

Research Article

A Comparative Study of Soft Computing Models for Prediction of Permeability Coefficient of Soil

Binh Thai Pham ^{1,2} **Manh Duc Nguyen**,³ **Nadhir Al-Ansari** ⁴ **Quoc Anh Tran**,⁵
Lanh Si Ho ^{1,2} **Hiep Van Le**,¹ and **Indra Prakash** ⁶

¹University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi 100000, Vietnam

²Civil and Environmental Engineering Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan

³University of Transport and Communications, Hanoi 100000, Vietnam

⁴Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea 971 87, Sweden

⁵Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway

⁶DDG (R) Geological Survey of India, Gandhinagar 382010, India

Correspondence should be addressed to Binh Thai Pham; binhpt@utt.edu.vn, Nadhir Al-Ansari; nadhir.alansari@ltu.se, and Lanh Si Ho; lanhs@utt.edu.vn

Received 11 May 2021; Accepted 2 November 2021; Published 19 November 2021

Academic Editor: Amin Jajarmi

Copyright © 2021 Binh Thai Pham et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Determination of the permeability coefficient (K) of soil is considered as one of the essential steps to assess infiltration, runoff, groundwater, and drainage in the design process of the construction projects. In this study, three cost-effective algorithms, namely, artificial neural network (ANN), support vector machine (SVM), and random forest (RF), which are well-known as advanced machine learning techniques, were used to predict the permeability coefficient (K) of soil (10^{-9} cm/s), based on a set of simple six input parameters such as natural water content w (%), void ratio (e), specific density (g/cm^3), liquid limit (LL) (%), plastic limit (PL) (%), and clay content (%). For this, a total of 84 soil samples data collected from the detailed design stage investigations of Da Nang-Quang Ngai national road project in Vietnam was used to generate training (70%) and testing (30%) datasets for building and validating the models. Statistical error indicators such as RMSE and MAE and correlation coefficient (R) were used to evaluate and compare performance of the models. The results show that all the three models performed well ($R > 0.8$) for the prediction of permeability coefficient of soil, but the RF model (RMSE = 0.0084, MAE = 0.0049, and $R = 0.851$) is more efficient compared with the other two models, namely, ANN (RMSE = 0.001, MAE = 0.005, and $R = 0.845$) and SVM (RMSE = 0.0098, MAE = 0.0064, and $R = 0.844$). Thus, it can be concluded that the RF model can be used for accurate estimation of the permeability coefficient (K) of the soil.

1. Introduction

The permeability of soil is one of the most important factors that govern the fluid flow characteristics of the soil. Generally, the permeability is represented by an amount of water transmit via interconnected void of a soil mass in a certain period, and it can be determined using field and laboratory tests. It is accepted fact that determination of the soil permeability coefficient is very crucial, and this task is difficult, time-consuming, and expensive [1, 2].

In geotechnical point of view, the soil permeability depends on many factors such as the soil density, water content, void ratio, mineralogy, soil structures, and others. The permeability coefficient is used in many geotechnical problems such as slope stability, the failure of structures related to the ground settlement, seepage, and leakage. Thus, many authors have tried to establish empirical relationships between influencing factors with the permeability coefficient [3–5]. There are several direct relationships between grain size and the permeability coefficient of soil. Hazen [6]

indicated that the permeability is proportional to the square of the effective grain size for the sand with uniform particles. Other authors proposed a regression that considers porosity, percentage of clay, and sand particle to estimate the permeability of soil [7]. Some other authors predicted soil permeability based on bulk density and grain-size particle and shape of the particle [8, 9]. As mentioned above, the permeability of soil is strongly dependent on the particle size distribution; however, it is not applicable for a wide range of soil [1, 10]. The study indicated that these empirical relationships have certain limitations as well as uncertainties.

Nowadays, machine learning (ML) and artificial intelligence (AI) techniques have been applied successfully in many fields including civil engineering. The ML techniques could enable engineers to estimate the unknown parameters relating to these problems with superior approximation abilities. Soft computer methods, for example, fuzzy logic, artificial neural networks (ANNs), and support vector machine (SVM) are now being used in geotechnical engineering for predicting soil compressive and shear strength, load bearing capacity of foundation, and so on [11–13]. Several authors have used ML techniques to estimate tensile strength of rock as well as flyrock caused by blasting [14]. AI and ML techniques are also being used frequently for the landslide studies, flood management, and infrastructure development.

Regarding the prediction of the soil permeability coefficient, there are several studies using the ML method, for instance, ANN, adaptive neuro-fuzzy system (ANFIS), and hybrid optimization model of genetic algorithm-ANFIS (GA-ANFIS) [1, 2, 15–17]. Sezer et al. [17] used an ANFIS to estimate the permeability of granular soil; they concluded that the ANFIS algorithm is superior to estimate the permeability of granular soil considering grain-size distribution and particle shape [2]. However, the hybrid model GA-ANFIS outperformed in terms of prediction accuracy compared with single ANN, ANFIS model, and hybrid GA-ANN model [15]. In general, soft computing-based models are great tools for the prediction of the properties of soil.

Random forest (RF) was firstly proposed by Breiman to solve unsupervised learning, regression, and classification problems [18–20], which is known as a powerful algorithm, which has been successfully employed and applied in many problems of geotechnical engineering field [21, 22]. For example, RF has been utilized successfully in predicting soil parameters such as prediction of shear strength of soil and soil permeability coefficient [20, 23]. The RF algorithm has important merits in handling with large databases, and it can also deal with thousand input variables [24].

Based on the literature survey, it can be concluded that these ML techniques have many advantages in predicting soil parameters. To the best of the authors' knowledge, there is no study on estimating permeability coefficient of soil using these techniques in Vietnam condition. Main difference with earlier studies is that here we have used different datasets to compare the performance of different models to select the best model for the estimation of permeability coefficient of soil (K). Moreover, first time, the RF model has been used in the determination of ' K ' in the study area of Vietnam.

Therefore, main objective of this study is to apply popular soft computing techniques (ANN, SVM, and RF) at the Da Nang-Quang Ngai expressway project site of Vietnam for the estimation of the permeability coefficient (K) of soil and to select the best model for the prediction of " K ." Various statistical evaluation indicators such as RMSE and MAE and correlation coefficient (R) were used to validate and evaluate the models. Matlab software was used for the data processing and to simulate the models: ANN, SVM, and RF.

2. Materials and Methods

2.1. Data Used. The dataset consists of 84 soil samples collected from the detailed design state investigations of Da Nang-Quang Ngai expressway development project near Da Nang, central Vietnam (Figure 1). To predict the " K " of soil, the input data related to the permeability are selected, such as water content (%), void ratio, specific density (g/cm^3), liquid limit (LL), plastic limit (PL), and clay content (%). All these input data are highly related to permeability especially void ratio which is the critical parameters for having a relationship with hydraulic conductivity in both Darcy's equation and Kozeny-Carman's equations (1) and (2).

Initial statistical analysis of the dataset is presented in Table 1. The natural water content values vary from 15.1% to 99%. The void ratio varies from 0.46 to 2.63. The distribution of the specific density ranges from $2.58 \text{ g}/\text{cm}^3$ to $2.74 \text{ g}/\text{cm}^3$. The liquid limit is from 18.9% to 88.93%, the plastic limit is from 12.2% to 54.8%, and finally the clay content is from 5.7% to 64%. Figure 2 shows the histogram of the input parameters.

2.2. Methods Used

2.2.1. Artificial Neural Network (ANN). ANN is known as a common and powerful technique that imitates the activity and performance of the human brain and nervous system [15–17]. This technique has many crucial abilities such as generalization and learning from data and can deal with a large variable. It was reported that the major characteristics of ANN comprises continuous nonlinear dynamics, high fault tolerance, collective computation, self-learning, self-organization, and real-time treatment [25]. Thus, this algorithm has been widely employed and applied successfully to solve many problems in geotechnical engineering. In both linear and nonlinear patterns, ANN is generally adopted to determine the hidden layer between output and input neurons; as a result, ANN could decide analyzing relationships and patterns by itself in data. In order to predict the permeability coefficient of soil, a multilayer perceptron (MLP) was adopted as a regression technique. To calculate the weights of the input through the activation function, the sigmoid function is used in neurons.

$$h_i = L_i(x) = \frac{1}{1 + e^{-x}}, \quad (1)$$

where h_i indicate the permeability coefficient (output) and $x = (x_1, x_2, \dots, x_i)$ denote input parameters (i.e., affected factors of permeability coefficient).

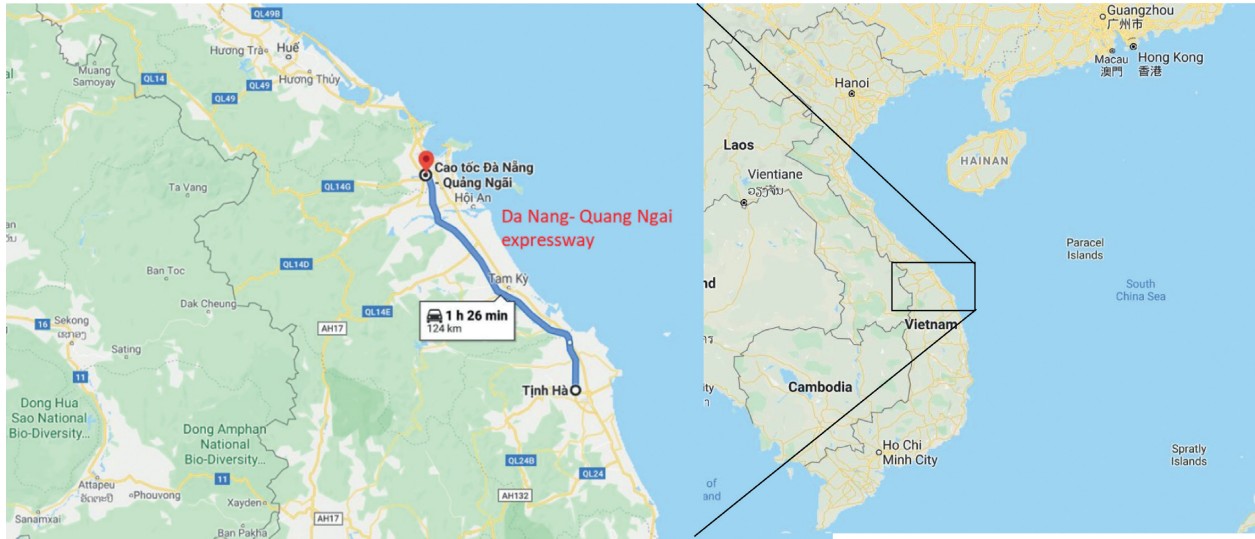


FIGURE 1: Location map of the studied site: Da Nang-Quang Ngai expressway project.

TABLE 1: Statistical analysis of the inputs and outputs in this study.

Parameters	Notations	Unit	Minimum	Maximum	Average	Std*
Natural water content	w	%	15.1	99.9	34.23	16.5
Void ratio	e	—	0.46	2.63	0.97	0.42
Specific density	γ	g/cm^3	2.58	2.74	2.68	0.02
Liquid limit	LL	%	18.9	88.93	37.27	13.04
Plastic limit	PL	%	12.2	54.8	22.21	7.04
Clay content		%	5.7	64	25.17	11.5
Permeability coefficient	K	$10^{-9} cm/s$	0.3	7.1	1.45	0.94

* Standard deviation.

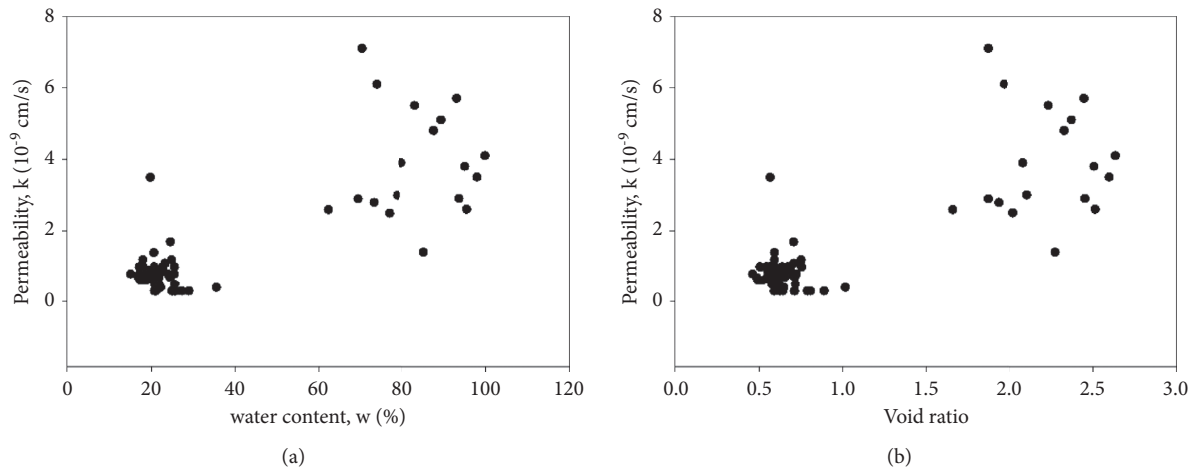


FIGURE 2: Continued.

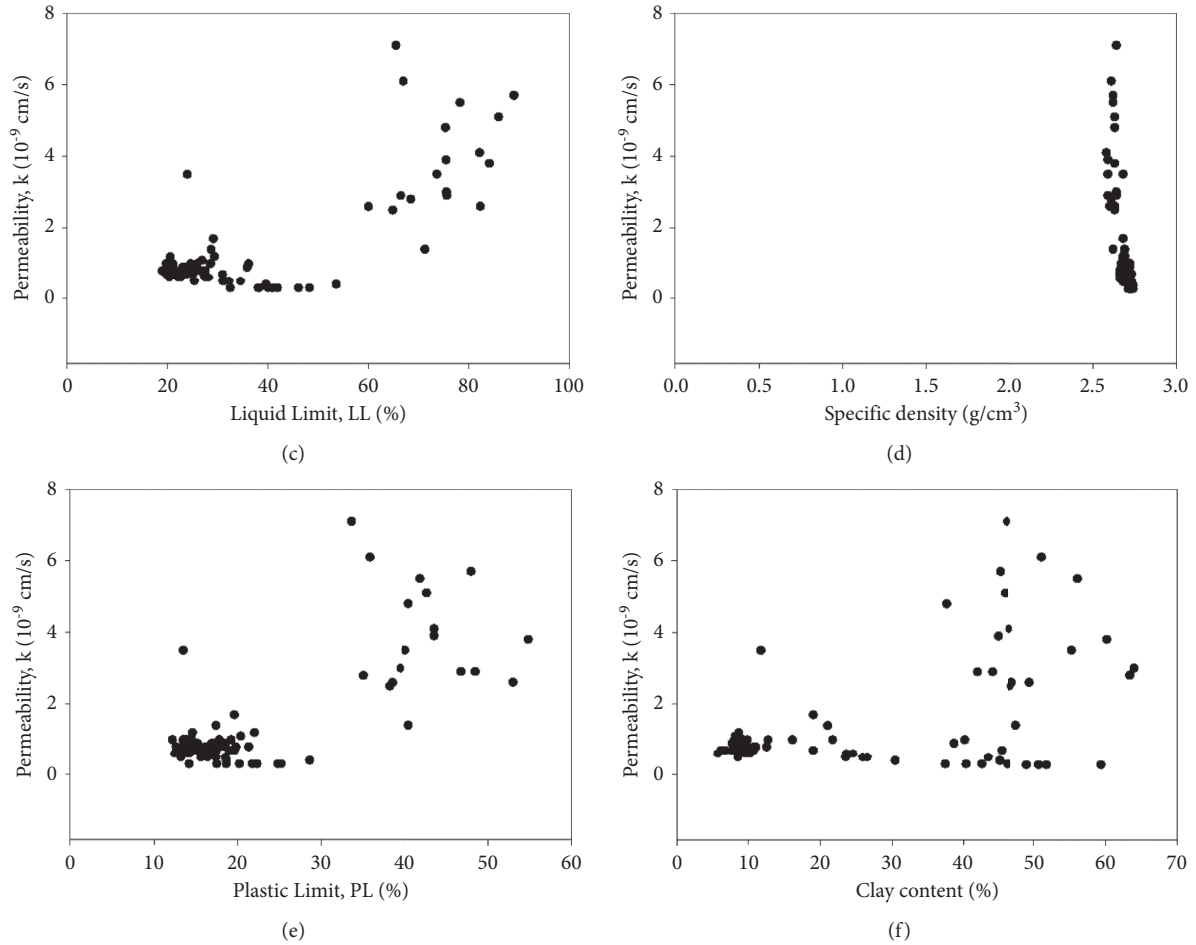


FIGURE 2: Histogram of the input parameters used in the models: (a) water content vs. permeability; (b) void ratio vs. permeability; (c) liquid limit vs. permeability; (d) specific density vs. permeability; (e) plastic limit vs. permeability; (f) clay content vs. permeability.

2.2.2. Support Vector Machine (SVM). SVM is known as a statistical-based learning algorithm that was firstly proposed by Vapnik to deal with the nonlinear problems with high dimension such as regression and classification [26, 27]. The concept of SVM is to build a hyperplane to separate the dataset into different classes. In the SVM, the original input space is transferred to a high-dimensional feature space using the training dataset [28–30].

Then, the optimum plane is defined via optimizing the class boundary. Thus, the support vectors are defined as the trained points that are placed the most adjacent to the optimal plane [28, 29]. SVM has been popularly used in landslide prediction, and the results showed that this technique has high accuracy [28, 31]. In this study, the SVM was employed as a regression method by propositioning a function of δ -insensitive loss [32].

2.2.3. Random Forest (RF). RF is known as a prevailing algorithm, which was firstly suggested by Breiman to solve classification, regression, and unsupervised learning problems [18, 22]. This algorithm is being employed commonly in different fields of civil engineering containing geotechnical engineering [21, 33, 34]. This machine learning

comprises several merits such as high performance with complex datasets using small calibrating and can deal with high noise variables [35, 36]. In addition, it was reported that this algorithm is very user-friendly because it has only two parameters (including a number of variables and trees) and it is usually not sensitive to their values [22].

In a random forest, the bagging technique is always used to randomly select the variables from the whole dataset for model calibration. In this study, two kinds of errors, including reduction in Gini and reduction in accuracy, and an Out-of-Bag (OOB) were computed because these error factors can be employed to rank and choose variables [37, 38]. For each variable, when the values of the variable are transferred over the OOB observations, the error of the estimable model will be decided by the function.

2.2.4. Relief F for Attribute Importance. In general, evaluating attribute quality (feature quality) is known as a crucial task for both regression and classification problems in machine learning such as constructive induction, regression and decision tree, and feature selection [39, 40]. Each input variable in a huge number of a learning problem is governed by thousands of attribute (feature). Generally, many learning

techniques cannot deal with this situation because of lack of information of features or variables with many irrelevances. An attribute (feature) selection is known as a task to choose a small subset, which is adequate to pronounce the target purpose. In order to decide which features need to be kept and which ones need to be removed, it is necessary to have a practical and reliable method for evaluating the related information to the target goal.

In recent years, many researchers have paid much effort to evaluate feature estimation. There exist many methods for estimating the quality of attributes. For the regression problem, mean square and mean absolute error [41] and Relief F [39] are used as estimation heuristics. Almost the heuristic methods used for evaluating the attribute quality of the attributes made the assumption of the conditional independence of the features. These methods are thus less suitable for problems that have much feature interaction. In opposite, Relief F does not assume the condition for the attribute. This algorithm is effective, to understand the circumstantial information, and can appropriately predict the attribute quality of problems with a high dependence between features [42]. It was reported that Relief F has been widely considered as an attribute selection method, which is used as the pre-processing step beforehand the model is learned and trained [43]. This method is known as one of the most effective algorithms until now [44]. Finally, Relief F could provide a unified assessment on evaluating the quality of features in regression problems. The detail of this algorithm can be found in the previous studies [39, 40].

2.2.5. Validation Indices. In this research, to assess, compare, and validate the performance of the model, RMSE, MAE, and R were employed. Generally, RMSE can be used to measure the mean squared difference between actual and estimated values, while MAE is used to determine the average error amplitude. When the values of RMSE and MAE are smaller, the model will have higher predictive ability. By contrast, higher values of R indicate the higher prediction ability of the model. These indicators (RMSE, MAE, and R) are usually applied for the regression problem that can be determined by using the following formulas [45, 46]:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^M (q_1 - q_2)^2}{M}}, \quad (2)$$

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |q_1 - q_2|, \quad (3)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^M (q_1 - q_2)^2}{\sum_{i=1}^M (q_1 - \bar{q}_i)^2}}, \quad (4)$$

where q_1 and q_2 correspond to the measured and modeled values, \bar{q}_i indicates the average permeability coefficient value, and M is the summation of input.

2.3. Methodology. In this research, there are few main steps carried out to predict the “ K ” of soil as indicated in Figure 3.

Step 1. First, the input dataset is generated and loaded, and then these datasets are randomly divided into testing (30%) and training (70%) groups. The split of this dataset in 70 : 30 ratio was done for the training and testing of the models, respectively, based on the experience of authors and similar studies carried out by other researchers for obtaining the best performance of the models [47]. In this step, the Relief F feature selection method was applied to validate the importance of the input variables on which the important parameters were selected for the generation of final training and testing datasets after removing irrelevant parameters.

Step 2. In this step, a training dataset was used to train the soft computing-based models (ANN, SVM, and RF). To get the best performance of these models, the optimization of the hyperparameters used in each model was carried out using the trial-error process. In this study, the ANN was trained with 10 hidden layers with sigma loss function, the SVM was trained with Radial Basis Function (RBF) kernel function using the gamma value of 0.25, and the RF was trained with 100 iterations.

Step 3. Validation of the models (ANN, SVM, and RF) was done in this step using testing dataset. Various statistical indicators (RMSE, MAE, and R) were calculated using both training datasets. While the values of these indicators using the training dataset indicate the goodness of fit of these models with the data used, the one using the testing dataset indicates the predictive capability of these models.

3. Results

3.1. Attribute Importance Using Relief F. We evaluated the importance of the input parameter by using the Relief F technique for the six input parameters including the water content, void ratio, specific density, liquid limit, plastic limit, and clay content (Table 2). The clay content was found to be the less important variables of the permeability with the weight value of merely 0.025. The weights of the other index parameters including plastic limit, liquid limit, and specific density are 0.0753, 0.0762, and 0.0877, respectively. Finally, the water content and void ratio are shown to be the most important parameters with a weight of 0.096 and 0.0942, correspondingly.

3.2. Validation and Comparison of the Models. Validation of the models (ANN, SVM, and RF) was done using both training and testing datasets as indicated in Figures 4–6 and summarized in Table 3. With respect to the training dataset, the RF has the highest value of R (0.972), followed by the ANN (0.948) and the SVM (0.861), respectively. In contrast, the RF has the lowest value of RMSE (0.0035) and MAE (0.0023), followed by the ANN (0.0047 and 0.0027) and the

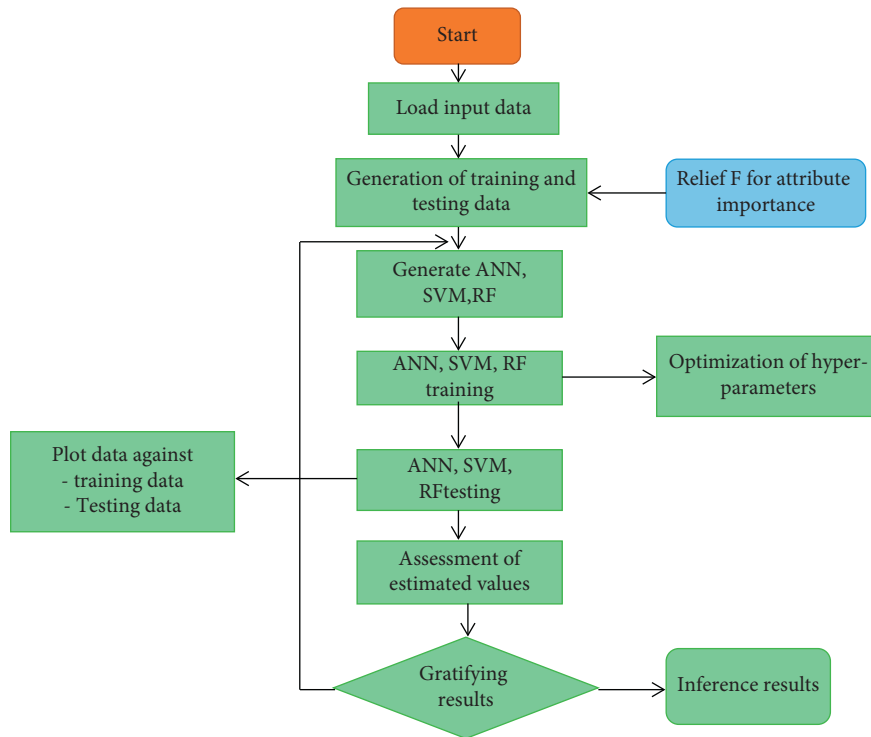


FIGURE 3: Methodological flowchart used in this study.

TABLE 2: Importance of input parameters using Relief F.

No.	Input parameters	Weights
1	Natural water content	0.096
2	Void ratio	0.0942
3	Specific density	0.0877
4	Liquid limit	0.0762
5	Plastic limit	0.0753
6	Clay content	0.025

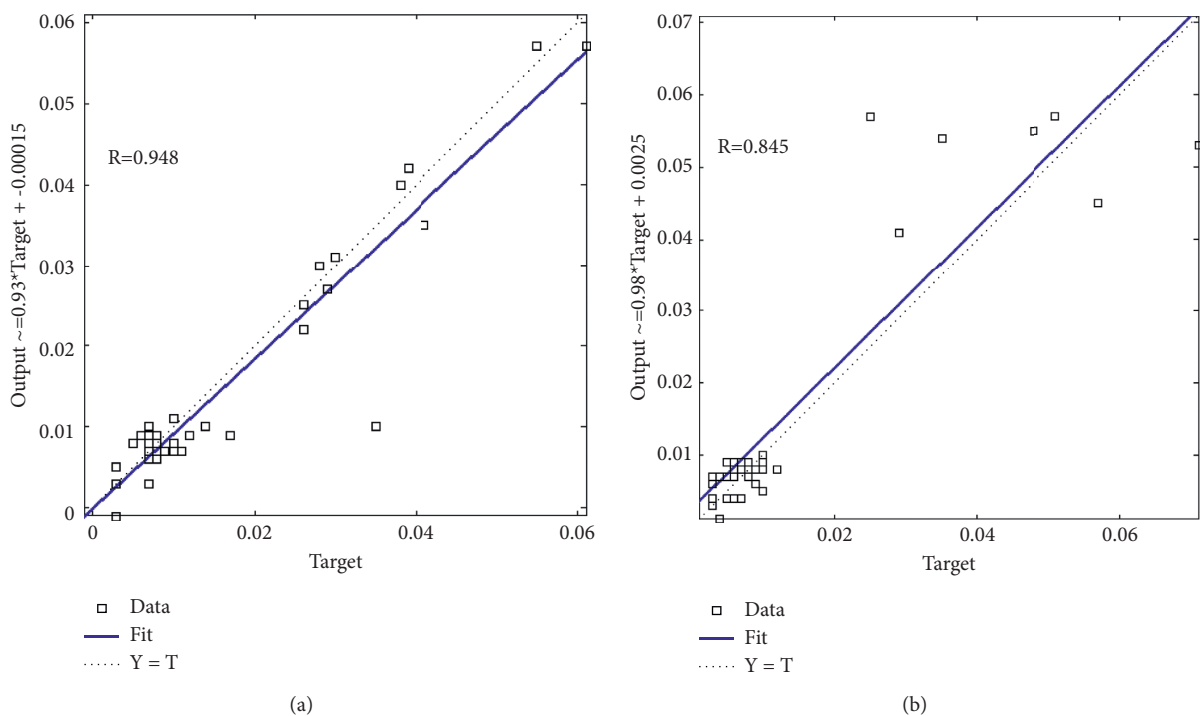


FIGURE 4: Continued.

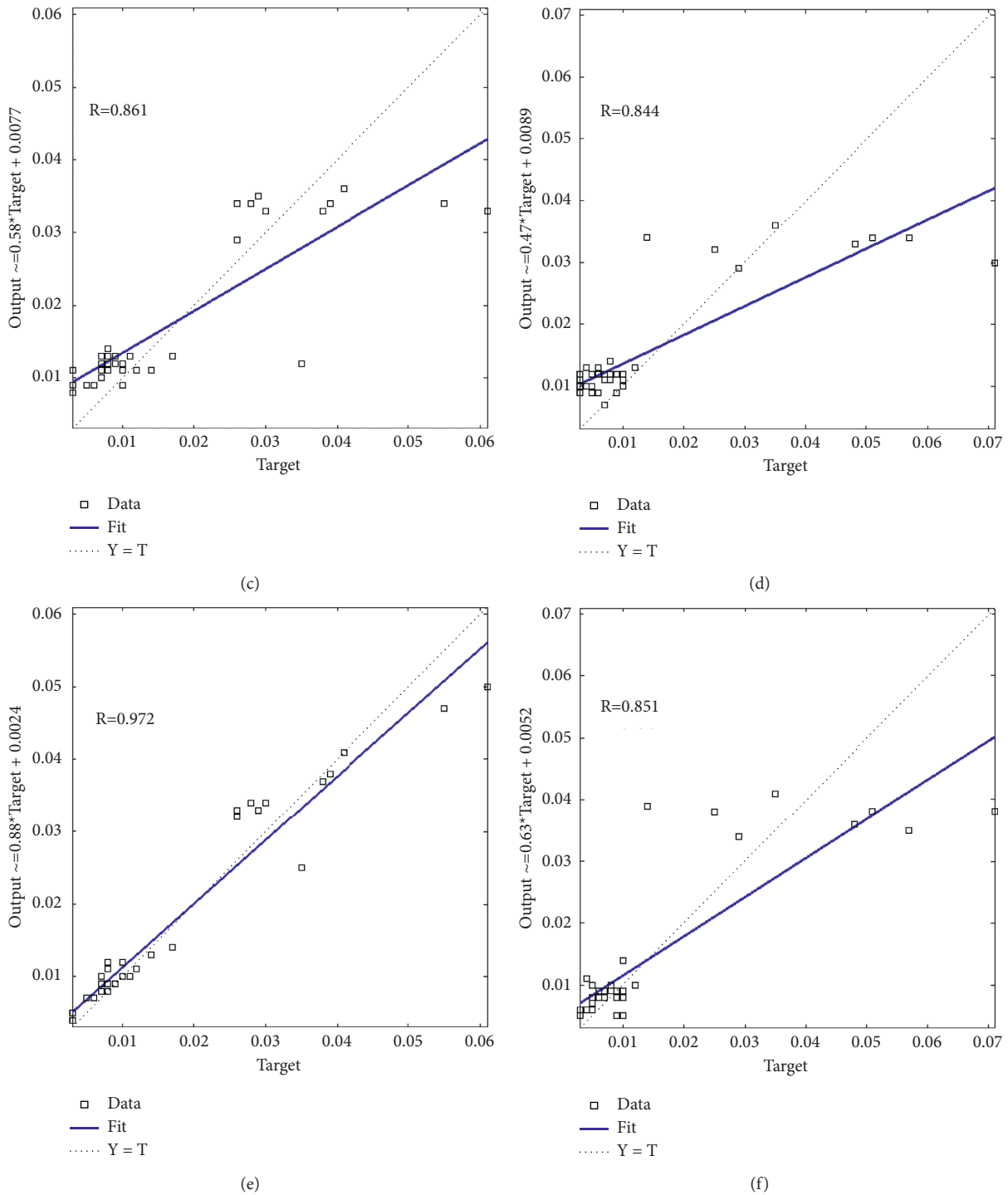


FIGURE 4: Correlation analysis of actual and predicted outputs: (a) ANN with training dataset, (b) ANN with testing dataset, (c) SVM with training dataset, (d) SVM with testing dataset, (e) RF with training dataset, and (f) RF with testing dataset.

SVM (0.0078 and 0.0056), respectively. These results on the training dataset show that the RF has the highest goodness of fit with the data used compared with other models (SVM and ANN). In terms of the testing dataset, similarly, the RF has the highest value of R (0.851), followed by the ANN (0.845) and the SVM (0.844), respectively. However, the ANN has the

lowest value of RMSE (0.001), followed by the RF (0.0084) and the SVM (0.0098), respectively, and the RF has the lowest value of MAE (0.0049), followed by the ANN (0.005) and the SVM (0.0064), respectively. Figure 5 shows the visualization of the actual and predicted values of the permeability coefficient of soil through experiments and models, respectively.

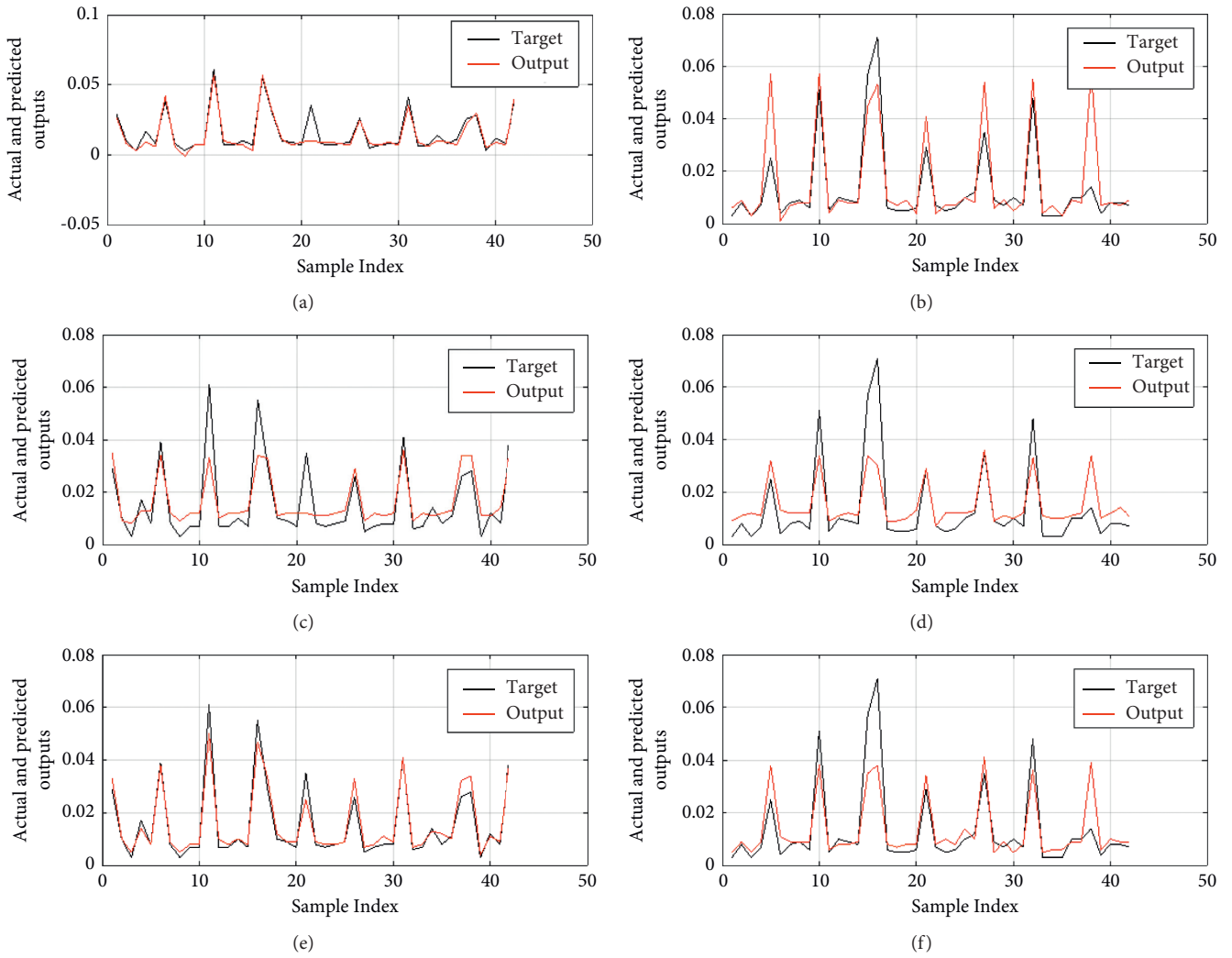


FIGURE 5: Analysis of errors of the models: (a) ANN with training dataset, (b) ANN with testing dataset, (c) SVM with training dataset, (d) SVM with testing dataset, (e) RF with training dataset, and (f) RF with testing dataset.

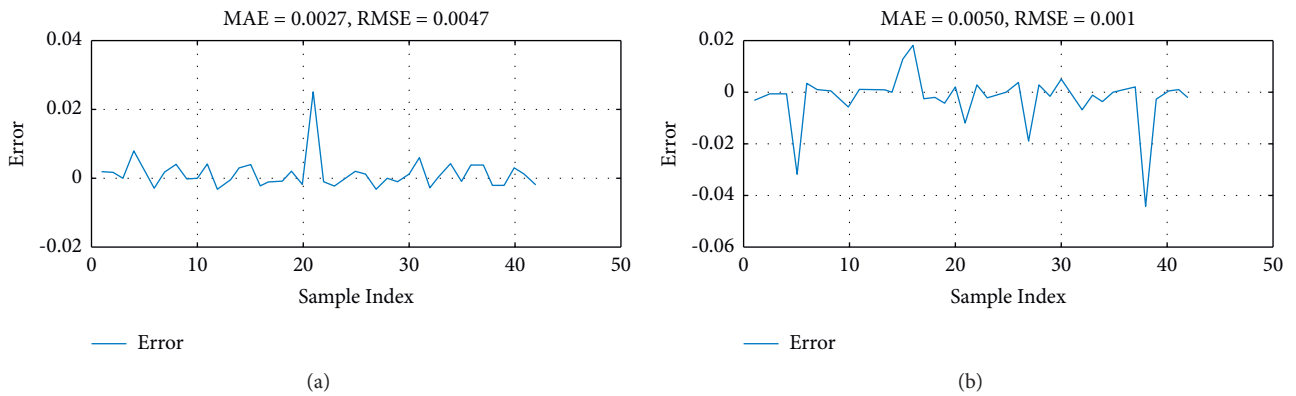


FIGURE 6: Continued.

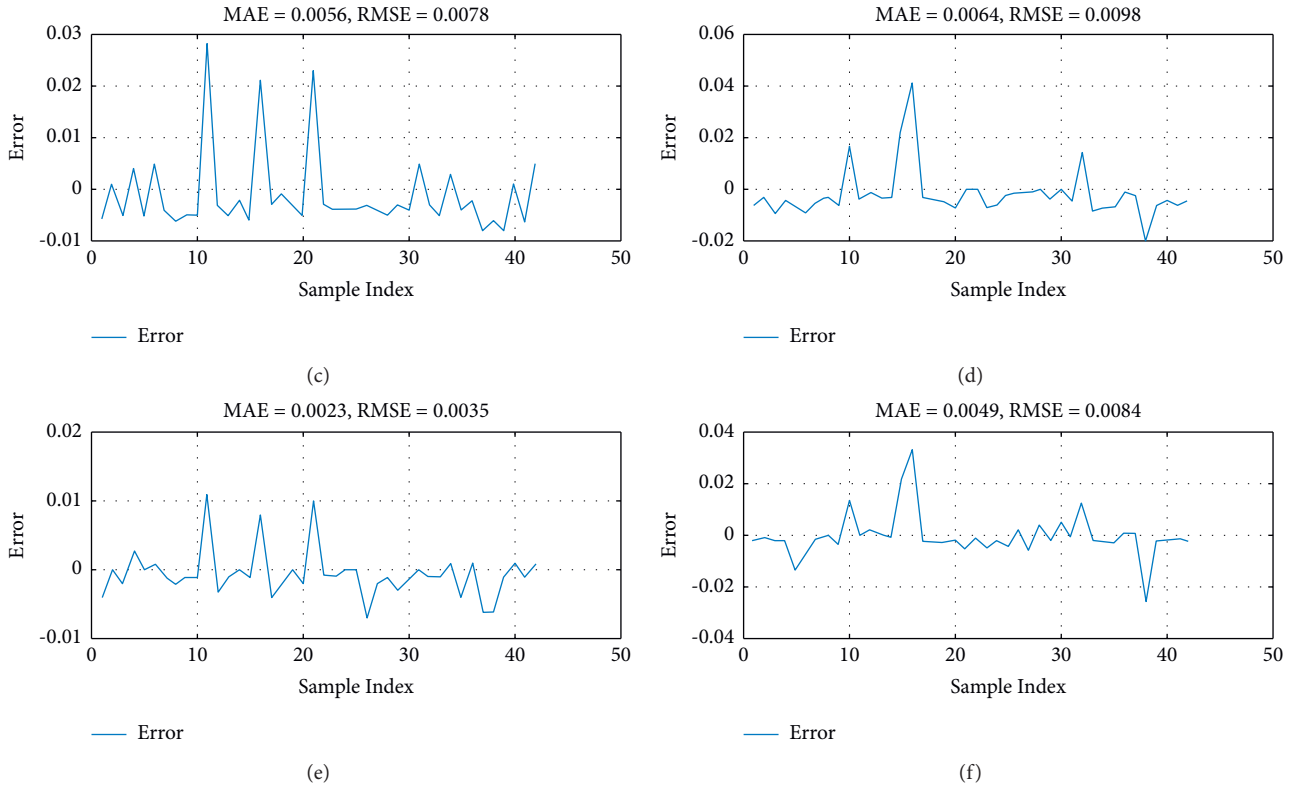


FIGURE 6: Analysis of errors of the models: (a) ANN with training dataset, (b) ANN with testing dataset, (c) SVM with training dataset, (d) SVM with testing dataset, (e) RF with training dataset, and (f) RF with testing dataset.

TABLE 3: Validation and comparison of the ANN, SVM, and RF.

Indicators	ANN		SVM		RF	
	Training	Testing	Training	Testing	Training	Testing
RMSE	0.0047	0.001	0.0078	0.0098	0.0035	0.0084
MAE	0.0027	0.005	0.0056	0.0064	0.0023	0.0049
<i>R</i>	0.948	0.845	0.861	0.844	0.972	0.851

4. Discussion and Conclusion

In the geotechnical study, the permeability coefficient (K) of soil is an important factor for designing civil engineering structures on soil. However, determining the “ K ” in the laboratory or in the field is time-consuming and expensive. Indirect estimation of “ K ” using empirical equation and correlating with other engineering properties of soils may not be accurate [3–5]. Moreover, they may be applicable to specific soil only. Therefore, in this study, we have applied three popular cost-effective soft computing-based models such as ANN, SVM, and RF to predict “ K ” of the Da Nang-Quang Ngai expressway development project site soil by using six soil parameters, namely, water content, void ratio, specific density, liquid limit, plastic limit, and clay content as input in the studied models.

The Relief F feature selection method results showed that the void ratio and the water content were found to be the most important input variables (parameters) in the prediction of the “ K ” of the soil. It is reasonable because the void

ratio is highly correlated to the permeability in several studies [48]. On the other hand, the water content represents the level of saturation, which directly links to the fluid flow in the porous media [49].

The validation results showed that all three models are good at estimating the prediction of soil coefficient of permeability. However, the RF is found to be the most accurate method to predict the “ K ” of soil in comparison with SVM and ANN. This can be attributed to the ability of the RF algorithm in processing large databases with a large number of input parameters also [49]. The results of this study also are in a good agreement with the results of other studies on estimating the shear strength of soil where performance of the RF model was the best in comparison with other ML models [20, 23].

In general, the soft computing-based models developed in this study contribute a powerful tool to estimate the permeability coefficient of the soil accurately. However, the performance of the model depends on the input parameters, so it is necessary to carry out the various strategies to

improve the input samples to improve the performance of the model. In addition, it is necessary to consider the overfitting problem [50]. Therefore, the data for the training are crucial for accurate prediction. Once need to make sure that the data are required to be reliable and sufficient to apply the machine learning technique into practice. In this study, we have used 70% of the total data as training data for obtaining optimum results based on the earlier studies [51–53].

Development and improvement of the performance of models are a continuous process. The findings of this study are that the RF model can be used to estimate accurate permeability coefficient of the soil using limited soil parameters but more studies at different sites are required for confirming its wider applicability.

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 there are no conflicts of interest.

Acknowledgments

The authors would like to thank the support of the Department of Science, Technology, and Environment (Ministry of Education and Training), University of Transport and Communications, and other agencies for providing data used in this research. This study was funded by the Ministry of Education and Training under grant number B2020-GHA-03 chaired by the University of Transportation.

References

- [1] S. K. Sinha and M. C. Wang, "Artificial neural network prediction models for soil compaction and permeability," *Geotechnical & Geological Engineering*, vol. 26, no. 1, pp. 47–64, 2008.
- [2] I. Yilmaz, M. Marschalko, M. Bednarik, O. Kaynar, and L. Fojtova, "Neural computing models for prediction of permeability coefficient of coarse-grained soils," *Neural Computing & Applications*, vol. 21, no. 5, pp. 957–968, 2012.
- [3] I. Garcia-Bengochea, A. G. Altschaeffl, and C. W. Lovell, "Pore distribution and permeability of silty clays," *Journal of the Geotechnical Engineering Division*, vol. 105, no. 7, pp. 839–856, 1979.
- [4] J. K. Mitchell, D. R. Hooper, and R. G. Campenella, "Permeability of compacted clay," *Journal of the Soil Mechanics and Foundations Division*, vol. 91, no. 4, pp. 41–65, 1965.
- [5] R. E. Olson, "Effective stress theory of soil compaction," *Journal of the Soil Mechanics and Foundations Division*, vol. 89, no. 2, pp. 27–45, 1963.
- [6] J. B. Burland, M. C. Burbidge, E. J. Wilson et al., "Discussion. Settlement of foundations on sand and gravel," *Proceedings-Institution of Civil Engineers*, vol. 80, pp. 1625–1648, 1986.
- [7] W. J. Rawls and D. L. Brakensiek, *Estimation of Soil Water Retention and Hydraulic Properties*, pp. 275–300, Springer, Berlin, Germany, 1989, Unsaturated Flow Hydrol. Model.
- [8] I. Lebron, M. G. Schaap, and D. L. Suarez, "Saturated hydraulic conductivity prediction from microscopic pore geometry measurements and neural network analysis," *Water Resources Research*, vol. 35, pp. 3149–3158, 1999.
- [9] J. M. Sperry and J. J. Peirce, "A model for estimating the hydraulic conductivity of granular material based on grain shape, grain size, and porosity," *Ground Water*, vol. 33, pp. 892–898, 1995.
- [10] V. L. Hauser, "Seepage control by particle size selection," *Transactions of the American Society of Agricultural Engineers*, vol. 21, pp. 691–695, 1978.
- [11] M. Marjanović, M. Kovačević, B. Bajat, and V. Voženilek, "Landslide susceptibility assessment using SVM machine learning algorithm," *Engineering Geology*, vol. 123, pp. 225–234, 2011.
- [12] B. T. Pham, D. T. Bui, I. Prakash, and M. B. Dholakia, "Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS," *Catena*, vol. 149, pp. 52–63, 2017.
- [13] 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 Computers*, vol. 35, pp. 955–965, 2019.
- [14] M. Hasanipanah, B. Keshtegar, D.-K. Thai, and N.-T. Troung, "An ANN-adaptive dynamical harmony search algorithm to approximate the flyrock resulting from blasting," *Engineering Computers*, pp. 1–13, 2020.
- [15] H. Ganjidoost, S. J. Mousavi, and A. Soroush, "Adaptive network-based fuzzy inference systems coupled with genetic algorithms for predicting soil permeability coefficient," *Neural Processing Letters*, vol. 44, pp. 53–79, 2016.
- [16] H. I. Park, "Development of neural network model to estimate the permeability coefficient of soils," *Marine Georesources & Geotechnology*, vol. 29, pp. 267–278, 2011.
- [17] A. Sezer, A. B. Göktepe, and S. Altun, "Estimation of the permeability of granular soils using neuro-fuzzy system," in *Proceedings of the AIAI Workshop*, pp. 333–342, Citeseer, Thessaloniki, Greece, April 2009.
- [18] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [19] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, pp. 18–22, 2002.
- [20] 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, Article ID 2218, 2020.
- [21] 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 Computers*, vol. 36, pp. 603–616, 2020.
- [22] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, pp. 217–222, 2005.
- [23] V. K. Singh, D. Kumar, P. S. Kashyap, P. K. Singh, A. Kumar, and S. K. Singh, "Modelling of soil permeability using different data driven algorithms based on physical properties of soil," *Journal of Hydrology*, vol. 580, Article ID 124223, 2020.
- [24] J. Dou, A. P. Yunus, D. T. Bui et al., "Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan," *The Science of the Total Environment*, vol. 662, pp. 332–346, 2019.

- [25] J. L. McClelland, D. E. Rumelhart, and P. R. Group, *Parallel Distributed Processing*, MIT press, Cambridge, MA, USA, 1986.
- [26] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [27] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer science & business media, Berlin, Germany, 2013.
- [28] B. Pradhan, "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS," *Computers & Geosciences*, vol. 51, pp. 350–365, 2013.
- [29] D. Tien Bui, B. Pradhan, O. Lofman, and I. Revhaug, "Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models," *Mathematical Problems in Engineering*, vol. 2012, Article ID 974638, 2012.
- [30] J. Tinoco, A. G. Correia, and P. Cortez, "Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns," *Computers and Geotechnics*, vol. 55, pp. 132–140, 2014.
- [31] D. T. Bui, T. A. Tuan, H. Klempe, B. Pradhan, and I. Revhaug, "Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree," *Landslides*, vol. 13, pp. 361–378, 2016.
- [32] P. Samui, "Support vector machine applied to settlement of shallow foundations on cohesionless soils," *Computers and Geotechnics*, vol. 35, pp. 419–427, 2008.
- [33] V. R. Kohestani, M. Hassanlourad, and A. Ardakani, "Evaluation of liquefaction potential based on CPT data using random forest," *Natural Hazards*, vol. 79, pp. 1079–1089, 2015.
- [34] S. S. Matin, L. Farahzadi, S. Makaremi, S. C. Chelgani, and G. H. Sattari, "Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest," *Applied Soft Computing*, vol. 70, pp. 980–987, 2018.
- [35] H. Hong, H. R. Pourghasemi, and Z. S. Pourtaghi, "Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models," *Geomorphology*, vol. 259, pp. 105–118, 2016.
- [36] A. Stumpf and N. Kerle, "Object-oriented mapping of landslides using Random Forests," *Remote Sensing of Environment*, vol. 115, pp. 2564–2577, 2011.
- [37] K. J. Archer and R. V. Kimes, "Empirical characterization of random forest variable importance measures," *Computational Statistics & Data Analysis*, vol. 52, pp. 2249–2260, 2008.
- [38] G. Biau, L. Devroye, and G. Lugosi, "Consistency of random forests and other averaging classifiers," *Journal of Machine Learning Research*, vol. 9, 2008.
- [39] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Proceedings of the Fourteenth International Conference on Machine Learning ICML'97*, pp. 296–304, San Francisco, CA, USA, July 1997.
- [40] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, pp. 23–69, 2003.
- [41] W.-Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, pp. 14–23, 2011.
- [42] I. Kononenko, *Estimating Attributes: Analysis and Extensions of RELIEF*, pp. 171–182, Springer, Berlin, Germany, 1994, European Conference on Machine Learning.
- [43] K. Kira and L. A. Rendell, *A Practical Approach to Feature Selection*, pp. 249–256, Elsevier, Amsterdam, Netherlands, 1992, Machine Learning Proceedings 1992.
- [44] T. G. Dietterich, "Machine-learning research," *AI Magazine*, vol. 18, p. 97, 1997.
- [45] 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, Article ID 4715, 2019.
- [46] B. T. Pham, M. D. Nguyen, D. Van Dao et al., "Development of artificial intelligence models for the prediction of Compression Coefficient of soil: an application of Monte Carlo sensitivity analysis," *The Science of the Total Environment*, vol. 679, pp. 172–184, 2019.
- [47] Q. H. Nguyen, H.-B. Ly, L. S. Ho et al., "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Mathematical Problems in Engineering*, vol. 2021, Article ID 4832864, 2021.
- [48] R. Walker, B. Indraratna, and C. Rujikiatkamjorn, "Vertical drain consolidation with non-Darcian flow and void-ratio-dependent compressibility and permeability," *Géotechnique*, vol. 62, pp. 985–997, 2012.
- [49] S. A. Berilgen, M. M. Berilgen, and I. K. Ozaydin, "Compression and permeability relationships in high water content clays," *Applied Clay Science*, vol. 31, pp. 249–261, 2006.
- [50] H.-D. Nguyen, V.-D. Pham, Q.-H. Nguyen et al., "An optimal search for neural network parameters using the Salp swarm optimization algorithm: a landslide application," *Remote Sensing Letters*, vol. 11, pp. 353–362, 2020.
- [51] Q. H. Nguyen, H.-B. Ly, L. S. Ho et al., "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Mathematical Problems in Engineering*, vol. 2021, Article ID 4832864, 15 pages, 2021.
- [52] V. Q. Tran and H. Q. Do, "Prediction of California bearing ratio (CBR) of stabilized expansive soils with agricultural and industrial waste using light gradient boosting machine," *Journal of Science and Transport Technology*, vol. 1, pp. 1–9, 2021.
- [53] C. Qi, L. Guo, H.-B. Ly, H. Van Le, and B. T. Pham, "Improving pressure drops estimation of fresh cemented paste backfill slurry using a hybrid machine learning method," *Minerals Engineering*, vol. 163, Article ID 106790, 2021.