysis of Figure 3, we see that our input features of the data set are not the same value.

#### 4. Methodology

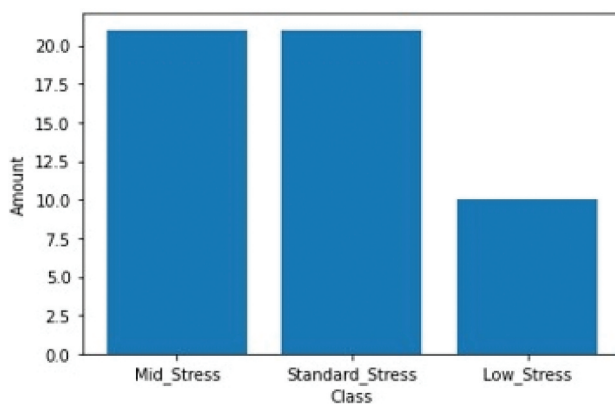
The basic model put forward is divided into several components. It begins by gathering data. In order to forecast the UCS of stabilised soft soil using polypropylene, the data are then sent to various machine learning models. The final prediction and analysis are finished after the execution of numerous algorithms. There are significant differences in performance metrics between the various artificially intelligent algorithms that were presented. Ten conventional machine learning algorithms were created and assessed by the models. The optimum algorithm for forecasting the UCS of stabilised soft soil with polypropylene was then recommended after comparing their results.

After the pre-processing, data is divided is 90% for training and 10% for validation. Implementation model is the crucial tasks and then result is analysed to predict UCS of stabilised soft soil with polypropylene. The main process flow of the model is shown in Figure 4.

**Table 3. Statistics of our data set**

Describing	No. of samples	Type of polypropylene column	Column diameter (mm)	Area replacement ratio	Column penetration ratio	Max. deviator stress
Min.	4.00	0.00	0.00	0.00	0.00	12.400
25%	16.00	1.00	10.00	4.00	0.60 & 0.80	14.968
50%	28.00	1.00	16.00	10.24	0.60 & 0.80	17.200
75%	40.00	3.00	10.00	12.00	0.80 & 1.00	20.305
Max.	52.00	3.00	16.00	30.72	0.80 & 1.00	23.800

**Figure 2. Amount of data per target class.**



#### 4.1. Machine learning algorithms

We applied several machine learning algorithms and one deep learning algorithm to predict UCSs of stabilised soft soil with polypropylene. Here we introduced all the algorithms and its implementation process with advantages to predict UCS of stabilised soft soil with polypropylene.

##### 4.1.1. Bernoulli naive Bayes

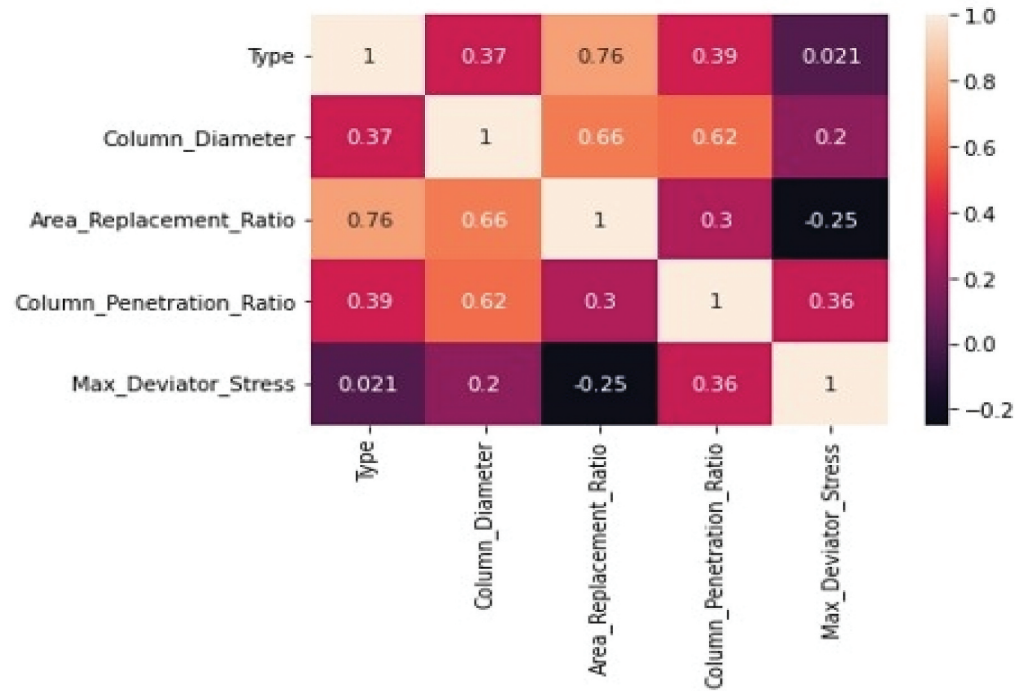
Bernoulli A variation of naive Bayes (Artur, 2021) is naive Bayes. So let us quickly discuss naive Bayes first. Naïve Bayes classifier is a machine learning classification technique based on the Bayes theorem that estimates how likely an event is to occur. As a probabilistic classifier, the naive Bayes classifier estimates the likelihood that a given input will fall into each of the classes. Another name for it is conditional probability. Naive Bayes classifier is based on two key presumptions: the first is that the qualities are unrelated to one another and do not affect one another’s performance, which is why it is named ‘naive’. The second is that equal weight is given to each feature.

##### 4.1.2. K neighbours

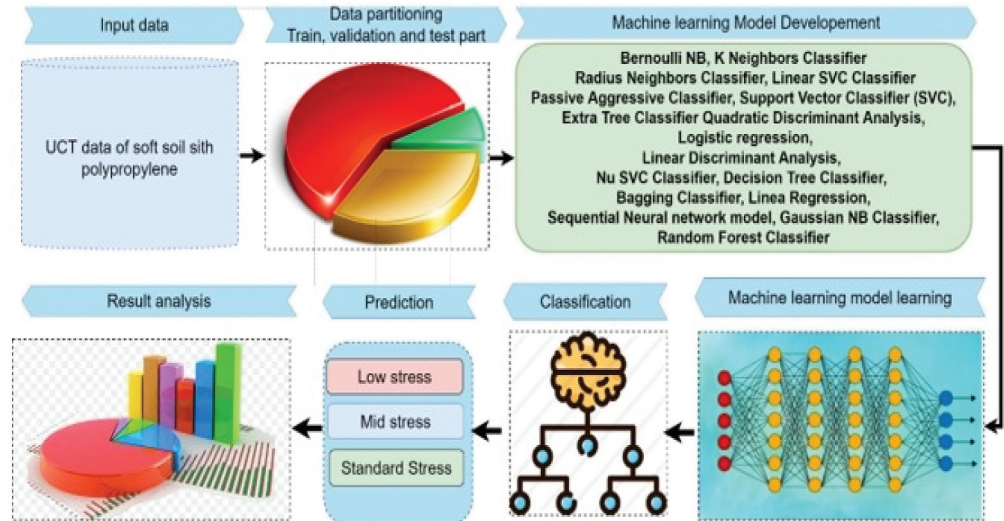
A straightforward, simple-to-implement strategy that may be used to solve both classification and regression problems is the supervised learning method known as k-nearest neighbours (KNN). According to the KNN (Moreira et al., 2007) method, related objects are clustered together. Or to put it another way, comparable things are nearby. K means, a method of clustering, divides observations into k groups. It may be easily applied to classification because the number of nodes can be specified. The goal of the k-means clustering technique is to group together k clusters from a given anonymous data set (one that lacks information about class identification). At first, k centroids are chosen in total. A centroid is a hypothetical or real piece of data that sits in the middle of a clustering.



**Figure 3. Correlation matrix among target class.**



**Figure 4. Proposed machine learning model for prediction of unconfined compressive strength of stabilised soft soil with polypropylene.**



#### 4.1.3. Naive Bayes

The Bayes theorem (WEBB et al., 2010) is used to construct a naive Bayes classifier. It calculates member probabilities for each categorisation, including the likelihood that a specific record or piece of data falls into that category. The class with the highest probability is the most likely.

#### 4.1.4. Radius neighbours classifier

The radius neighbours classifier (Robert 2017) is a machine learning technique for classification. It is a modification to the k-nearest neighbours approach that uses all instances within a new example's radius to produce predictions rather than just the k-closest neighbours. The k-nearest neighbours algorithm, or KNN, is the foundation of it. The full training data set is taken and stored

while using KNN. Then, for each new case for which we wish to make a prediction, the k-closest examples in the training data set are found. The new example is then given the mode (most popular value) class label from the k neighbours.

#### 4.1.5. *Linear SVC classifier*

With a high number of data, the linear support vector classifier (SVC) (Guyon, 2000) method performs well. It uses a linear kernel function to perform classification. When compared to the SVC model, the linear SVC adds more parameters, including the error function and penalty normalisation, which applies “L1” or “L2.” Because linear SVC is based on the kernel linear technique, the kernel technique cannot be modified.

#### 4.1.6. *Passive aggressive classifier*

The category of online learning machine learning techniques includes passive aggressive classifiers (Chang et al., 2010). It operates by acting passively in response to accurate classifications and aggressively in response to incorrect classifications. The reason why passive aggressive algorithms are thus named is passive: maintain the model and make no changes if the prediction is accurate. In other words, the example data are insufficient to alter the model in any way. Aggressive: modify the model if the prediction turns out to be inaccurate. In other words, a model modification could make it right.

#### 4.1.7. *SVC classifier*

SVC stands for C-support vector classification, and libsvm is used to implement it. Scikit-Learn makes use of the sklearn, svm SVC module. A linear SVC's (Guyon, 2000) goal is to split or categorise the data you supply by returning a “best fit” hyperplane. You may then feed some features to your classifier to get the “predicted” class after acquiring the hyper plane.

#### 4.1.8. *Extra tree classifier*

This class implements a meta estimator that employs averaging to increase predictive accuracy and reduce overfitting. The meta estimator fits a number of randomised decision trees (also known as extra-trees) on different sub-samples of the data set. Extremely randomised trees classifier (Wang et al., 2009), also known as extra trees classifier, is a form of ensemble learning method that combines the findings of various de-correlated decision trees gathered in a “forest” to produce its classification outcome.

#### 4.1.9. *Quadratic discriminant analysis*

A generative model is quadratic discriminant analysis (QDA) (Tharwat et al., 2017). Each class is thought to have a Gaussian distribution according to QDA. The proportion of data points that belong to the class is the class-specific prior. The average of the input variables that are part of the class makes up the mean vector particular to that class. Although QDA has more parameters to estimate, it tends to fit the data better than LDA because it gives the covariance matrix greater freedom. With QDA, the number of parameters considerably rises because each class will have a unique covariance matrix with QDA.

#### 4.1.10. *Logistic regression*

A linear regression model is a categorisation model (WRIGHT, 1995). A statistical analysis technique called logistics regression uses historical data set observations to forecast data. It takes advantage of a logistical property to represent binary variables despite the expansions being more complex.

#### 4.1.11. *Linear discriminant analysis*

Prior to classification, the set of attributes is generally reduced using linear discriminant analysis (LDA) to a more manageable quantity. Each of the extra dimensions is a template created by a linear combination of spatial domain. Five steps can be used to execute LDA (Xanthopoulos 2013): Create the mean vectors from the data set for each class. Determine the scatter matrices (in-

between-class and within-class scatter matrices). Calculate the scatter matrices' eigenvectors and accompanying eigenvalues.

#### 4.1.12. Nu SVC classifier

Nu support vector machines (SVM) (Pérez-Cruz et al., 2003) is a supervised learning technique that may be applied to both classification and regression issues. We can utilise the SVC and SVM-based classifier to solve classification issues. This link states that the nu parameter for the one-class SVM model is specified by nu. The nu parameter serves as a lower and upper bound for the proportion of samples that are support vectors and samples that are on the incorrect side of the hyperplane, respectively. 0.1 is the default.

#### 4.1.13. Decision tree classifier

In decision tree classifier (Safavian and Landgrebe, 1991), the pattern is shown as a binary tree but we utilised it to classify three types of UCT. The root node was in fact the very first node. A flow path is created by a series of data set-related questions (characteristics as well as associated information that goes along with them), the pathway is then divided and the result is then predicted. The leaves stand in for the various data set classifications. Most often, the decision tree is used in multiclass classification.

#### 4.1.14. Linear regression classifier

Once the classes can be distinguished in the feature space by linear bounds, logistic regression is typically utilised as a linear regression classifier (Jifu Hao, 2012). However, if we happen to have a better understanding of the form of the decision border, that can be fixed. There are two reasons why linear regression is ineffective for classifying data. The first is that classification issues demand discrete data while linear regression deals with continuous values. The second issue is with how the threshold value changes as more data points are entered. Continuous data are employed with prediction model. With categorical data, classification algorithms are applied. We look for the best fit line in regression in order to more precisely forecast the result.

#### 4.1.15. Bagging classifier

An ensemble meta-estimator called a bagging classifier (ZAREAPOOR & SHAMSOLMOALI, 2015) fits base classifiers one at a time to random subsets of the original data set, and then it aggregates the individual predictions (either by voting or by averaging) to produce a final prediction. An ensemble learning technique called bagging, often referred to as Bootstrap aggregating, aids in enhancing the efficiency and precision of machine learning algorithms. It lowers the variance of a prediction model and is used to handle bias-variance trade-offs. The benefit of bagging is that it enables several weak learners to work together to outperform a single strong student. Additionally, it aids in lowering variance, which eliminates overfitting of models during the process. Bagging has the drawback of making a model's interpretability more difficult.

#### 4.1.16. Gaussian NB classifier

A supervised machine learning technique is the Gaussian processes classifier i.e (Ali Haghpanah Jahromi 2017). A powerful non-parametric machine learning system for classification and regression can be built using Gaussian processes, a generalisation of the Gaussian probability distribution. Continuous valued features are supported by Gaussian naive Bayes, which also models each as following a Gaussian (normal) distribution. Each method for developing a simple model is to assume that the data has a Gaussian distribution with really no covariance (independent dimensions) among dimensions.

#### 4.1.17. Random forest

Random forests (GUO et al., 2021) are regarded as a highly resilient and accurate strategy because of the significant number of decision trees utilised in the process. It is unaffected by the overfitting issue. The main justification for this is that it averages every forecast, eliminating any biases. A random forest produces correct estimates that are easy to understand. Large data sets can be

handled effectively. In comparing to the method proposed, the random forest algorithm predicts outcomes with greater accuracy. Depending on the Random Forest Regression Algorithm, Intelligent Model was used to predict the compressibility prediction of cement-stabilised dredged soil (GUO et al., 2021).

#### 4.1.18. Neural nets

A paradigm for machine learning that is based on brain activity. In every stage of the network, neurons are present, including those from the central nervous system. The model is able to comprehend any information that travels through these synapses and provide the intended outcome. A non-linear framework is frequently employed when there is a complex relationship between the inputs and outcomes. In our implementation we use and manually implemented sequential neural network model to predict UCS of stabilised soft soil with polypropylene. We use dense layer, drop out of 0.5, batch size 32, total epoch 1000 and Adam optimiser with soft max activation function. Stress-strain modelling of sands used ANNs (ELLIS et al., 1995).

#### 4.2. Main algorithm

This section discusses the operation of the main algorithm. The method works to forecast the UCS of stabilised soft soil with polypropylene using raw information from the database. This fundamental algorithm describes the operation of our model at each phase. After receiving input, the forecast of the UCS of stabilised soft soil with polypropylene is forwarded to various levels.

##### Algorithm 1

- (1) *Begin*
- (2) *Input: Raw data*
- (3) *Output: Predicted class of for prediction of unconfined compressive strength of stabilized soft soil with polypropylene.*
- (4) *Initialization of parameters: Target Class C: = 3, Number of models M: = 10*
- (5) *Data splitting for train and test*
- (6) *Getting input single sample from S from raw data*
- (7) *Pre-processing of samples S: Scaling and transformation*
- (8) *Hyper parameter tuning of models*
- (9) *Number of models M: = 10, Correct prediction: = 0*
- (10) *For each I in range I to M*
- (11) *Train I-th model with train data (80% of raw data)*
- (12) *P:=Prediction of test S data by the I-th model*
- (13) *If P is True for any of target C class*
- (14) *Correct prediction:= Correct prediction + 1*
- (15) *Evaluate performance metrics of the I-th model using Correct prediction*
- (16) *End If*
- (17) *End For*
- (18) *End*

#### 4.3. System setup

On a laptop with a solid internet connection, the model was created. A free cloud-based solution for machine research and education called Google Collaboratory was implemented. It uses the same Jupyter interface. It includes a fully set up environment for in-depth research and unrestricted use of a potent GPU. The default setting for the remaining settings was determined to be device acceleration. On a computer running Windows 10 and equipped with five Intel (R) 3.60 GHz

**Table 4. Testing performance using different machine learning models**

Model	MAE	MSE	RMSE	$r^2$	Accuracy
Bernoulli NB	0.1624	0.1645	0.49451	0.6551	0.4731
K neighbours classifier	0.3750	0.3750	0.6123	0.3642	0.6250
Radius neighbours classifier	0.3750	0.3750	0.6123	0.3642	0.6251
Linear SVC classifier	0.3125	0.3125	0.5590	0.4701	0.6875
Passive aggressive classifier	0.3751	0.5001	0.7071	0.1523	0.6875
Support vector classifier (SVC)	0.2500	0.2500	0.5000	0.5761	0.7510
Extra tree classifier	0.2500	0.2500	0.5000	0.5761	0.7511
Quadratic discriminant analysis	0.3750	0.5001	0.7071	0.1523	0.6875
Logistic regression	0.1250	0.1250	0.3535	0.7881	0.8750
Linear discriminant analysis	0.1250	0.1250	0.3535	0.7881	0.8750
Nu SVC classifier	0.1250	0.1250	0.3535	0.7881	0.8750
Decision tree classifier	0.1251	0.1251	0.3535	0.7880	0.8750
Bagging classifier	0.1250	0.1250	0.3535	0.7881	0.8751
Linear regression	0.2876	0.1266	0.3559	0.7852	0.8901
Sequential neural network model	0.1592	0.0771	0.2777	0.3472	0.9091
Gaussian NB classifier	0.0626	0.0626	0.2501	0.8941	0.9374
Random forest classifier	0.0625	0.0625	0.2500	0.8942	0.9375

processors, 16 GB of RAM, the categorisation was done. Tables 4 displays all of the experimental studies' improved results.

#### 4.4. Experimental setup

The suggested method developed with different machine learning method to demonstrate the performance. We have done the fine-tuning on our model. As a matter of fact, we divided the data set into two parts: 80% for training and 20% for validation. At every round of training, binary cross-validation has been used to train a model. And then 20% random data is used only for testing as well as assessing the framework.

#### 4.5. Evaluation metrics

We use the following evaluation metric to judge the performance of our model. To assess the method, we used MAE, MSE, RMSE,  $r^2$  value and accuracy score for the purpose of validation. used in checking our model performance is calculated with Equations (1)–(5). The mean amount of the errors in a collection of forecasts is known as the MAE. We are aware that an error is essentially the

difference in absolute terms between the true or actual values and the expected values. Because of the absolute distinction, results with a negative sign are ignored. The difference between the actual values and the predicted values, squared, is what is known as the MAE. The standard deviation of the errors that happen when a prediction is made based on a data set is known as RMSE. This is the same as MSE, but when assessing the model's correctness, the value's root is taken into account. As the coefficient of determination, it has additional names. This measure indicates how well a model matches a specific data set. It shows how closely the anticipated values (i.e. the regression line) matches the actual data values. If machine learning algorithm predicts correctly then it is called true. If machine learning algorithm predict wrongly then it is called false.

$$\text{Accuracy} = \frac{\text{Number of Prediction}}{\text{Total Number of Prediction}} \quad (1)$$

$$\text{MAE} = \sum_{i=0}^n (y_i - x'_i) \quad (2)$$

$$\text{MSE} = [\sum_{i=0}^n (y_i - y'_i)]/n \quad (3)$$

$$\text{RMSE} = \sqrt{[\sum_{i=0}^n (y_i - y'_i)]/n} \quad (4)$$

$$r^2 = [1 - [\sum_{i=0}^n (y_i - y'_i)] / [\sum_{i=0}^n (y_i - y^{\wedge})]] \quad (5)$$

## 5. Result analysis

This section describes the results analysis in detail, including qualitative and comparative analysis. Because the performance of the training set is slightly better than that of the compared methods, we fine-tuned the parameters of our model.

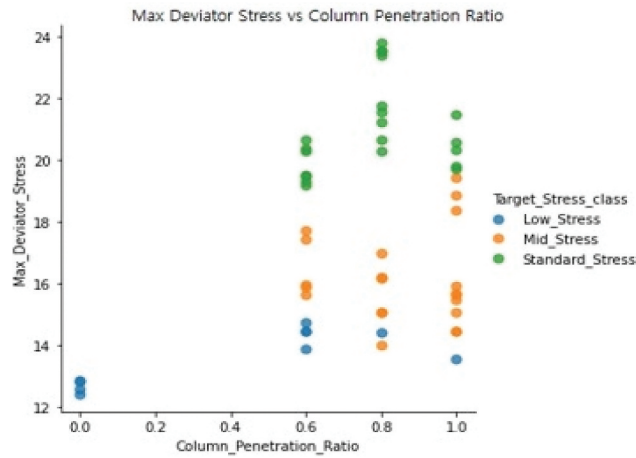
### 5.1. Qualitative result analysis

In the qualitative result analysis, we analyse the data to extract features, correlation and other statistical analysis. This analysis helps to understand the validity of the proposed method and its performance. Here we added Figures 5–7 to show the correlation among the input features and target features. In Figure 5, the scattering of the correlation among “max deviator stress” and “column penetration ratio” is shown. In Figure 6, the scattering of the correlation among “Column diameter” and “Area replacement ratio” is shown. In Figure 7, the scattering of the correlation among all columns of the data is shown using a TNSE plot. We also have analysed principal component (PC). PCs in PCA are the vectors representing direction of variance of data. PC corresponding to highest eigenvalue is the direction of max variance. In Figure 8, we show the amount of PC in PCA analysis. Another important Figure 9 is added to present the correlation among different features of data individually. This figure clearly illustrates each features value based correlation and visualisation.

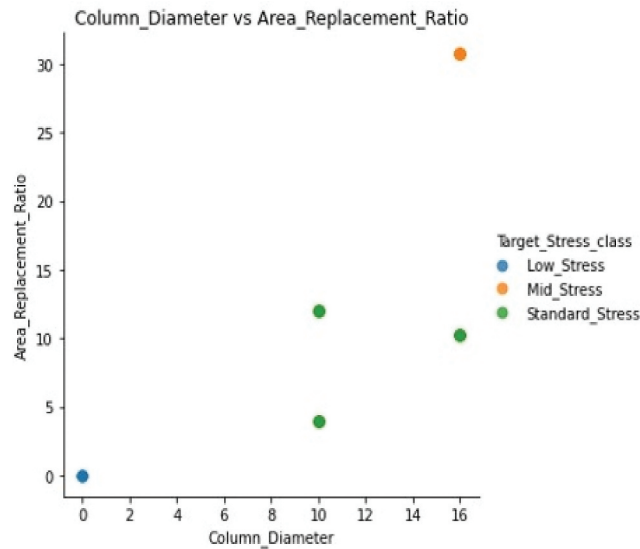
### 5.2. Comparative result analysis

In the qualitative result analysis, we analyse the data to give overall quantitative result analysis with comparison. This analysis helps to understand the validity of the proposed method and its performance. Based on the comparison of performance review parameters, the Random Forest Classifier modelling approach outperforms other modelling approaches for this data set. As a result of MAE 0.0625, MSE 0.0625, RMSE 0.2500,  $r^2$  value 0.8942 and accuracy of 0.9375 were obtained for testing the data set using Random Forest model.

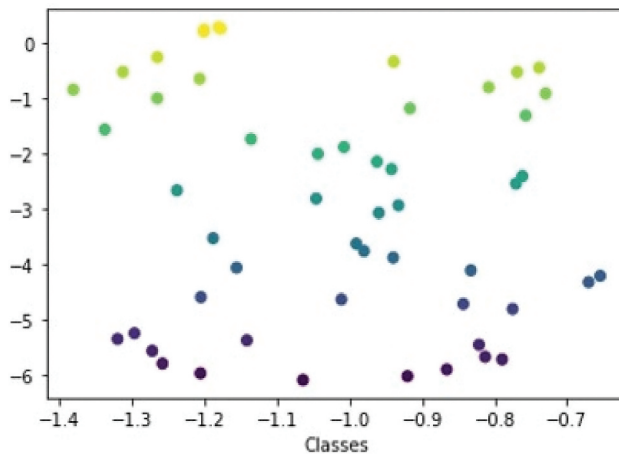
**Figure 5. Correlation matrix among first two target class.**



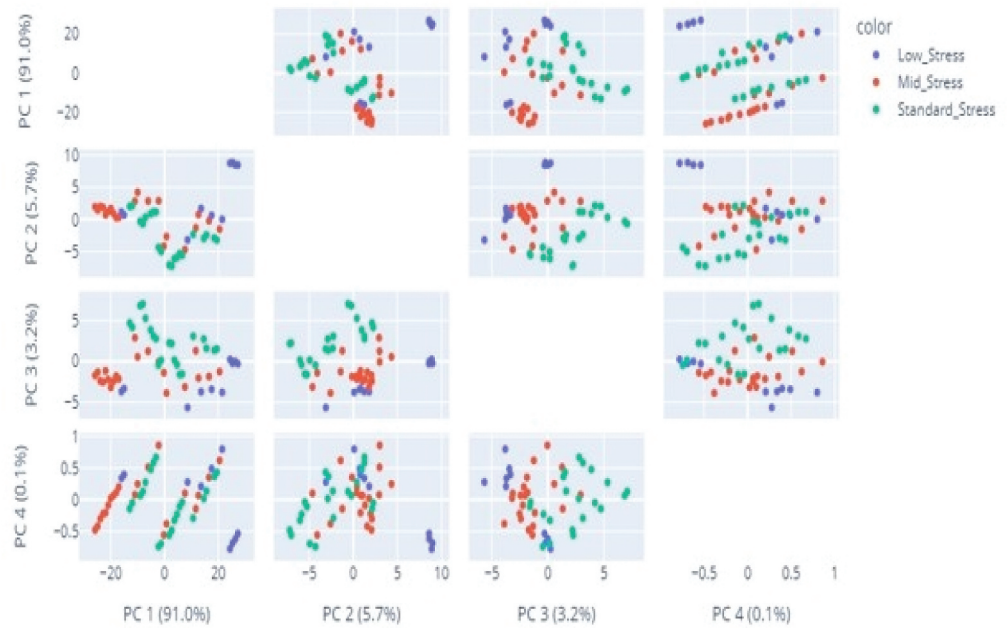
**Figure 6. Correlation matrix among last two target class.**



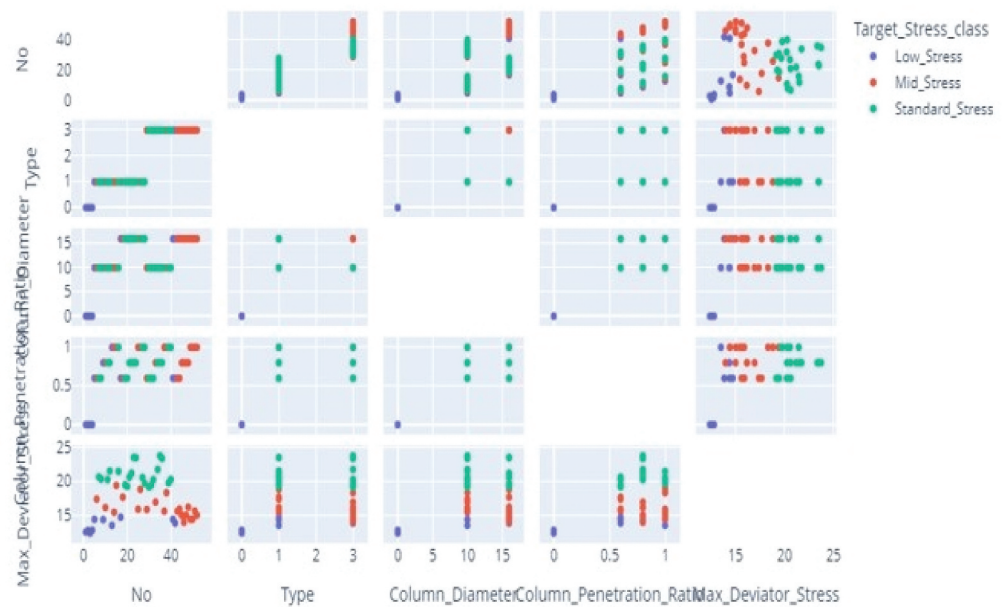
**Figure 7. Correlation among the input features and target features.**



**Figure 8. PCA among target class.**



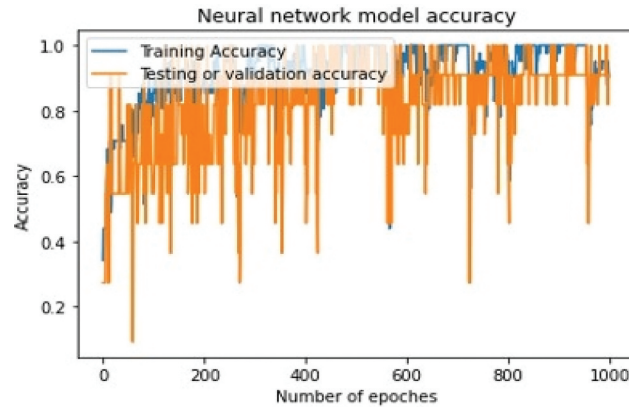
**Figure 9. PCA among target class among each features and target class.**



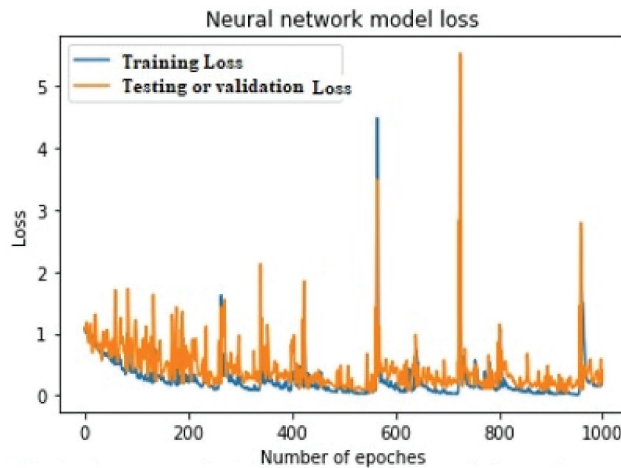
From the analysis result of different model from Table 4, we can easily decide that Random Forest model gives better result than all other compared model. We choose this model as a best performed model for our UCT analysis and prediction of soft soil with polypropylene. Here we added Figure 9 to present the training and validation accuracy of the sequential deep learning model. Figures 10 and 11 present the training and validation accuracy and loss of the sequential deep learning model. Our sequential deep learning model got training loss of 0.2535 and validation loss of 0.4056. ROC curve of the sequential model is clearly illustrated in Figure 12; its ROC area is 94.6011.



**Figure 10. Accuracy of sequential neural network model.**



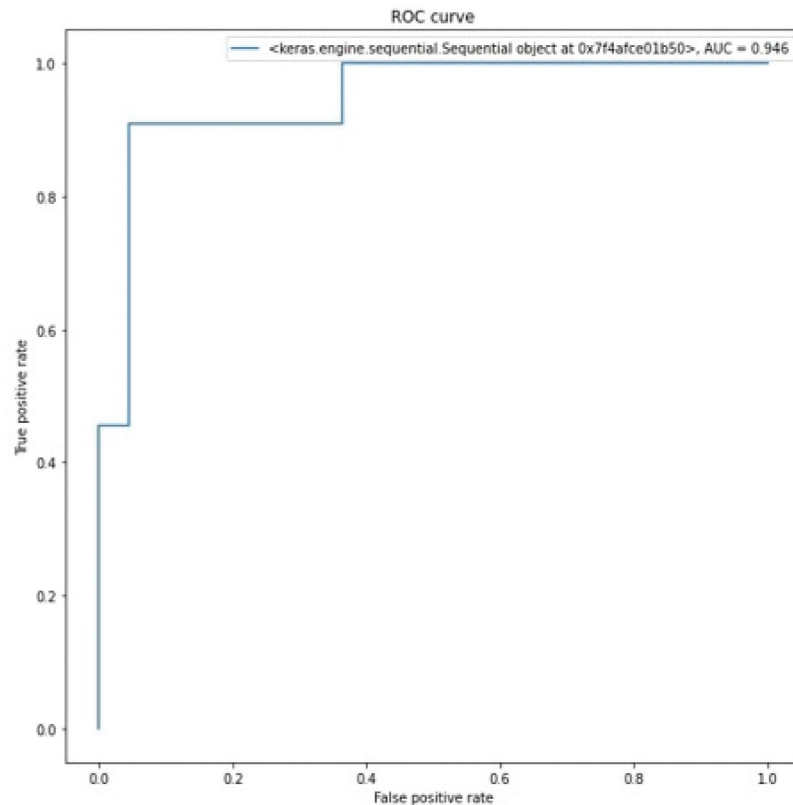
**Figure 11. Loss of sequential neural network model.**



A multi-layer perceptron (MLP) model was created using a novel methodology (Kim et al., 2021) to spatially predict and evaluate the range of effective cohesion for residual soils, which are typically associated with rainfall-induced slope failures in Singapore. As a result, the appropriate effective cohesions may be estimated using the MLP network by employing the stated values of observed effective cohesion as learning data and four index soil parameters as input information. The effectiveness of the soil factors, including effective cohesion, friction and other variables, was the focus of another recent study (Satyanga and Rahardjo, 2021). These results demonstrate the benefits of auxiliary variables for mapped shear strength characteristics. Unsaturated and saturated hydraulic and mechanical soil parameters are included in a novel technique that was developed (Li et al., 2022). Another proposed data set can be used in regional GIS-based water balance as well as slope stability evaluation to take into account geographical variability. A method (Suthar, 2020) that has the machine learning modelling methods to predict the UCS of stabilised soft soil with polypropylene. Depending on the Random Forest Regression Algorithm predicts Compressibility of cement-stabilised dredged soil (Guo et al., 2021).

We generate our unique data in the laboratory with five soft soil features (Column Reinforcement Type, Column Diameter, Column Penetration Ratio, Max\_Deviator Stress). We predict the UCS with 17 machine learning models. Our method got higher accuracy of 93.75% with the Random Forest model. Our method is to create a technique for unconfined compression testing that can quickly calculate the compressive strengths of soft soils with sufficient cohesiveness to permit testing in an open environment. We created different machine learning algorithms and

**Figure 12. ROC curve of sequential neural network model.**



examined those techniques for forecasting the UCS of stabilised soft soil with polypropylene. Finally we recommend that our method is cost-effective and robustness of the suggested strategy, which requires less computing.

## 6. Discussion

Our implementation and its outcome suggest that ANN model named as Random Forest model is performing better than all other machine learning models. As a result of MAE 0.0625, MSE 0.0625, RMSE 0.2500,  $r^2$  value 0.8942 and accuracy of 0.9375 were obtained for testing the data set using Random Forest model. Additionally, the sequential model got loss 0.2535, training accuracy of 0.9024, validation loss 0.4056 and validation accuracy 0.9091. Random Forest is better than all other methods; it is clearly expressed in Table 4. Based on sensitivity analysis, the relevant parameters for estimating the UCS of stabilised soft soil with polypropylene are the column penetration ratio, column diameter and maximum deviator stress. The results showed that the suggested model can predict the UCS of stabilised soft soil with polypropylene with a high degree of accuracy. However, the proposed modelling method shows that it is more practical and less time-consuming than laborious laboratory work. We have got some potential guidelines for research idea from a paper of SHUKLA, 2022. The application of our method will be applied in real world for the following purpose: (1) increase the bearing capacity of soft soil; (2) improve the stability of soft soil; (3) minimise the settlement and lateral deformation of soft soil; (4) solve the problem of building infrastructure on soft soil and (5) reduce the expense of soft soil.

## 7. Limitation of the study

In our method we use five input features for the prediction of UCS of stabilised soft soil. Soft soil is just kaolin (kaolin is the powder form of China clay) mixed with 20% water. In the future we will

focus on different valuable features of soft soil with different mixture to get better prediction of UCS. Developing more adaptable hybrid deep learning model is also our concern for the methodological improvement. Our method works with only 52 samples that will be increased in the future for the better prediction of UCS.

## 8. Conclusions

In this study, several machine learning methods were developed for the prediction of UCS of stabilised soft soil. The research not only provides several approaches for the same data set, but it also compares the model's achievement. The model's performance was assessed by estimating the RMSE and MAE. Based on the comparison of performance evaluation parameters, the Random Forest classification model structure outperforms other modelling approaches with this data set. As a result of MAE 0.0625, MSE 0.0625, RMSE 0.2500,  $r^2$  value 0.8942 and accuracy of 0.9375 were obtained for testing the data set using Random Forest model. Additionally, the sequential model got loss 0.2535, training accuracy of 0.9024, validation loss 0.4056 and validation accuracy 0.9091. We implemented each model in our data set with 52 samples for the prediction of three classes of stress. Our method used very relevant five input features (Column Reinforcement Type, Column Diameter, Column Penetration Ratio, Max\_Deviator Stress) of soft soil to predict stress of soft soil. In the future we have plans to use features and multiple target class to analyse and predict UCT of soft soil with polypropylene. The key research findings of our method are given as follows:

- (i) Developed an UCT method that quickly estimates the compressive strengths of soft soils with adequate cohesion to allow testing in an unconfined condition.
- (ii) Developed various machine learning algorithms and analysed those methods for prediction of UCS of stabilised soft soil with polypropylene.
- (iii) Validate robustness of the proposed method that is cost effective with less computational complexity.

In conclusion, the analysis' findings demonstrated that the suggested technique is novel and can be utilised to estimate the UCS value with a respectable level of accuracy while being less complicated and expensive than time-consuming lab work. In the future, we have plans to develop different features of soft soil and high-performance machine and deep learning model.

### Acknowledgments

The authors acknowledge Universiti Malaysia Pahang (UMP) for financing this research through the Research Grant Scheme, Project Number RDU 223309. The authors thank the staff of the Geotechnical Engineering Laboratory at UMP for providing facilities and cooperative behaviour during the tests.

### Author details

Md. Ikramul Hoque<sup>1</sup>  
E-mail: [ikramul3300@becm.kuet.ac.bd](mailto:ikramul3300@becm.kuet.ac.bd)

Muzamir Hasan<sup>1</sup>  
Md Shofiqul Islam<sup>2</sup>

Moustafa Houda<sup>3</sup>  
Mirvat Abdallah<sup>3</sup>

Md. Habibur Rahman Sobuz<sup>4</sup>

<sup>1</sup> Faculty of Civil Engineering Technology, Universiti Malaysia Pahang, Gambang, Malaysia.

<sup>2</sup> Faculty of Computing, Universiti Malaysia Pahang, Gambang, Malaysia.

<sup>3</sup> College of Engineering and Technology, American University of the Middle East, Egaila, Kuwait.

<sup>4</sup> Department of Building Engineering and Construction Management, Khulna University of Engineering & Technology, Khulna, Bangladesh.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Citation information

Cite this article as: Machine Learning Methods to Predict and Analyse Unconfined Compressive Strength of Stabilised Soft Soil with Polypropylene Columns, Md. Ikramul Hoque, Muzamir Hasan, Md Shofiqul Islam, Moustafa Houda, Mirvat Abdallah & Md. Habibur Rahman Sobuz, *Cogent Engineering* (2023), 10: 2220492.

### References

- Ali-Haghpanah, J. (2017). A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. *Proceedings of the 2017 Artificial Intelligence and Signal Processing Conference (AISP)*, Shiraz, Iran. <https://doi.org/10.1109/AISP.2017.8324083>
- Artur, M. (2021). "Review the performance of the Bernoulli Naive Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features", *Procedia computer science*, 190, 564–570. <https://doi.org/10.1016/j.procs.2021.06.066>

- Aslani, M., Nazariafshar, J., & Ganjian, N. (2019). "Experimental study on shear strength of cohesive soils reinforced with stone columns", *Geotechnical and Geological Engineering*, 37, 2165–2188. 3 <https://doi.org/10.1007/s10706-018-0752-z>
- Ayyappa, Y., Bekkanti, A., Krishna, A., Neelakanteswara, P., & Basha, C. Z. (2020, July). Enhanced and effective computerized multi layered perceptron based back propagation brain tumor detection with Gaussian filtering. *Proceedings of the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 58–62). IEEE. <https://ieeexplore.ieee.org/document/9182921>
- Black, J., Sivakumar, V., & Bell, A. (2011). "The settlement performance of stone column foundations", *Géotechnique*, 61(11), 909–922. <https://doi.org/10.1680/geot.9.P.014>
- Chien-Chung, C. (2010). A Passive-Aggressive Algorithm for Semi-supervised Learning. *Proceedings of the International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, Hsinchu, Taiwan. IEEE. <https://ieeexplore.ieee.org/abstract/document/5695474>
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Mach Learn*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Ellis, G., Yao, C., Zhao, R., & Penumadu, D. (1995). "Stress-strain modeling of sands using artificial neural networks", *Journal of Geotechnical Engineering*, 121, 429–435. 5 [https://doi.org/10.1061/\(ASCE\)0733-9410\(1995\)121:5\(429\)](https://doi.org/10.1061/(ASCE)0733-9410(1995)121:5(429))
- Guo, Q., LI, B., Chen, Y., Chen, G., & Chen, L. (2021). "Intelligent model for the compressibility prediction of cement-stabilized dredged soil based on random forest regression algorithm", *KSCE Journal of Civil Engineering*, 25, 3727–3736. 10 <https://doi.org/10.1007/s12205-021-2202-3>
- Guyon, I. (2000). "Linear discriminant and support vector", *Advances in Large Margin Classifiers*, 9, 147.
- Hao, J., Wang, F., Wang, X., Zhang, D., Bi, Y., Gao, Y., Zhao, X., & Zhang, Q. (2012). Development and optimization of baicalin-loaded solid lipid nanoparticles prepared by coacervation method using central composite design. *European Journal of Pharmaceutical Sciences*, 47(2), 497–505. <https://doi.org/10.1016/j.ejps.2012.07.006>
- Hasan, M. (2018). "THE undrained shear strength of soft clay reinforced with group encapsulated lime bottom ash columns", *International Journal of GEOMATE*, 14, 46–50. 46 <https://doi.org/10.21660/2018.46.45208>
- Kim, Y., Satyanaga, A., Rahardjo, H., Park, H., & Sham, A. W. L. (2021). Estimation of effective cohesion using artificial neural networks based on index soil properties: A Singapore case. *Engineering Geology*, 289, 106163. <https://doi.org/10.1016/j.enggeo.2021.106163>
- Lamorski, K., Pachepsky, Y., Sławiński, C., & Walczak, R. (2008). "Using support vector machines to develop pedotransfer functions for water retention of soils in Poland", *Soil Science Society of America Journal*, 72, 1243–1247. 5 <https://doi.org/10.2136/sssaj2007.0280N>
- Lee, C.-Y., & Chern, S.-G. (2013). Application of a support vector machine for liquefaction assessment. *Journal of Marine Science and Technology*, 21(3), 10. <https://doi.org/10.6119/JMST-012-0518-3>
- Li, Y., Rahardjo, H., Satyanaga, A., Rangarajan, S., & Lee, D. T. (2022). Soil database development with the application of machine learning methods in soil properties prediction. *Engineering Geology*, 306, 106769. <https://doi.org/10.1016/j.enggeo.2022.106769>
- Mahesh, P., & Deswal, S. (2010). "Modelling pile capacity using gaussian process regression", *Computers and Geotechnics*, 37, 942–947. 7–8 <https://doi.org/10.1016/j.compgeo.2010.07.012>
- Majdi, A., & Rezaei, M. (2013). "Prediction of unconfined compressive strength of rock surrounding a roadway using artificial neural network", *Neural Computing & Applications*, 23, 381–389. 2 <https://doi.org/10.1007/s00521-012-0925-2>
- Moreira, A. (2007). Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. *Proceedings of the International Conference on Computer Graphics Theory and Applications*, Spain. <https://doi.org/10.5220/0002080800610068>
- Pal, M., & Deswal, S. (2009). M5 model tree based modelling of reference evapotranspiration, *Hydrological Processes: An International Journal*, 23(10), 1437–1443. <https://doi.org/10.1002/hyp.7266>
- Pérez-Cruz, F., A, A.-R. J., & Giner, J. (2003). Estimating GARCH models using support vector machines\*. *Quantitative Finance*, 3(3), 163–172. <https://doi.org/10.1088/1469-7688/3/3/302>
- Robert, C. (2017). A Standardised Benchmark for Assessing the Performance of Fixed Radius Near Neighbours. *Proceedings of the European Conference on Parallel Processing*, France (pp. 311–321). LNEE, Springer. [https://link.springer.com/chapter/10.1007/978-3-319-58943-5\\_25](https://link.springer.com/chapter/10.1007/978-3-319-58943-5_25)
- Sabat, A. K. (2015). "Prediction of California bearing ratio of a stabilized expansive soil using artificial neural network and support vector machine", *Electronics Journal Geotechnical engineering*, 20(3), 981–991.
- Safavian, S. R., & Landgrebe, D. (1991). "A survey of decision tree classifier methodology", *IEEE Transactions on Systems, Man, and Cybernetics*, 21, 660–674. 3 <https://doi.org/10.1109/21.97458>
- Samui, P. (2008). "Support vector machine applied to settlement of shallow foundations on cohesionless soils", *Computers and Geotechnics*, 35, 419–427. 3 <https://doi.org/10.1016/j.compgeo.2007.06.014>
- Sathyapriya, S., Arumairaj, P., & Ranjini, D. (2017). "Prediction of unconfined compressive strength of a stabilised expansive clay soil using ANN and regression analysis (SPSS)", *Asian Journal of Research in Social Sciences and Humanities*, 7, 109–123. 2 <https://doi.org/10.5958/2249-7315.2017.00075.2>
- Shukla, S. K. (2022). Seven research mantras: A Short Guide for Researchers. *International Journal of Geosynthetics & Ground Engineering*, 8(6), 75. <https://doi.org/10.1007/s40891-022-00419-6>
- Solomatine, D. P., & Dulal, K. N. (2003). "Model trees as an alternative to neural networks in rainfall—run-off modelling", *Hydrological Sciences Journal*, 48, 399–411. 3 <https://doi.org/10.1623/hysj.48.3.399.45291>
- Solomatine, D. P., & Y, X. U. E. (2004). "M 5 model trees and neural networks: Application to flood forecasting in the upper reach of the huai river in China", *Journal of Hydrologic Engineering*, 9, 491–501. 6 [https://doi.org/10.1061/\(ASCE\)1084-0699\(2004\)9:6\(491\)](https://doi.org/10.1061/(ASCE)1084-0699(2004)9:6(491))
- Suthar, M. (2020). "Applying several machine learning approaches for prediction of unconfined compressive strength of stabilized pond ashes", *Neural Computing & Applications*, 32, 9019–9028. 13 <https://doi.org/10.1007/s00521-019-04411-6>
- Tharwat, A. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, 30, 169–190. <https://content.iospress.com/articles/ai-communications/aic729>
- Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). "Naïve Bayes", *Encyclopedia of Machine Learning*, 15(1), 713–714. <https://doi.org/10.1007/978-0-387-30164-8>

Wright, R. E., Grimm, L. G., & Yarnold, P. R. (1995). Logistic regression. *American Psychological Association*, 217–244.

Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2013). Linear discriminant analysis. *Robust data mining*, 27–33.

Zahmatkesh, A., & Choobbasti, A. J. (2010). “Settlement evaluation of soft clay reinforced by stone columns, considering the effect of soil compaction”,

*International Journal of Research and Reviews in Applied Sciences*, 3(2), 159–166. <https://doi.org/10.1007/s12517-010-0145-y>

Zareapoor, M., & Shamsolmoali, P. (2015). “Application of credit card fraud detection: Based on bagging ensemble classifier”, *Procedia computer science*, 48, 679–685. <https://doi.org/10.1016/j.procs.2015.04.201>