Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4832864, 15 pages
https://doi.org/10.1155/2021/4832864

Research Article

Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil

Quang Hung Nguyen,[1] **Hai-Bang Ly**,[2] **Lanh Si Ho**,[2,3] **Nadhir Al-Ansari**,[4] **Hiep Van Le**,[5] **Van Quan Tran**,[2] **Indra Prakash**,[6] **and Binh Thai Pham**[2]

[1]*Thuyloi University, Hanoi 100000, Vietnam*
[2]*University of Transport Technology, Hanoi 100000, Vietnam*
[3]*Civil and Environmental Engineering Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan*
[4]*Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden*
[5]*Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam*
[6]*Bhaskaracharya Institute for Space Applications and Geo-Informatics (BISAG), Gandhinagar 382002, India*

Correspondence should be addressed to Quang Hung Nguyen; hungwuhan@tlu.edu.vn, Hai-Bang Ly; banglh@utt.edu.vn, Nadhir Al-Ansari; nadhir.alansari@ltu.se, and Binh Thai Pham; binhpt@utt.edu.vn

Received 24 June 2020; Revised 17 December 2020; Accepted 27 January 2021; Published 8 February 2021

Academic Editor: Yu-Sheng Shen

The main objective of this study is to evaluate and compare the performance of different machine learning (ML) algorithms, namely, Artificial Neural Network (ANN), Extreme Learning Machine (ELM), and Boosting Trees (Boosted) algorithms, considering the influence of various training to testing ratios in predicting the soil shear strength, one of the most critical geotechnical engineering properties in civil engineering design and construction. For this aim, a database of 538 soil samples collected from the Long Phu 1 power plant project, Vietnam, was utilized to generate the datasets for the modeling process. Different ratios (i.e., 10/90, 20/80, 30/70, 40/60, 50/50, 60/40, 70/30, 80/20, and 90/10) were used to divide the datasets into the training and testing datasets for the performance assessment of models. Popular statistical indicators, such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (R), were employed to evaluate the predictive capability of the models under different training and testing ratios. Besides, Monte Carlo simulation was simultaneously carried out to evaluate the performance of the proposed models, taking into account the random sampling effect. The results showed that although all three ML models performed well, the ANN was the most accurate and statistically stable model after 1000 Monte Carlo simulations (Mean $R = 0.9348$) compared with other models such as Boosted (Mean $R = 0.9192$) and ELM (Mean $R = 0.8703$). Investigation on the performance of the models showed that the predictive capability of the ML models was greatly affected by the training/testing ratios, where the 70/30 one presented the best performance of the models. Concisely, the results presented herein showed an effective manner in selecting the appropriate ratios of datasets and the best ML model to predict the soil shear strength accurately, which would be helpful in the design and engineering phases of construction projects.

1. Introduction

Soil is a crucial material in civil engineering, as most of the structures are built on soil ground [1]. The failure of the ground and collapse of the buildings are often associated with soil shear strength. Under different loading conditions, the soil shear strength, or the shear resistance, is dependent on the cohesion, friction, and interlocking between particles [1]. The mechanical property of soil is complex due to the fact that soil often contains different particle sizes, high water content, and large voids [1]. Soil shear strength is dominated by basic parameters such as soil mineralogy, overburden pressure, water content, density, and void. Commonly, the soil shear strength is calculated by

determining the effective stress and soil parameters, such as internal friction angle and cohesion [1, 2]. These soil parameters can be determined in the field by Standard Penetration Test (SPT) or shear vane test and in the laboratory by conducting direct shear test, ring shear test, triaxial test, and unconfined compression [3, 4]. These tests are time-consuming and involve a lot of cost on conducting tests on an important number of samples.

Over the last decades, many researchers have tried to improve and find alternative methods to determine the shear strength of soil [3, 5–10]. Nam et al. [11] used a multistage direct shear test for determining the shear strength of unsaturated and saturated soils. Such a method could reduce some disadvantages of conventional direct shear tests and produced high accuracy results. Besides, many researchers have attempted to establish a relationship between soil indexes, such as clay fraction, liquid limit, plastic limit, and clay mineralogy [9, 12]. Also, many efforts have been made to evaluate the shear strength of soil through other soil parameters, such as establishing a correlation between suction and shear strength [10, 13]. In addition, several conventional procedures were introduced to estimate the shear strength of soil, where the relationship between the water content and suction is employed as a tool in the prediction process of unsaturated soil shear strength [6, 14–16]. Another effort has been carried out to estimate the soil shear strength in situ through shear wave velocity [16–18]. Overall, the conventional and traditional techniques possess some disadvantages and limitations, such as limitations in using basic soil parameters or considering a small range of soils. As an example, Kaya [2] indicated that the empirical formula, as suggested by Wright [19], is only limited to the soil containing a clay fraction superior to 50%.

In the recent time, Machine Learning (ML) techniques have been developed expeditiously and successfully applied in many fields of civil engineering [20–27] and Earth sciences [28–31], including geotechnical engineering such as landslide susceptibility [32–41] and estimation of soil parameters [42–47] including shear strength of soil [47–52]. In the work of Das et al. [53], the authors successfully applied an Artificial Neural Network (ANN) for estimating the residual friction angle of tropical soil in a specified area. Besides, it is found that the Support Vector Machine (SVM) showed a better performance than ANN for estimating the shear strength of soil using basic soil parameters, such as liquid limit, plastic limit, and clay fraction. In another work, Besalatpour et al. [54] showed that Adaptive-Network-based Fuzzy Inference System (ANFIS) and ANN models had higher ability than conventional regression methods. In another study, three new optimization techniques, namely, the Dragonfly Algorithm (DA), Invasive Weed Optimization (IWO), and Whale Optimization Algorithm (WOA), were employed to optimize the weights and biases of an ANN structure in estimating the shear strength of soil [50], where it was noticed that the learning error was significantly decreased. Thus, the IWO-ANN hybrid algorithm was found to be promising model instead of conventional methods in solving soil shear strength problems. Further, Moayedi et al. [49] used four neural-metaheuristic models for estimating the shear strength of soil and stated that the Salp Swarm Algorithm-Multilayer Perceptron (SSA-MLP) model is a potential alternative method for estimating the soil shear strength. In general, ML techniques have significantly improved the prediction ability compared to conventional methods.

Despite significant growing of researches in applying ML algorithms in soil science, it is surprising how few of these suggestions are dedicated to the investigation of the performance assessment under a combination of factors during the model development phase. These factors could be the choice of data splitting, the selection of sampling technique, or the ML algorithm. For instance, a study on the comparison of ML techniques in digital soil mapping found that sample design and model choice significantly affected the outputs [55]. With regard to the data splitting, the data sample is often divided into two datasets, including a training set for model training and a testing set for model validation. Many researchers proposed a ratio of 70/30 or 80/20 (training/testing set) for producing datasets in landslide susceptibility problems [56–61]. Regarding studies on estimating the residual strength of soil using ML algorithms, previous works mainly used ratios of 70/30, 80/20, and 90/10 (training/testing) for generating datasets [22, 43, 47–49, 51–53]. Recently, Pham et al. [47] conducted a study on estimating the shear strength of soil in varying the training dataset size from 30% to 90% using the Random Forest (RF) algorithm. The study revealed that the increase in the size of the training dataset improved the training performance and made the model more stable. For the testing performance, the increase in the training set's size from 30% to 80% could also enhance the testing performance. However, when training size increased from 80% to 90%, the opposite trend was found in testing performance. In general, the training set size had an important effect on the prediction ability of the ML models [62].

The main objective of the present study is to evaluate the performance of ML models considering different ratios of soil data splitting for the prediction of soil shear strength. In this research, three ML techniques, namely, ANN, Extreme Learning Machine (ELM), and Boosting algorithm, were adopted to estimate the soil shear strength based on different splitting ratios of input data for the training and testing phases. The main difference of this study compared with the previously published works is that it is the first time the influence of splitting strategy of training and testing datasets used in ML models was investigated to predict the soil shear strength. Results were evaluated using standard statistical measures, namely Mean Absolute Error (MAE), Correlation Coefficient (R), and Root Mean Squared Error (RMSE), for the selection of the best model in predicting the soil shear strength and study the influence of different ratios of training and testing data on the performance of models.

2. Research Significance

ML, which includes advanced soft computing based techniques, has been developed and applied successfully and efficiently to solve a lot of real-world problems [63–68].

The main advantage of ML is that it can subjectively analyze unlimited amounts of data and give reliable outcomes and assessment [69]. However, its performance depends significantly on the quality of data and the strategy of using the data [70–72]. Therefore, assessment of the influence of data splitting on ML models' performance has a high significance, which will pave the way on how to select a suitable data splitting for better ML-based modeling. In this study, we have selected three popular ML models, namely, ANN, ELM, and Boosted, for modeling. In addition, we have selected a research problem, "the prediction of soil shear strength," which is an important geotechnical engineering task [43, 46, 47, 73]. This will help the construction engineers and managers to quickly and accurately predict the soil shear strength, which can be used for the design and verification of construction projects.

3. Data Used

Soil investigation data of the Long Phu 1 power plant project, located in Soc Trang province, Vietnam (longitude of 9°59′07.3″N and latitude of 106°04′48.0″E), was used in this study for the development of the ML models. The construction of this power plant was started in June 2015, reflecting a key project under the Vietnamese Government's 2011–2020 National Power Development Plan [73]. A database of 538 soil samples was used to build the training and testing data sets. Soil parameters such as clay content (%), void ratio, moisture content (%), liquid limit (%), plastic limit (%), and specific gravity were used as input variables, whereas the soil shear strength (kg/cm^2) determined by direct shear test under the Undrain and Unconsolidated (UU) scheme was used as the output variable.

Statistical analysis of the input variables suggests that, in the samples, the clay content varied from 0 to 65 (%), plastic limit from 15 to 35 (%), liquid limit from 20 to 65 (%), specific gravity from 2.6 to 2.7, and void ratio from 0.5 to 1.0 (Figures 1(a)–1(g)), whereas the output variable varied from 0.45 to 0.7 (kg/cm^2) (Figure 1(g)).

Considering different ranges of variables (Figure 1), these values were scaled in the range of [0, 1] to avoid unexpected jumps and reduce fluctuations within the datasets used for modeling.

4. Methods Used

4.1. Artificial Neural Network (ANN). ANN has been known as a popular and powerful machine learning technique (computational model) [74, 75], based on structures and functions of biological neural networks: the nervous system of the human brain [20, 76–78]. This method has been used successfully in solving a wide range of civil engineering problems, including geotechnical engineering problems. ANN method is used to identify the relationship between input and output neurons in both linear and nonlinear patterns [21, 22, 79]. Thus, ANN could make a decision by analyzing patterns and relationships in data by itself [2, 43, 80]. In this study, a multilayered perceptron neural network, a popular ANN [81], was employed as a regression technique to estimate the soil shear strength.

4.2. Boosting Trees (Boosted). Boosted (Trees) is a hybrid method that combines the decision trees and boosting method. In this ensemble-type method, decision trees are employed to link input and output variables through recursive dual separations, while the boosting method is adopted to associate many individual models for improving the performance of the hybrid model [82]. The Boosted method, having the merits of tree-based techniques, can overcome the disadvantages of a sole tree model because of the following reasons. Firstly, this ensemble can choose a proper variable to match the appropriate functions. Secondly, it is suitable for various types of data using random boosting, and finally, this method can mitigate both bias and variance via model averaging [83].

4.3. Extreme Learning Machine (ELM). ELM was firstly suggested by Huang et al. [84, 85], which is a modern algorithm and employed as a Single hidden Layer Feedforward Neuron Network (SLFN) [86]. ELM algorithm produces better performance in terms of learning speed compared to a conventional algorithm, for instance, backpropagation and least-square support vector machine [61, 84, 87]. The main aim of ELM is to get the smallest norm of weights on which the smallest training error can be reached for optimization of the model performance. A detailed description of ELM algorithm is available in published papers [84, 88–90].

4.4. Monte Carlo Approach. Monte Carlo method has been widely introduced to solve problems relating to the variability of input parameters in various fields, including geotechnical engineering [45, 91, 92]. Monte Carlo methods are a broad class of computational algorithms that rely on the repeated random sampling process to obtain numerical results. Basically, this technique could produce a high ability to compute, statistically, the relationship in data for both linear and nonlinear problems [45, 91]. Monte Carlo technique is implemented by repeating randomly input variables based on the distribution of probability density, and the outputs are computed correspondingly via a simulated model [93, 94]. A concept of the Monte Carlo method includes the following: (i) variability of input parameter could be completely spread by predetermined models and (ii) sensitivity analysis of inputs can be evaluated using statistical analysis of the output results.

4.5. Performance Evaluation Criteria. In this paper, standard statistical measures, namely, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (R), were used to compare and validate the performance of ML models [47, 95]. In general, RMSE is the mean squared difference between the estimated and actual values, while MAE is the mean amplitude of errors. Lower values of RMSE and MAE mean higher prediction ability of the models. Besides, R is employed to evaluate the

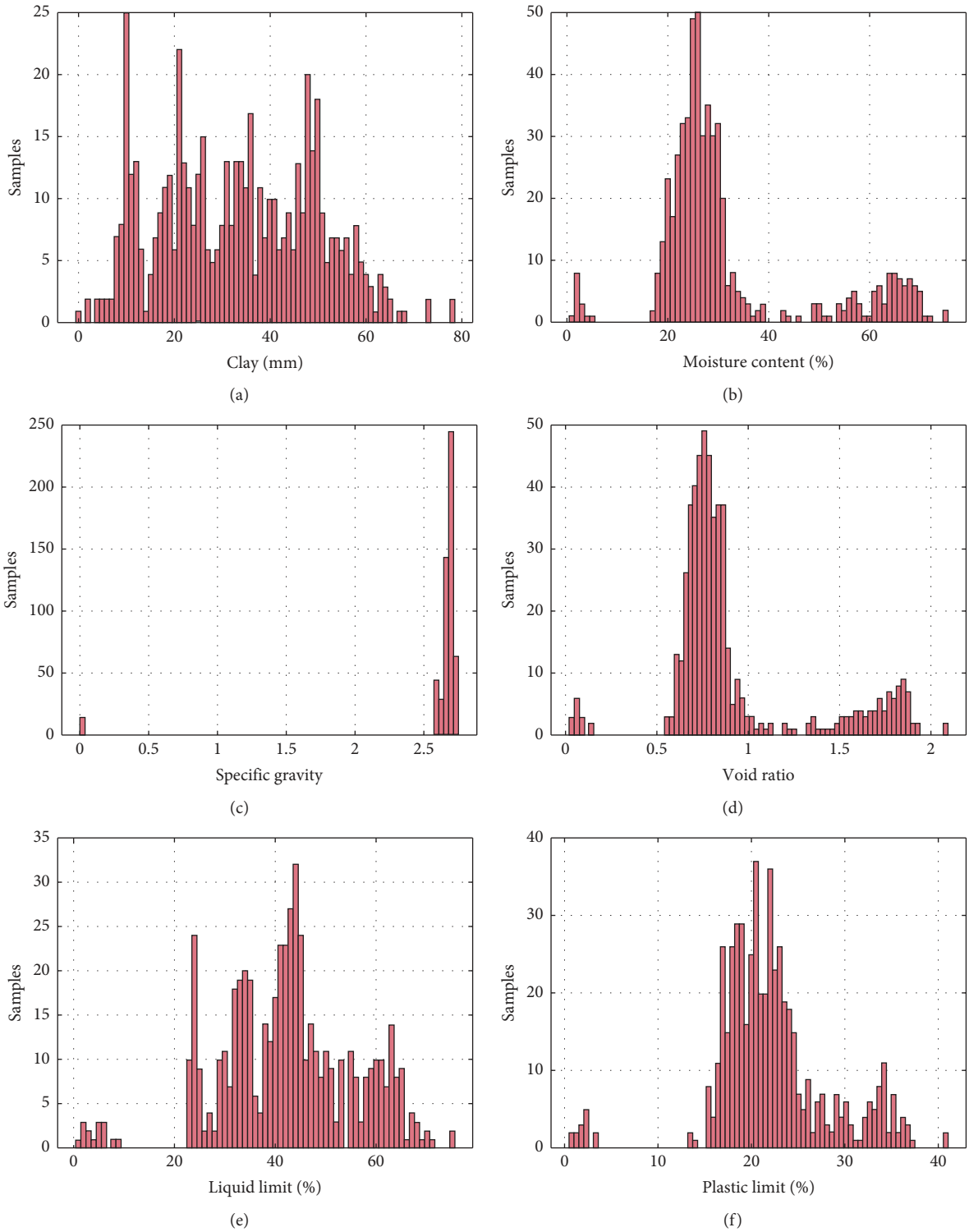
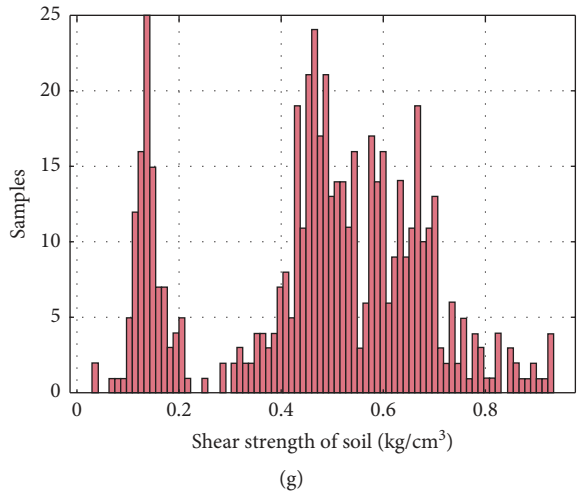


Figure 1: Continued.



(g)

Figure 1: Histograms of the parameters used in this study: (a) clay; (b) moisture content; (c) specific gravity; (d) void ratio; (e) liquid limit; (f) plastic limit; (g) shear strength of soil.

correlation of the predicted and actual values of soil shear strength. The values of R are between −1 and +1, where the absolute values of R close to 1 mean higher prediction ability. These indicators can be computed using the following formulas [45, 96]:

$$R = \frac{\sum_{i=1}^{n} \left(y_{coi} - \overline{y_{co}}\right)\left(y_{aci} - \overline{y_{ac}}\right)}{\sqrt{\sum_{i=1}^{n} \left(y_{coi} - \overline{y_{co}}\right)^2 \left(y_{aci} - \overline{y_{ac}}\right)^2}},$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_{coi} - y_{aci}\right)^2}, \tag{1}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left|y_{coi} - y_{aci}\right|,$$

where y_{coi} and $\overline{y_{co}}$ represent the output value of the ith sample and the corresponding output mean value computed by the ML model, respectively; y_{aci} and $\overline{y_{ac}}$ denote the measured value of the ith sample and the measured mean value, respectively; and n indicates the total number of samples.

5. Results and Analysis

In this section, the prediction results of the soil shear strength are presented using various ML models (ANN, ELM, and Boosted). In the modeling, clay content, void ratio, moisture content, liquid limit, plastic limit, and specific gravity were considered as input variables, whereas soil shear strength was considered as the output variable. As a first step, the influence of training and testing ratio on the performance of the ML models is presented, followed by the study of the random sampling effects on the performance of ML models, and finally, comparisons of different ML models are performed.

5.1. Influence of Different Training and Testing Ratios on the Performance of the ML Models. To evaluate the influence of different ratios on the performance of ML models, ANN model was used to select the best train-to-test ratio for the estimation of soil shear strength. Using ANN to perform the study, six parameters (Table 1) were selected using trial and error tests to train the model. The dataset was divided into two parts, with different ratios: 10 : 90, 20 : 80, 30 : 70, 40 : 60, 50 : 50, 60 : 40, 70 : 30, 80 : 20, and 90 : 10 train/test split. Basically, a training dataset was used to construct the model, whereas the testing dataset was used to assess the model's predictive capability. Finally, the performance of ANN model on different ratio-based training and testing datasets using various statistical indices was evaluated, as shown in Figure 2.

It can be seen that as the number of data in the training datasets increased, the errors (RMSE and MAE) of the ANN model increased, and R values of the ANN model decreased, showing the accuracy of ANN decreased (Figures 2(a), 2(c), and 2(e)). In contrast, as the number of data in the testing datasets increased, the errors (RMSE and MAE) of ANN decreased, and R values increased, reflecting an increase of the ANN accuracy (Figures 2(b), 2(d), and 2(f)). It can be observed that the performance of the ANN model on both training and testing datasets was the best on the training/testing ratio of 70/30, based on the values of mean, standard deviation, and quantile levels of the three criteria.

5.2. Random Sampling Effects on the Performance of ANN. To validate the random sampling effects on the performance of the ML models, the ANN model was used and trained on different training/testing ratios using Monte Carlo simulation. In this process, the 1000 simulation was carried out to validate the statistical convergence of the model, as shown in Figure 3. It can be seen that RMSE and MAE values were stable at 10% of the average values with only 10 iterations,

Table 1: Parameters of the ANN algorithm used in this study.

No.	Parameters	Setting
1	Number of hidden layers	1
2	Number of neurons in the hidden layer	8
3	Activation function for the hidden layer	Sigmoid
4	Activation function for the output layer	Linear
5	Training algorithm	Levenberg-Marquardt
6	Cost function	MSE

whereas these values were stable at 5% average from 20 Monte Carlo iterations. Besides, the values of R were statistically stable at 2% average with 8 iterations and at 1% average from 50 iterations.

In addition, the analysis of the probability density of R, RMSE, and MAE values was also carried out to study the random sampling effects on the performance of ANN model (Figure 4). It can be observed that the distribution of the probability density of R, RMSE, and MAE values was different on various training/testing ratios.

In general, it can be stated that the performance of the ANN model is sensitive to the random selection of data in the datasets used for training and validating the model. In this study, the ANN model was converged with above 700 Monte Carlo simulations, and the train-to-test ratio of 70 : 30 was found as the best option for ML modeling.

5.3. Validation and Comparison of Different ML Models. Validation and comparison of three ML models (i.e., ANN, ELM, and Boosted) were conducted using the best ratio of 70/30 of training and testing datasets. The ANN was trained with the parameters provided in Table 1, whereas ELM was trained with the network constructed by one input layer (6 neurons), one hidden layer (8 neurons), and one output (1 neuron). Regarding Boosted algorithm, the minimum leaf size was taken as 8, the number of learning cycles was 20, and the learning rate was set at 0.1. Values of R, RMSE, and MAE of the models using the testing dataset are shown in Figures 4–6. On the basis of RMSE indicator, it can be observed that the range of RMSE of ANN model was from about 0.05 to 0.1, whereas this value ranged from about 0.08 to 0.125 for Boosted algorithm and from 0.07 to 0.3 for ELM model over 1000 Monte Carlo simulations (Figure 5). Regarding MAE indicator, it can be seen that the range of MAE of ANN model was from 0.04 to 0.07, whereas this value ranged from 0.06 to 0.09 for the Boosted model and from 0.075 to 0.25 for ELM model over 1000 Monte Carlo simulations (Figure 6). In terms of R indicator, ANN model had the R values ranging from 0.95 to 0.97, from 0.88 to 0.95 for Boosted model, and from 0.62 to 0.95 for ELM model (Figure 7). Based on these results, it can be generally seen that the ANN model got the lowest error values (RMSE and MAE) and highest R values compared with other models (Boosted and ELM), whereas the EML got the most unstable values of RMSE, MAE, and R. ELM also got the highest values of errors and lowest values of R over 1000 Monte Carlo simulations. A summary of the main results of the three methods is presented in Table 2. Overall, it can be stated that the ANN model is the best and most stable model compared with other models (Boosted and ELM) for the prediction of soil shear strength.

6. Discussion

ML models are known as advanced techniques and approaches for quick and accurate prediction of real-world problems. These models, based on the objective computational algorithms, can handle complex relationships between input and output variables [97]. However, it is observed that ML models are quite sensitive to the quality of data and the way they are used in the modeling process, especially the ratio used to divide the datasets for training and validating the ML models [98]. In this study, this problem is analyzed by investigating the influence of training/testing ratio on the performance of three different popular ML models, namely, ANN, EML, and Boosted, to predict the soil shear strength.

Overall, the results showed that the ML models' performance was significantly changed under different training/testing ratios. The results showed that the training/testing ratio of 70/30 was the most suitable one for training and validating the ML models. This finding is in line with other published works, such as Pham et al. [99], who investigated different training/testing ratios for training and validating various ML models (SVM, Logistic Regression, ANN, and Naıve Bayes) for spatial prediction of landslides and proved that 70/30 was the best training/testing ratio for getting the best performance of the ML models. Other studies and researches also confirmed the finding of this study [100–105]. In addition, it is noticed that when the percentage of data in the training dataset increased, the errors (RMSE and MAE) of the models increased, and R values decreased. Thus, an increase of data (or samples) in the training dataset might have a negative influence on the prediction accuracy and difficulty in applying the models.

Besides, the validation and comparison results showed that all the ML models performed well, but ANN was the best model for the prediction of soil shear strength. It can be stated that ANN model has been reaffirmed as the best single ML model for solving most of the real-word problems [106, 107]. ANN has several advantages compared with other ML models, such as (i) capable of extracting the essential process information from data for analyzing and prediction, (ii) an ability of generalization of data, (iii) able to correctly process information that only broadly resembles the original training data, and (iv) its essential features being related to nonlinearity, fault tolerance, independent assumptions, and universality. Thus, ANN algorithm is particularly reasonable for extremely complex data. Last but not least, ANN is an

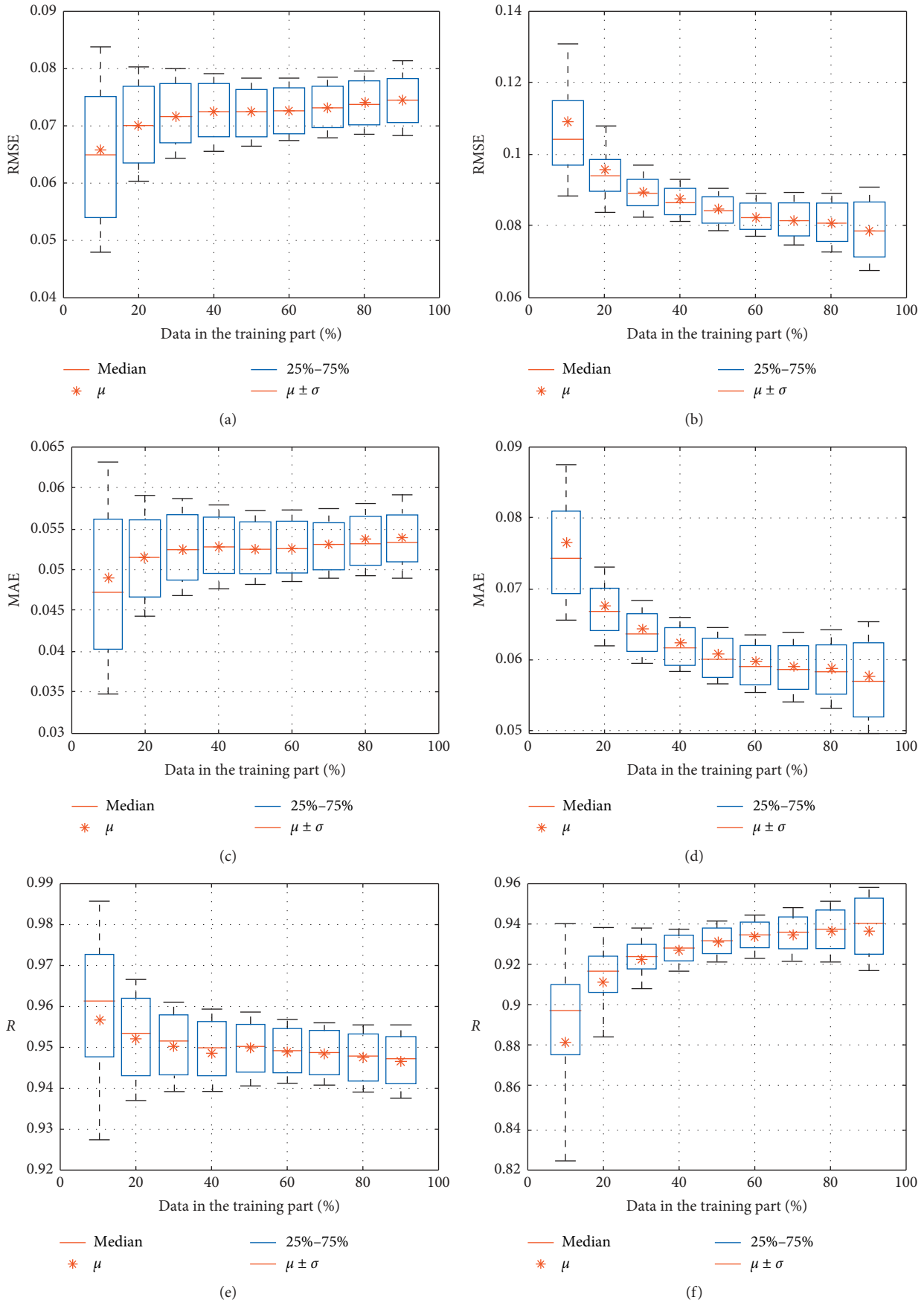


Figure 2: Validation of the ANN model's performance under different ratio (percentage) of data in the training part: (a) RMSE for the training part; (b) RMSE for the testing part; (c) MAE for the training part; (d) MAE for the testing part; (e) R for the training part; and (f) R for the testing part.

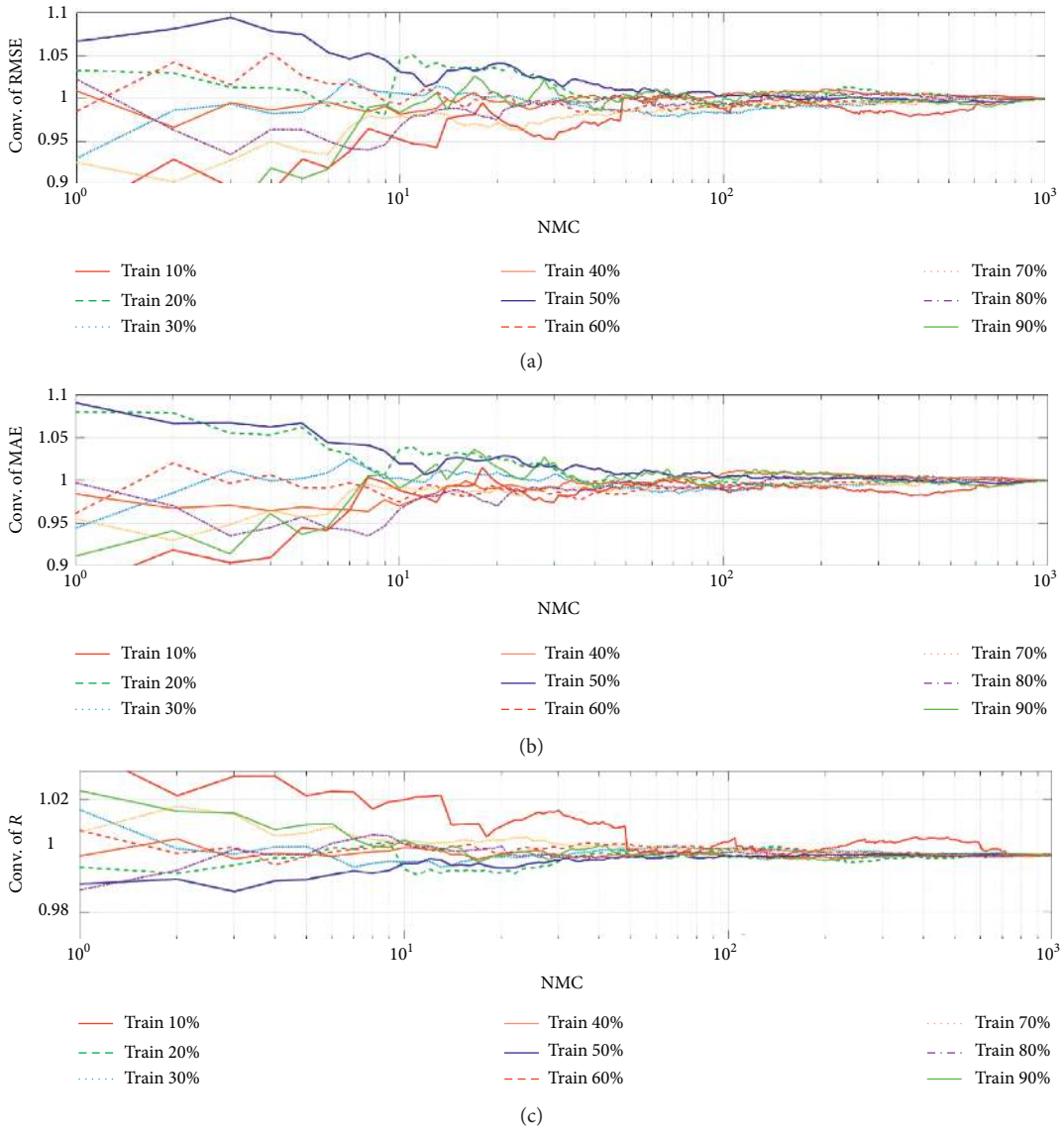


Figure 3: Statistical convergence results for 1000 Monte Carlo simulations for the testing part: (a) RMSE; (b) MAE; and (c) R.

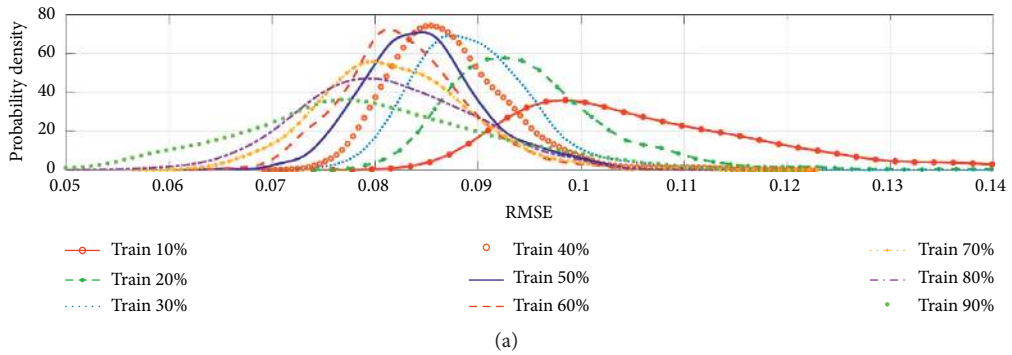


Figure 4: Continued.

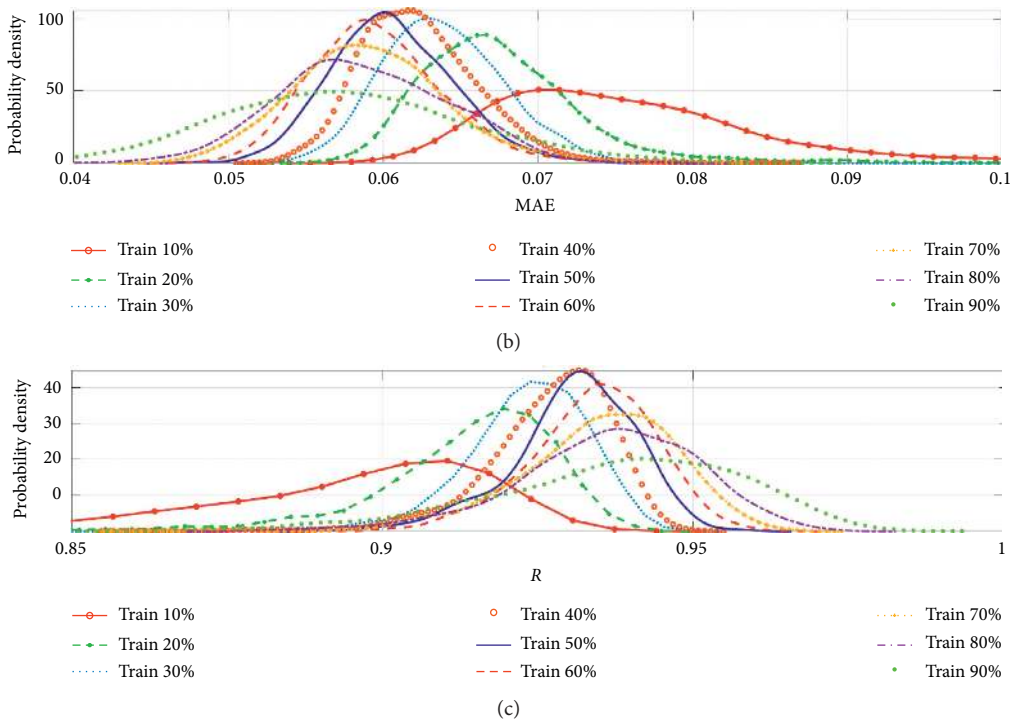


Figure 4: Probability density results for 1000 Monte Carlo simulations for the testing part with different indicators: (a) RMSE; (b) MAE; and (c) R. It should be noticed that the legends, for instance, Train 10%, indicates that the results obtained with 10% of the total data were used to construct the training dataset.

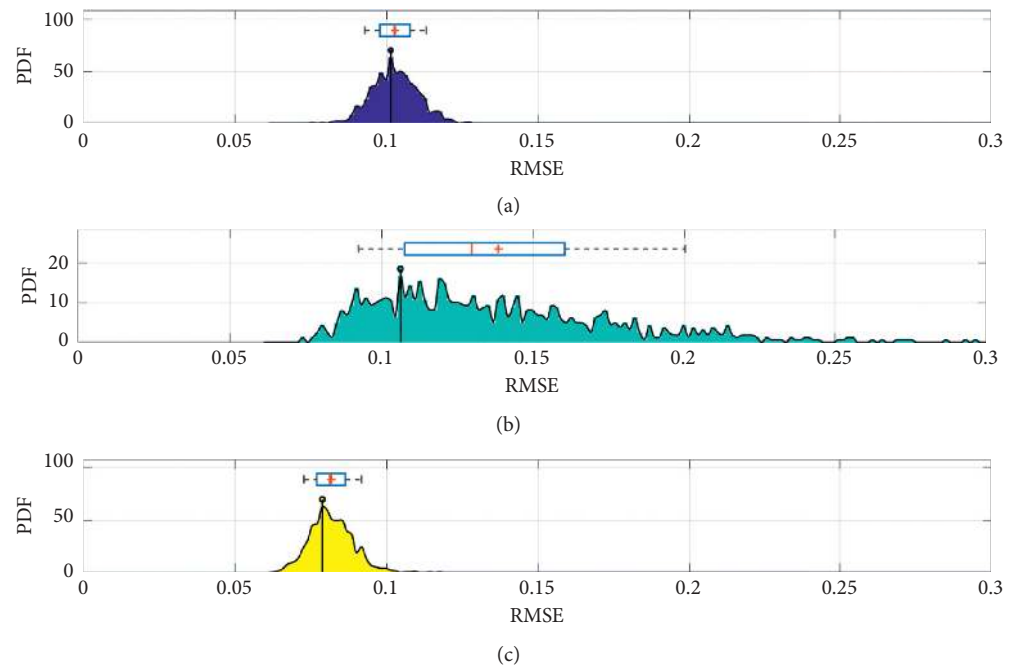


Figure 5: Comparison of Boosted, ELM, and ANN in terms of probability density results for 1000 Monte Carlo simulations for the testing part in terms of RMSE.

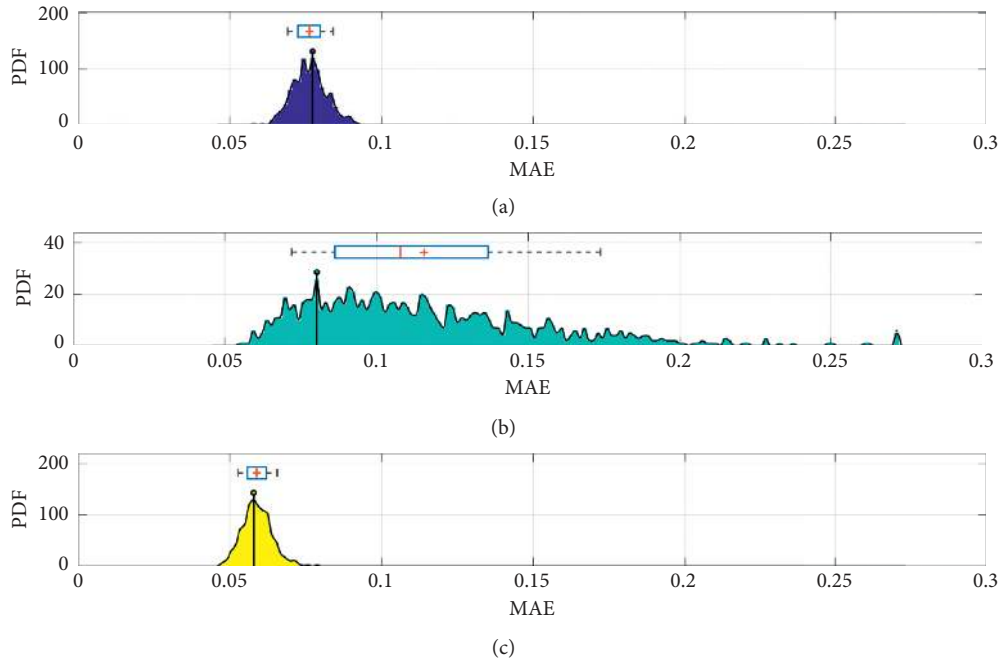


Figure 6: Comparison of Boosted, ELM, and ANN in terms of probability density results for 1000 Monte Carlo simulations for the testing part in terms of MAE.

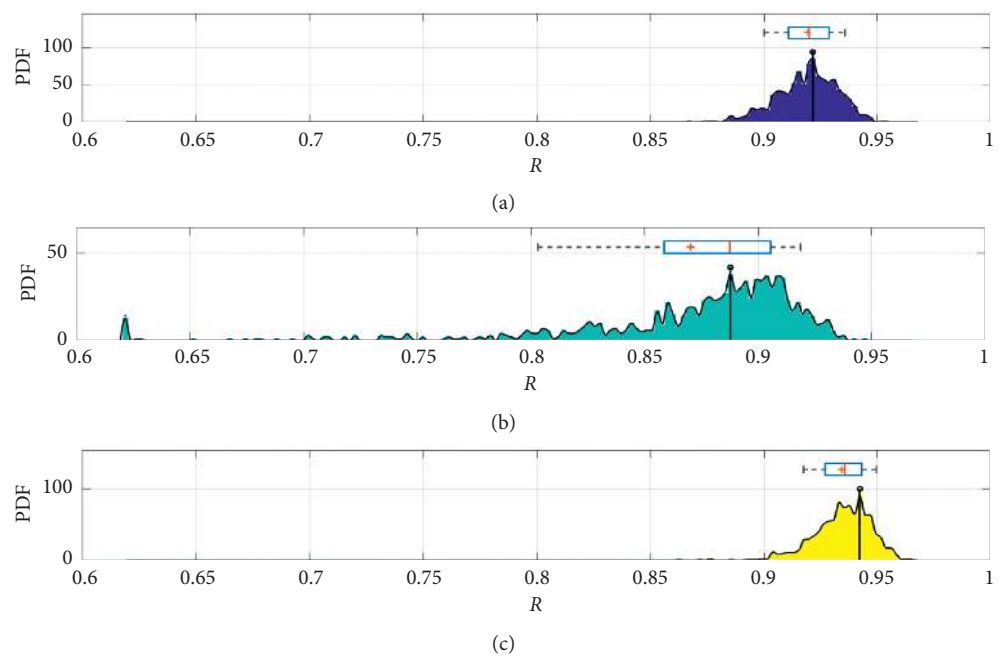


Figure 7: Comparison of Boosted, ELM, and ANN in terms of probability density results for 1000 Monte Carlo simulations for the testing part in terms of R.

Table 2: Summary of the results obtained over 1000 Monte Carlo simulations for ANN, Boosted, and ELM algorithms in this study.

Criteria	ANN	Boosted	ELM
Mean R	0.9348	0.9192	0.8703
Std. R	0.0129	0.0134	0.0624
Mean RMSE	0.0820	0.1029	0.1386
Std. RMSE	0.0073	0.0076	0.0412
Mean MAE	0.0591	0.0763	0.1157
Std. MAE	0.0049	0.0056	0.0388

adaptive algorithm, so that the learning process can be more effective [108, 109]. Therefore, it can be stated that the ANN was the best predictor for the prediction of soil shear strength.

7. Conclusions

Soil shear strength is one of the most critical geotechnical engineering properties used for designing and constructing civil engineering structures and constructions. Prediction of this parameter using advanced ML models might help in saving time and reducing cost for construction projects. In this study, three popular ML models, including ANN, ELM, and Boosted, were applied and compared to predict the soil shear strength using a database collected from Long Phu 1 power plant project, Vietnam. In addition, the performance of these models was also investigated under the influence of different training and testing ratios over 1000 Monte Carlo simulations.

Validation and comparison results showed that even the performance of all models was good and the performance of ANN was the best compared with other models. It can also be observed that the performance of the models was significantly changed under the different training and testing ratios used for training and validating the models. Based on the statistical analysis, a ratio of 70/30 for training and testing datasets was considered as the best ratio for training and validating the models. In addition, Monte Carlo simulations showed that the performance of the models is different under the random sampling effect over 1000 simulations. ANN was found as the best and most stable method under the variability of the input space.

In short, civil engineers can use the results of this study for quick and accurate prediction of soil shear strength for designing purposes, for instance, road, bridges, retaining walls, and other geotechnical and civil structures. Although the one group of data used in this study is sufficient for the development of the ML models, it is recommended that these ML models should be applied and validated with various data in different regions for better justification and verification. However, it is noticed that these applied models are considered as black-box models and do not provide the equations for engineer's calculation; therefore, other ML models like GEP, GMDH, and EPR, which can provide the equations, can be considered for future application and comparison.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Ministry of Transport, project titled "Building Big Data and Development of Machine Learning Models Integrated with Optimization Techniques for Prediction of Soil Shear Strength Parameters for Construction of Transportation Projects" under Grant no. DT 203029.

References

[1] B. M. Das and K. Sobhan, *Principles of Geotechnical Engineering*, Cengage Learning, Boston, MA, USA, 2013.

[2] A. Kaya, "Residual and fully softened strength evaluation of soils using artificial neural networks," *Geotechnical and Geological Engineering*, vol. 27, no. 2, pp. 281–288, 2009.

[3] H. Hettiarachchi and T. Brown, "Use of SPT blow counts to estimate shear strength properties of soils: energy balance approach," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 135, no. 6, pp. 830–834, 2009.

[4] H. Motaghedi and A. Eslami, "Analytical approach for determination of soil shear strength parameters from CPT and CPTu data," *Arabian Journal for Science and Engineering*, vol. 39, no. 6, pp. 4363–4376, 2014.

[5] M. Cha and G.-C. Cho, "Shear strength estimation of sandy soils using shear wave velocity," *Geotechnical Testing Journal*, vol. 30, no. 6, pp. 484–495, 2007.

[6] E. A. Garven and S. K. Vanapalli, "Evaluation of empirical procedures for predicting the shear strength of unsaturated soils," in *Proceedings of the Fourth International Conference on Unsaturated Soils*, pp. 2570–2592, Carefree, AZ, USA, April 2006.

[7] B.-S. Kim, S. Shibuya, S.-W. Park, and S. Kato, "Application of suction stress for estimating unsaturated shear strength of soils using direct shear testing under low confining pressure," *Canadian Geotechnical Journal*, vol. 47, no. 9, pp. 955–970, 2010.

[8] J. O. Ohu, G. S. V. Raghavan, E. McKyes, and G. Mehuys, "Shear strength prediction of compacted soils with varying added organic matter contents," *Transactions of the ASAE*, vol. 29, no. 2, pp. 351–355, 1986.

[9] B. Tiwari and H. Marui, "A new method for the correlation of residual shear strength of the soil with mineralogical composition," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 131, no. 9, pp. 1139–1150, 2005.

[10] O. M. Vilar, "A simplified procedure to estimate the shear strength envelope of unsaturated soils," *Canadian Geotechnical Journal*, vol. 43, no. 10, pp. 1088–1095, 2006.

[11] S. Nam, M. Gutierrez, P. Diplas, and J. Petrie, "Determination of the shear strength of unsaturated soils using the multistage direct shear test," *Engineering Geology*, vol. 122, no. 3-4, p. 272, 2011.

[12] A. W. Skempton, "Residual strength of clays in landslides, folded strata and the laboratory," *Géotechnique*, vol. 35, no. 1, pp. 3–18, 1985.

[13] D. W. Rassam and D. J. Williams, "A relationship describing the shear strength of unsaturated soils," *Canadian Geotechnical Journal*, vol. 36, no. 2, pp. 363–368, 1999.

[14] M. A. Tekinsoy, C. Kayadelen, M. S. Keskin, and M. Söylemez, "An equation for predicting shear strength envelope with respect to matric suction," *Computers and Geotechnics*, vol. 31, no. 7, pp. 589–593, 2004.

[15] Y. F. Xu, "Fractal approach to unsaturated shear strength," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 130, no. 3, pp. 264–273, 2004.

[16] Y. F. Xu and D. A. Sun, "A fractal model for soil pores and its application to determination of water permeability," *Physica A: Statistical Mechanics and Its Applications*, vol. 316, no. 1–4, pp. 56–64, 2002.

[17] C. R. McGann, B. A. Bradley, M. L. Taylor, L. M. Wotherspoon, and M. Cubrinovski, "Development of an empirical correlation for predicting shear wave velocity of Christchurch soils from cone penetration test data," *Soil Dynamics and Earthquake Engineering*, vol. 75, pp. 66–75, 2015.

[18] M. S. Nam and C. Vipulanandan, "Roughness and unit side resistances of drilled shafts socketed in clay shale and limestone," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 134, no. 9, pp. 1272–1279, 2008.

[19] S. G. Wright, *Evaluation of Soil Shear Strengths for Slope and Retaining Wall Stability Analyses with Emphasis on High Plasticity Clays*, Federal Highway Administration, Washington, DC, USA, 2005.

[20] H.-B. Ly, T.-T. Le, H.-L. T. Vu, V. Q. Tran, L. M. Le, and B. T. Pham, "Computational hybrid machine learning based prediction of shear capacity for steel fiber reinforced concrete beams," *Sustainability*, vol. 12, no. 7, p. 2709, 2020.

[21] D. V. Dao, H.-B. Ly, H.-L. T. Vu, T.-T. Le, and B. T. Pham, "Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete," *Materials*, vol. 13, no. 5, p. 1072, 2020.

[22] D. V. Dao, H. Adeli, H.-B. Ly et al., "A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation," *Sustainability*, vol. 12, no. 3, p. 830, 2020.

[23] T.-T. Le, B. T. Pham, H.-B. Ly, A. Shirzadi, and L. M. Le, "Development of 48-hour precipitation forecasting model using nonlinear autoregressive neural network," in *CIGOS 2019, Innovation for Sustainable Infrastructure*, pp. 1191–1196, Springer, Berlin, Germany, 2020.

[24] H.-B. Ly, L. M. Le, L. V. Phi et al., "Development of an AI model to measure traffic air pollution from multisensor and weather data," *Sensors*, vol. 19, no. 22, p. 4941, 2019.

[25] H.-B. Ly, B. T. Pham, D. V. Dao, V. M. Le, L. M. Le, and T.-T. Le, "Improvement of ANFIS model for prediction of compressive strength of manufactured sand concrete," *Applied Sciences*, vol. 9, no. 18, p. 3841, 2019.

[26] H.-B. Ly, T.-T. Le, L. M. Le et al., "Development of hybrid machine learning models for predicting the critical buckling load of I-shaped cellular beams," *Applied Sciences*, vol. 9, no. 24, p. 5458, 2019.

[27] B. T. Pham, L. M. Le, T.-T. Le et al., "Development of advanced artificial intelligence models for daily rainfall prediction," *Atmospheric Research*, vol. 237, Article ID 104845, 2020.

[28] W. Chen, Y. Li, W. Xue et al., "Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods," *Science of the Total Environment*, vol. 701, Article ID 134979, 2020.

[29] V.-H. Nhu, A. Mohammadi, H. Shahabi et al., "Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment," *International Journal of Environmental Research and Public Health*, vol. 17, no. 14, p. 4933, 2020.

[30] W. Chen, B. Pradhan, S. Li et al., "Novel hybrid integration approach of bagging-based Fisher's linear discriminant function for groundwater potential analysis," *Natural Resources Research*, vol. 28, no. 4, pp. 1239–1258, 2019.

[31] Y. Wang, H. Hong, W. Chen et al., "Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm," *Journal of Environmental Management*, vol. 247, pp. 712–729, 2019.

[32] M. Abedini, B. Ghasemian, A. Shirzadi et al., "A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment," *Geocarto International*, vol. 34, no. 13, pp. 1427–1457, 2019.

[33] Pham, Shirzadi, Shahabi et al., "Landslide susceptibility assessment by novel hybrid machine learning algorithms," *Sustainability*, vol. 11, no. 16, p. 4386, 2019.

[34] Nguyen, Tuyen, Shirzadi et al., "Development of a novel hybrid intelligence approach for landslide spatial prediction," *Applied Sciences*, vol. 9, no. 14, p. 2824, 2019.

[35] B. T. Pham, I. Prakash, K. Khosravi et al., "A comparison of support vector machines and Bayesian algorithms for landslide susceptibility modelling," *Geocarto International*, vol. 34, no. 13, pp. 1385–1407, 2019.

[36] B. T. Pham and I. Prakash, "A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 3, pp. 1911–1925, 2019.

[37] T. V. Phong, T. T. Phan, I. Prakash et al., "Landslide susceptibility modeling using different artificial intelligence methods: a case study at Muong Lay district, Vietnam," *Geocarto International*, pp. 1–24, 2019.

[38] B. T. Pham, T. V. Phong, T. Nguyen-Thoi et al., "Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers," *Geocarto International*, pp. 1–23, 2020.

[39] V.-H. Nhu, A. Mohammadi, H. Shahabi et al., "Landslide detection and susceptibility modeling on cameron highlands (Malaysia): a comparison between random forest, logistic regression and logistic model tree algorithms," *Forests*, vol. 11, no. 8, p. 830, 2020.

[40] V.-H. Nhu, A. Shirzadi, H. Shahabi et al., "Shallow landslide susceptibility mapping by random forest base classifier and its ensembles in a semi-arid region of Iran," *Forests*, vol. 11, no. 4, p. 421, 2020.

[41] G. Wang, X. Lei, W. Chen, H. Shahabi, and A. Shirzadi, "Hybrid computational intelligence methods for landslide susceptibility mapping," *Symmetry*, vol. 12, no. 3, p. 325, 2020.

[42] M. D. Nguyen, B. T. Pham, T. T. Tuyen et al., "Development of an artificial intelligence approach for prediction of consolidation coefficient of soft soil: a sensitivity analysis," *The Open Construction and Building Technology Journal*, vol. 13, no. 1, p. 178, 2019.

[43] B. T. Pham, L. H. Son, T.-A. Hoang, D.-M. Nguyen, and D. Tien Bui, "Prediction of shear strength of soft soil using machine learning methods," *Catena*, vol. 166, pp. 181–191, 2018.

[44] B. T. Pham, M. D. Nguyen, K.-T. T. Bui, I. Prakash, K. Chapi, and D. T. Bui, "A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil," *Catena*, vol. 173, pp. 302–311, 2019.

[45] B. T. Pham, M. D. Nguyen, D. V. Dao et al., "Development of artificial intelligence models for the prediction of Compression Coefficient of soil: an application of Monte Carlo sensitivity analysis," *Science of the Total Environment*, vol. 679, pp. 172–184, 2019.

[46] H.-B. Ly and B. T. Pham, "Prediction of shear strength of soil using direct shear test and support vector machine model," *The Open Construction and Building Technology Journal*, vol. 14, no. 1, p. 41, 2020.

[47] B. T. Pham, C. Qi, L. S. Ho et al., "A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil," *Sustainability*, vol. 12, no. 6, p. 2218, 2020.

[48] D. T. Bui, N.-D. Hoang, and V.-H. Nhu, "A swarm intelligence-based machine learning approach for predicting soil shear strength for road construction: a case study at Trung Luong National Expressway Project (Vietnam)," *Engineering with Computers*, vol. 35, no. 3, pp. 955–965, 2019.

[49] H. Moayedi, M. Gör, M. Khari, L. K. Foong, M. Bahiraei, and D. T. Bui, "Hybridizing four wise neural-metaheuristic paradigms in predicting soil shear strength," *Measurement*, vol. 156, Article ID 107576, 2020.

[50] H. Moayedi, D. Tien Bui, A. Dounis, L. Kok Foong, and B. Kalantar, "Novel nature-inspired hybrids of neural computing for estimating soil shear strength," *Applied Sciences*, vol. 9, no. 21, p. 4643, 2019.

[51] H. Moayedi, D. Bui, D. Anastasios, and B. Kalantar, "Spotted hyena optimizer and ant lion optimization in predicting the shear strength of soil," *Applied Sciences*, vol. 9, no. 22, p. 4738, 2019.

[52] V.-H. Nhu, N.-D. Hoang, V.-B. Duong, H.-D. Vu, and D. T. Bui, "A hybrid computational intelligence approach for predicting soil shear strength for urban housing construction: a case study at Vinhomes Imperia project, Hai Phong City (Vietnam)," *Engineering with Computers*, vol. 36, no. 2, pp. 1–14, 2019.

[53] S. Das, P. Samui, S. Khan, and N. Sivakugan, "Machine learning techniques applied to prediction of residual strength of clay," *Open Geosciences*, vol. 3, no. 4, pp. 449–461, 2011.

[54] A. Besalatpour, M. A. Hajabbasi, S. Ayoubi, M. Afyuni, A. Jalalian, and R. Schulin, "Soil shear strength prediction using intelligent systems: artificial neural networks and an adaptive neuro-fuzzy inference system," *Soil Science and Plant Nutrition*, vol. 58, no. 2, pp. 149–160, 2012.

[55] B. Heung, H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, and M. G. Schmidt, "An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping," *Geoderma*, vol. 265, pp. 62–77, 2016.

[56] D. T. Bui, B. Pradhan, O. Lofman, I. Revhaug, and O. B. Dick, "Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS," *Computers & Geosciences*, vol. 45, pp. 199–211, 2012.

[57] W. Chen, J. Peng, H. Hong et al., "Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China," *Science of the Total Environment*, vol. 626, pp. 1121–1135, 2018.

[58] F. Huang, K. Yin, J. Huang, L. Gui, and P. Wang, "Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine," *Engineering Geology*, vol. 223, pp. 11–22, 2017.

[59] B. T. Pham, D. Tien Bui, H. R. Pourghasemi, P. Indra, and M. B. Dholakia, "Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods," *Theoretical and Applied Climatology*, vol. 128, no. 1-2, pp. 255–273, 2017.

[60] K. Taalab, T. Cheng, and Y. Zhang, "Mapping landslide susceptibility and types using Random Forest," *Big Earth Data*, vol. 2, no. 2, pp. 159–178, 2018.

[61] N. N. Vasu and S.-R. Lee, "A hybrid feature selection algorithm integrating an extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon, South Korea," *Geomorphology*, vol. 263, pp. 50–70, 2016.

[62] C. Qi, A. Fourie, Q. Chen, and Q. Zhang, "A strength prediction model using artificial intelligence for recycling waste tailings as cemented paste backfill," *Journal of Cleaner Production*, vol. 183, pp. 566–578, 2018.

[63] J. Zhou, P. G. Asteris, D. J. Armaghani, and B. T. Pham, "Prediction of ground vibration induced by blasting operations through the use of the Bayesian Network and random forest models," *Soil Dynamics and Earthquake Engineering*, vol. 139, Article ID 106390, 2020.

[64] S. Lu, M. Koopialipoor, P. G. Asteris, M. Bahri, and D. J. Armaghani, "A novel feature selection approach based on tree models for evaluating the punching shear capacity of steel fiber-reinforced concrete flat slabs," *Materials*, vol. 13, no. 17, p. 3902, 2020.

[65] D. J. Armaghani and P. G. Asteris, "A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength," *Neural Computing and Applications*, pp. 1–32, 2020.

[66] P. G. Asteris, "A novel heuristic algorithm for the modeling and risk assessment of the COVID-19 pandemic phenomenon," *Computer Modeling in Engineering & Sciences*, vol. 125, no. 2, pp. 815–828, 2020.

[67] D. J. Armaghani, E. Momeni, and P. Asteris, "Application of group method of data handling technique in assessing deformation of rock mass," *Applied Metaheuristic Computing*, vol. 1, pp. 1–18, 2020.

[68] D. Jahed Armaghani, P. G. Asteris, B. Askarian, M. Hasanipanah, R. Tarinejad, and V. V. Huynh, "Examining hybrid and single SVM models with different kernels to predict rock brittleness," *Sustainability*, vol. 12, no. 6, p. 2229, 2020.

[69] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 67, 2016.

[70] P. G. Asteris, "On the metaheuristic models for the prediction of cement-metakaolin mortars compressive strength," *Metaheuristic Computing and Applications*, vol. 1, no. 1, p. 063, 2020.

[71] M. Apostolopoulou, P. G. Asteris, D. J. Armaghani et al., "Mapping and holistic design of natural hydraulic lime mortars," *Cement and Concrete Research*, vol. 136, Article ID 106167, 2020.

[72] H.-B. Ly, B. T. Pham, L. M. Le, T.-T. Le, V. M. Le, and P. G. Asteris, "Estimation of axial load-carrying capacity of concrete-filled steel tubes using surrogate models," *Neural Computing and Applications*, pp. 1–22, 2020.

[73] B. T. Pham, T. Nguyen-Thoi, H.-B. Ly et al., "Extreme learning machine based prediction of soil shear strength: a sensitivity analysis using Monte Carlo simulations and feature backward elimination," *Sustainability*, vol. 12, no. 6, p. 2339, 2020.

[74] M. Alizadeh, E. Alizadeh, S. Asadollahpour Kotenaee et al., "Social vulnerability assessment using artificial neural network (ANN) model for earthquake hazard in Tabriz city, Iran," *Sustainability*, vol. 10, no. 10, p. 3376, 2018.

[75] P. G. Asteris and V. G. Mokos, "Concrete compressive strength using artificial neural networks," *Neural Computing and Applications*, pp. 1–20, 2019.

[76] B. T. Pham, S. K. Singh, and H.-B. Ly, "Using Artificial Neural Network (ANN) for prediction of soil coefficient of consolidation," *Vietnam Journal of Earth Sciences*, vol. 42, 2020.

[77] V. M. Le, B. T. Pham, T.-T. Le, H.-B. Ly, and L. M. Le, "Daily rainfall prediction using nonlinear autoregressive neural network," *Micro-Electronics and Telecommunication Engineering*, pp. 213–221, 2020.

[78] T.-T. Le, B. T. Pham, V. M. Le, H.-B. Ly, and L. M. Le, "A robustness analysis of different nonlinear autoregressive networks using Monte Carlo simulations for predicting high fluctuation rainfall," in *Micro-electronics and Telecommunication Engineering*, pp. 205–212, Springer, Berlin, Germany, 2020.

[79] T. A. Pham, H.-B. Ly, V. Q. Tran, L. V. Giap, H.-L. T. Vu, and H.-A. T. Duong, "Prediction of pile axial bearing capacity using artificial neural network and random forest," *Applied Sciences*, vol. 10, no. 5, p. 1871, 2020.

[80] M. A. Behrang, E. Assareh, A. Ghanbarzadeh, and A. R. Noghrehabadi, "The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data," *Solar Energy*, vol. 84, no. 8, pp. 1468–1480, 2010.

[81] K. K. Peh, C. P. Lim, S. S. Quek, and K. H. Khoh, "Use of artificial neural networks to predict drug dissolution profiles and evaluation of network performance using similarity factor," *Pharmaceutical Research*, vol. 17, no. 11, pp. 1384–1389, 2000.

[82] M. Zięba, S. K. Tomczak, and J. M. Tomczak, "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction," *Expert Systems with Applications*, vol. 58, pp. 93–101, 2016.

[83] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802–813, 2008.

[84] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[85] G. Bin Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.

[86] A. Shirzadi, S. Asadi, H. Shahabi et al., "A novel ensemble learning based on Bayesian Belief Network coupled with an extreme learning machine for flash flood susceptibility mapping," *Engineering Applications of Artificial Intelligence*, vol. 96, Article ID 103971, 2020.

[87] D. T. Bui, P.-T. T. Ngo, T. D. Pham et al., "A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping," *Catena*, vol. 179, pp. 184–196, 2019.

[88] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2011.

[89] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, vol. 2, pp. 985–990, Budapest, Hungary, July 2004.

[90] H.-B. Ly, P. G. Asteris, and B. T. Pham, "Accuracy assessment of extreme learning machine in predicting soil compression soefficient," *Vietnam Journal of Earth Sciences*, vol. 42, 2020.

[91] L. M. Le, H.-B. Ly, B. T. Pham et al., "Hybrid artificial intelligence approaches for predicting buckling damage of steel columns under axial compression," *Materials*, vol. 12, no. 10, p. 1670, 2019.

[92] X. Wang, Z. Yang, and A. P. Jivkov, "Monte Carlo simulations of mesoscale fracture of concrete with random aggregates and pores: a size effect study," *Construction and Building Materials*, vol. 80, pp. 262–272, 2015.

[93] J. Guilleminot and C. Soize, "Generalized stochastic approach for constitutive equation in linear elasticity: a random matrix model," *International Journal for Numerical Methods in Engineering*, vol. 90, no. 5, pp. 613–635, 2012.

[94] S. Mordechai, *Applications of Monte Carlo Method in Science and Engineering*, InTechOpen, Rijeka, Croatia, 2012.

[95] H.-B. Ly, E. Monteiro, T.-T. Le et al., "Prediction and sensitivity analysis of bubble dissolution time in 3D selective laser sintering using ensemble decision trees," *Materials*, vol. 12, no. 9, p. 1544, 2019.

[96] H.-L. Nguyen, B. T. Pham, L. H. Son et al., "Adaptive network based fuzzy inference system with meta-heuristic optimizations for international roughness index prediction," *Applied Sciences*, vol. 9, no. 21, p. 4715, 2019.

[97] H. Q. Nguyen, H.-B. Ly, V. Q. Tran, T.-A. Nguyen, T.-T. Le, and B. T. Pham, "Optimization of artificial intelligence system by evolutionary algorithm for prediction of axial capacity of rectangular concrete filled steel tubes under compression," *Materials*, vol. 13, no. 5, p. 1205, 2020.

[98] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.

[99] B. T. Pham, I. Prakash, A. Jaafari, and D. T. Bui, "Spatial prediction of rainfall-induced landslides using aggregating one-dependence estimators classifier," *Journal of the Indian Society of Remote Sensing*, vol. 46, no. 9, pp. 1457–1470, 2018.

[100] C. Verma and Z. Illés, "Attitude prediction towards ICT and mobile technology for the real-time: an experimental study using machine learning," in *Proceedings of the 15th eLearning and Software for Education Conference - eLSE 2019*, vol. 3, pp. 247–254, Bucharest, Romania, April 2019.

[101] D. V. Dao, A. Jaafari, M. Bayat et al., "A spatially explicit deep learning neural network model for the prediction of landslide susceptibility," *Catena*, vol. 188, p. 104451, 2020.

[102] C. Qi, H.-B. Ly, Q. Chen, T.-T. Le, V. M. Le, and B. T. Pham, "Flocculation-dewatering prediction of fine mineral tailings using a hybrid machine learning approach," *Chemosphere*, vol. 244, Article ID 125450, 2020.

[103] B. T. Pham, M. Avand, S. Janizadeh et al., "GIS based hybrid computational approaches for flash flood susceptibility assessment," *Water*, vol. 12, no. 3, p. 683, 2020.

[104] P. T. Nguyen, D. H. Ha, M. Avand et al., "Soft computing ensemble models based on logistic regression for groundwater potential mapping," *Applied Sciences*, vol. 10, no. 7, p. 2469, 2020.

[105] P. T. Nguyen, D. H. Ha, A. Jaafari et al., "Groundwater potential mapping combining artificial neural network and real AdaBoost ensemble technique: the DakNong province case-study, Vietnam," *International Journal of Environmental Research and Public Health*, vol. 17, no. 7, p. 2473, 2020.

[106] W. Chen, H. R. Pourghasemi, A. Kornejady, and N. Zhang, "Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques," *Geoderma*, vol. 305, pp. 314–327, 2017.

[107] Z. H. Khan, T. S. Alin, and M. A. Hussain, "Price prediction of share market using artificial neural network (ANN)," *International Journal of Computer Applications*, vol. 22, no. 2, pp. 42–47, 2011.

[108] J. C. Gertrudes, V. G. Maltarollo, R. A. Silva, P. R. Oliveira, K. M. Honorio, and A. B. F. Da Silva, "Machine learning techniques and drug design," *Current Medicinal Chemistry*, vol. 19, no. 25, pp. 4289–4297, 2012.

[109] A.-L. Milac, S. Avram, and A.-J. Petrescu, "Evaluation of a neural networks QSAR method based on ligand representation using substituent descriptors: application to HIV-1 protease inhibitors," *Journal of Molecular Graphics and Modelling*, vol. 25, no. 1, pp. 37–45, 2006.