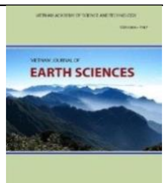
Vietnam Journal of Earth Sciences, 42(3), 255–264, Doi: 10.15625/0866-7187/42/3/15342

(VAST)

Vietnam Academy of Science and Technology

Vietnam Journal of Earth Sciences

http://www.vjs.ac.vn/index.php/jse



Prediction of soil unconfined compressive strength using Artificial Neural Network model

Hoang-Anh Le[1], Thuy-Anh Nguyen[1]*, Duc-Dam Nguyen[1], Indra Prakash[2]

[1]*University of Transport Technology, Hanoi 100000, Vietnam*
[2]*Bhaskarcharya Institute for Space Applications and Geo-Informatics (BISAG), Gandhinagar 382002, India*

Received 27 May 2020; Received in revised form 6 August 2020; Accepted 5 September 2020

ABSTRACT

The main objective of the present study is to apply Artificial Neural Network (ANN), which is one of the most popular machine learning models, to accurately predict the soil unconfined compressive strength (q_u) for the use in designing of foundations of civil engineering structures. For the development of model, data of 118 soil samples were collected from Long Phu 1 power plant project, Soc Trang Province, Vietnam. The database of physicomechanical properties of soils was prepared for the model study, where 70% data was used for the training and 30% for the testing of the model. Standard statistical indices, namely Root Mean Squared Error (RMSE) and Pearson Correlation Coefficient (R) were used in the validation of the model's performance. In addition, Partial Dependence Plots (PDP) was used to evaluate the importance of the input variables used for modeling. Results showed that the ANN model performed well for the prediction of the q_u (RMSE = 0.442 and R = 0.861). The PDP analysis showed that the liquid limit is the most important input factor for modeling of the q_u. The present study demonstrated that the ANN is a promising tool that can be used for quick and accurate prediction of the q_u, which can be used in designing the civil engineering structures like bridges, buildings, and powerhouses.

Keywords: soil unconfined compressive strength; Artificial Neural Network; machine learning.

1. Introduction

Soil unconfined compressive strength (q_u) has a vital role in designing and construction of the structures located in soil (Das and Sobhan, 2013). Normally, this parameter is measured directly in the laboratory using the unconfined compression tests (Das and Sobhan, 2013). However, sometimes, these tests face a formidable problem in getting the representative samples, which can significantly affect the accuracy of the tests. In addition, these tests are often time-consuming and involve high cost. Thus, the alternative methods should be developed and applied to accurately estimate the q_u based on other parameters in a time and cost-effective manner.

Traditionally, there are several empirical equations that have been explored and applied to alternatively predict the soil strength considering different independent variables. Busscher et al. (Busscher et al., 1987) estimated the soil strength in correlation with

*Corresponding author, Email: anhnt@utt.edu.vn

the bulk density and soil water content. Yılmaz (Yılmaz, 2000) predicted the clayey soil strength using a liquidity index. In addition, the soil-water characteristic curve has also been popularly used to predict the strength of the soil. In general, empirical equation-based approaches pave a useful alternative tool to estimate the soil strength, which can also be applied to predict the q_u. However, this approach, in some cases, considers a limited number of independent variables, sometimes, not representing the real site condition of the soil. Thus, this can affect the accuracy and reliability of predictive outcomes.

In recent years, soft computing-based machine learning or artificial intelligence approaches have emerged for modeling and in solving real-world problems (Ahmadlou et al., 2019; Dao et al., 2020a; Dou et al., 2020; Le et al., 2020; Ly et al., 2019b; Pham et al., 2020a, 2020b; Phong et al., 2019). The main principle of these approaches is based on computational algorithms, which enable the remarkable capability of the human brain in order to learn and discover the real problems related to data in an environment of uncertainty and imprecision (Dao et al., 2020b; Kalkan et al., 2009). In the field of geotechnical engineering, several studies have been effectively performed to predict the critical properties of soil materials (Pham et al., 2020c). Besides, Pham et al. (Pham et al., 2018) investigated various machine learning methods for the prediction of shear strength of the soil. Kirts et al. (Kirts et al., 2018) used Support Vector Machines (SVM) for prediction of soil compressibility. In another study, soil bulk density was predicted using a popular machine learning model, namely a decision tree (Bondi et al., 2018). Moayedi et al. (Moayedi et al., 2019) developed a novel hybrid machine learning models of Artificial Neural Network (ANN) and various optimization techniques namely Whale Optimization Algorithm (WOA), Dragonfly Algorithm (DA), or Invasive Weed Optimization (IWO) for accurate prediction of soil shear strength. In general, machine learning models are considered promising tools for accurate prediction of the soil properties.

In this study, the main objective is to apply the ANN-one of the most popular machine learning models to predict the q_u using the database collected from Long Phu 1 Thermal power plant project, Soc Trang Province,Vietnam. At present, the use of machine learning approaches is still limited in the prediction of soil strength (Das et al., 2011; Kalkan et al., 2009; Narendra et al., 2006). Various quantitative statistical indices, namely Pearson Correlation Coefficient (R) and Root Mean Squared Error (RMSE) were used in the validation of the model's performance. Matlab software was used in data processing and modeling.

2. Methodology

2.1. Construction of database

In this study, the database of 118 soil samples, including experimental results, was collected from Long Phu 1 power plant project, Vietnam (https://www.power-technology.com/projects/long-phu-1-thermal-power-plant-soc-trang-province/). Seven soil parameters were considered in the modeling study. Six parameters, namely clay content (%), void ratio, liquid limit (%), moisture content (%), plastic limit (%), and specific gravity, were used as input variables, whereas the q_u was considered as an output variable. The values of these variables were determined at the project laboratory as per standard laboratory procedures (Das and Sobhan, 2013) (Fig. 1). The results of the input and out variables are presented in Fig. 2.

For the simulation of model, 70% of data was randomly extracted to generate the training dataset, whereas 30% of the remaining data was used to generate the testing dataset. The training dataset was used to train and construct the ANN model, whereas the testing dataset was used to validate the predictive capability of the ANN model.

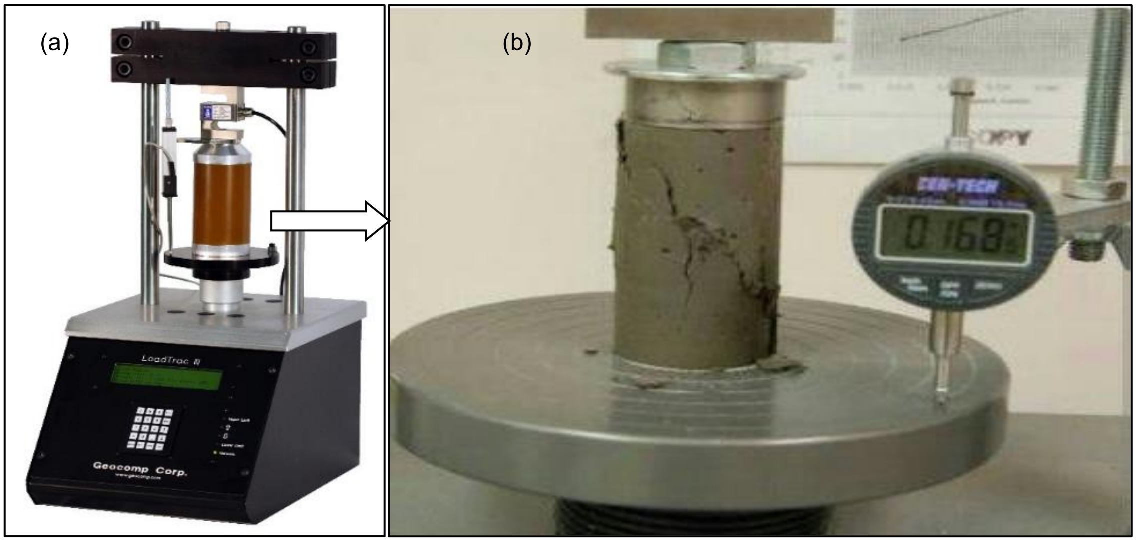


Figure 1. Soil unconfined compressive tests used in this study: (a) unconfined compressive equipment and (b) tested samples (Source: http://www.mocivilengineering.com/2016/08/soil-investigation.htmland https://www.quora.com/What-is-the-difference-between-confined-and-unconfined-in-soil)

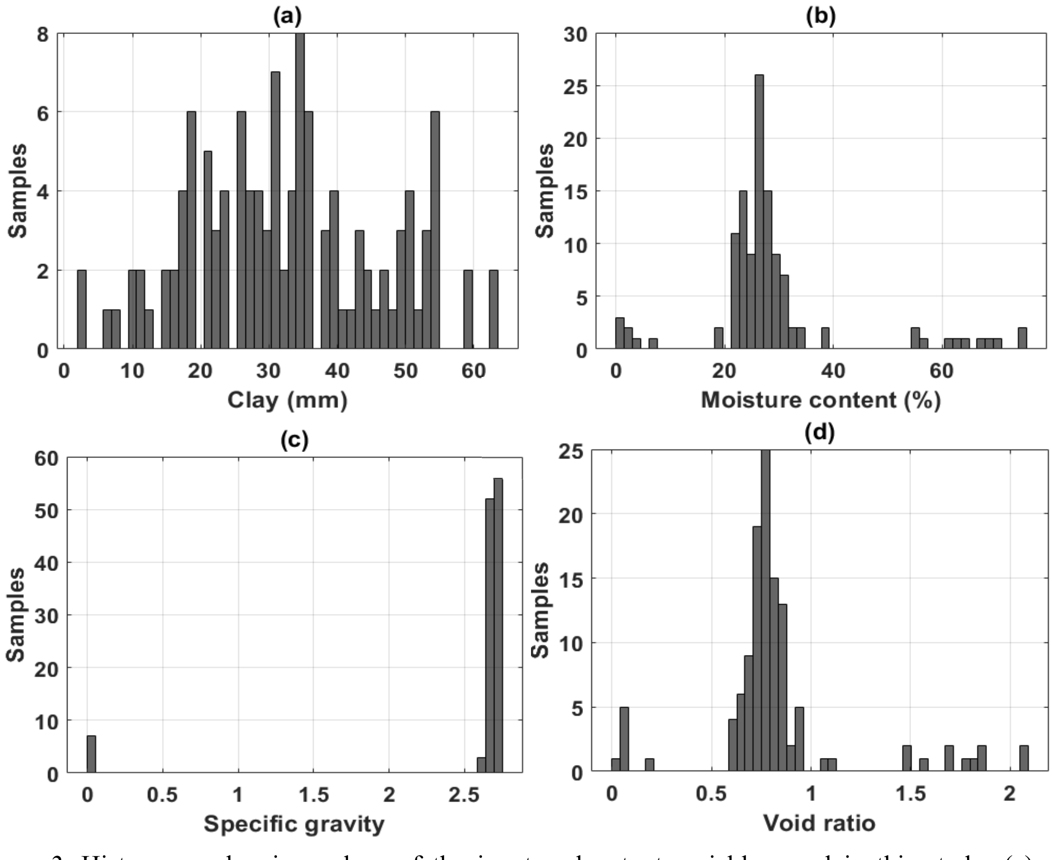
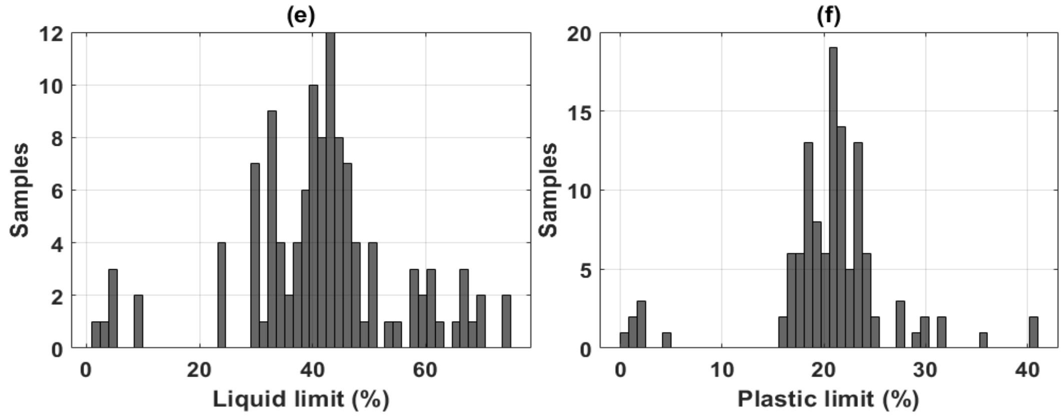
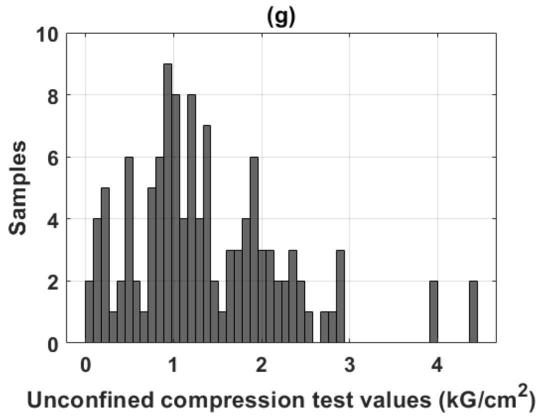


Figure 2. Histograms showing values of the input and output variables used in this study: (a) clay; (b) moisture content; (c) Specific gravity; (d) Void ratio; (e) Liquid limit; (f) Plastic limit; and (g) unconfined compressive strength (q_u)





← *Figure 2*. Continue

2.2. Model Used: Artificial Neural Network (ANN)

The ANN, which is one of the most popular machine learning models, is based on the behavior of the biological neural network of the human brain (Du et al., 2017). In the ANN method, a backpropagation neural network is usually utilized for analysis of regression and classification problems (Singh, 2012). In the ANN input, hidden and output layers are used to construct the networks which are connected by the neurons called network nodes (Abad et al., 2018; Khandelwal et al., 2018). The input layer includes all input variables, the output layer includes output variable, and the hidden layer(s) includes an activation function which is used to analyze the hidden relationship between input variables and output variable (Armaghani et al., 2019; Bejarbaneh et al., 2018). In this study, the ANN was applied to predict the q_u (output variable) based on the input variables (clay content (%), void ratio, liquid limit (%), moisture content (%), plastic limit (%), and specific gravity). and sigmoid activation function was used in the two hidden layers with [12-8] neurons.

To validate performance of the ANN model, two popular quantitative statistical indices, namely R and RMSE were used on both training and testing datasets. Following are the equations used to calculate these indices (Ly et al., 2019a; Nguyen et al., 2019):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(d_{o,i} - d_{t,i}\right)^2} \quad (1)$$

$$R = \sqrt{\frac{\sum_{i=1}^{n}\left(d_{o,i} - \bar{d}_o\right)\left(d_{t,i} - \bar{d}_t\right)}{\sqrt{\sum_{i=1}^{n}\left(d_{o,j} - \bar{d}_o\right)^2 \sum_{i=1}^{n}\left(d_{t,i} - \bar{d}_t\right)^2}}} \quad (2)$$

where n is the number of instances in the database, d_o and $\bar{d}_0$ are defined as the actual value and average actual value, d_t, and $\bar{d}_t$ are the predicted value and the average predicted value. Quantitatively, higher values of R indicate a better predictive capability of the model and vice versa, whereas lower values of RMSE show better performance of the model and vice versa.

3. Results and discussions

3.1. Prediction performance of the ANN model

Figure 3 shows the comparison between estimated ANN model results (output) and actual experimental results (target) for phases of training (a) and testing (b). Errors between output and the target results are depicted in Fig. 4. Analysis of results reveals that the ANN model could estimate the q_u almost accurately for most of the results of the experimental sample. The errors were relatively small, that is close to 0. Figure 4 shows frequency versus error values for the training and testing datasets. During the training and testing phases, the errors were mostly in the range of [-0.5, 0.5]. These values showed that the prediction capability of the proposed ANN model was excellent within the acceptable error value.

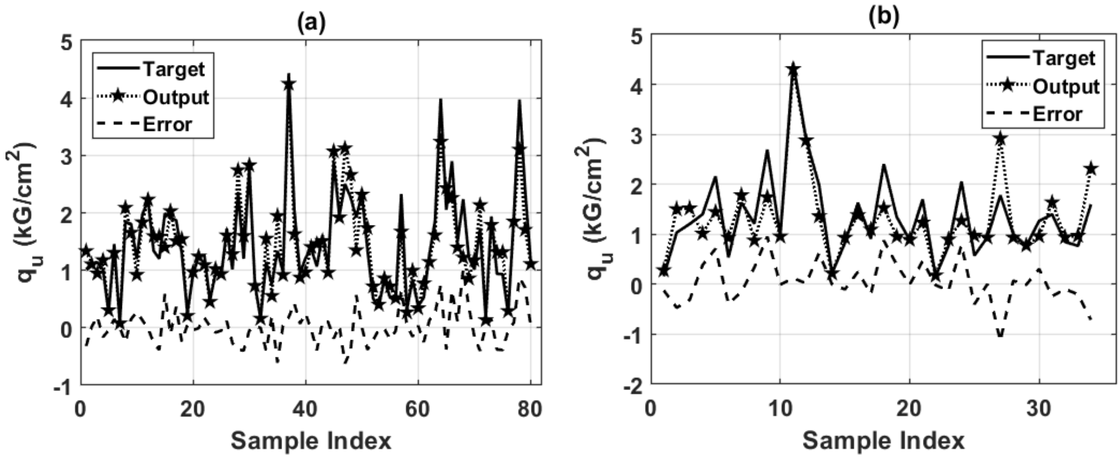


Figure 3. Compressive strength q_u of soil by ANN model (a) Training; (b) Testing

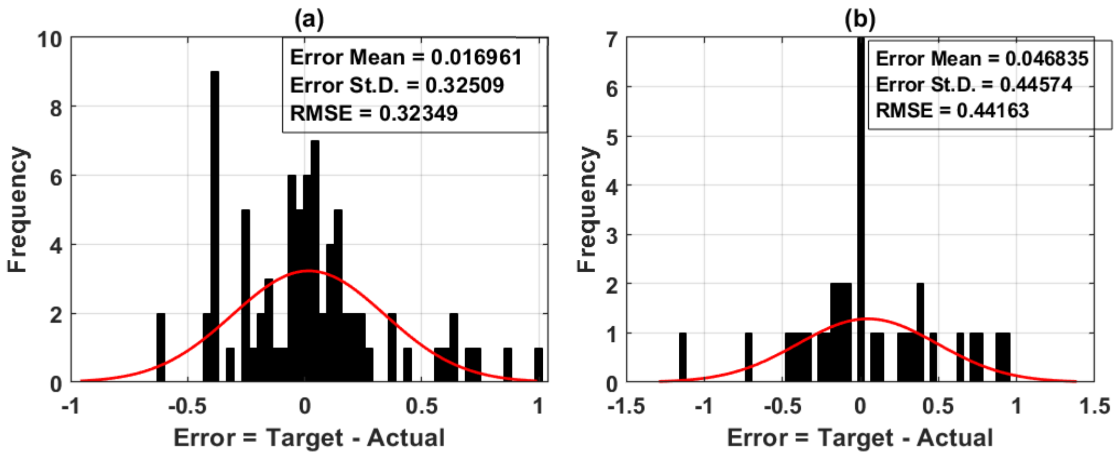


Figure 4. Errors of the ANN model (a) training; (b) testing

The regression analysis graphs between the predicted and actual values for the training, testing, and all (total) dataset are given in Fig. 5. The values of R were 0.928, 0.861, and 0.908 for the training, testing, and all (total) datasets, respectively. Regression analysis also showed that the ANN model is a good predictor in the present case. It is reasonable as the ANN is one of the best machine learning models used in prediction. It has several advantages such as (i) ANN algorithm is independent on the statistical distribution of the used data and (ii) it is objective in assigning weights to input variables, thus, it involves a minimum of human interference (Pradhan and Lee, 2010). The validation results are in good agreement with previously published works (Das et al., 2011; Mozumder and Laskar, 2015).

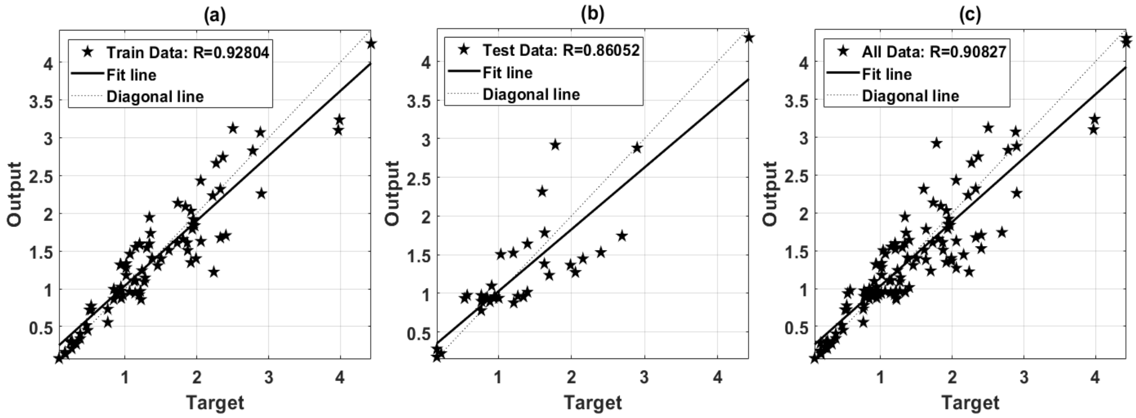


Figure 5. Regression graphs for different datasets: (a) training; (b) testing; and (c) all (total) data

3.2. Importance of input factors using Partial Dependence Plots (PDP)

The dependence between the predicted value by the ANN model and each selected input variable was estimated using Partial Dependence Plots (PDP) (Friedman et al., 2001). In this investigation, PDP estimated 6 input variables used in the ANN model, namely clay content, void ratio, liquid limit, moisture content, plastic limit, and specific gravity (Fig. 6). For clay content, PDP value varied from 0.6 to 2.4, from 3 to -1 for void ratio, from 2.0 to 0.6 for moisture content, and from 0.25 to 1.6 for plastic limit. Finally, when varying the specific gravity, the value of q_u was found slightly fluctuating, from 0.4 to 1.5. For the liquid limit, the predicted output ranged in -3.2 to 1.5. Overall, a variation equal to 4.7 was found for the liquid limit. Thus, it could be concluded that the latter was the most critical variable in the prediction of the q_u. Thus, based on the amplitude of the variation of PDP values, the order of the influence of variables on q_u are liquid limit, followed by void ratio, clay content, moisture content, plastic limit, and specific gravity. In other studies (Cokca et al., 2004; Spoor and Godwin, 1979) also stated that the inputs that depended on water content were found as the critical variables for the q_u prediction as the presence of water reduces the angle of friction and cohesion among soil particles.

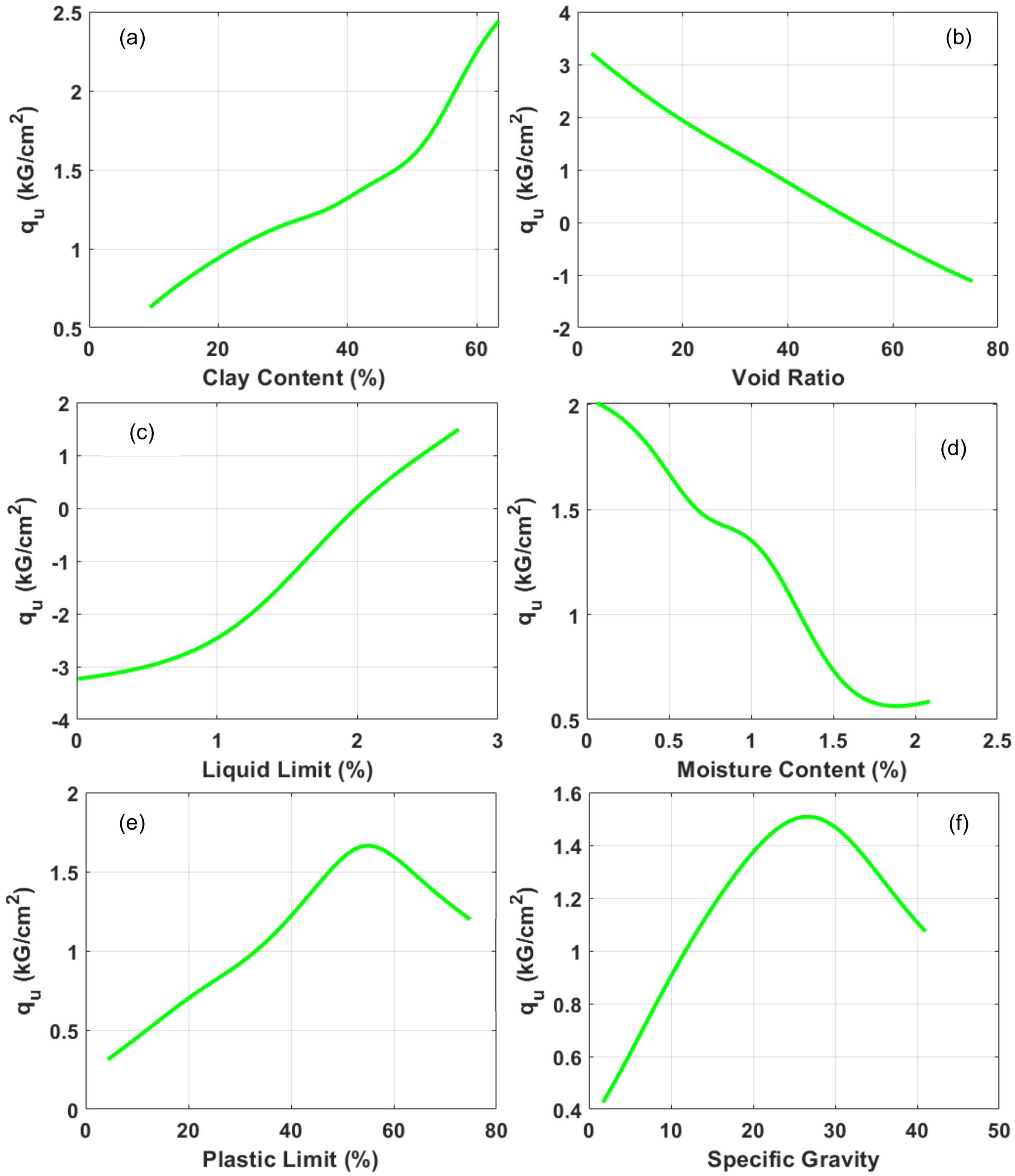


Figure 6. Partial dependence plots (PDP) of 6 input variables used in this investigation: (a) clay content, (b) void ratio, (c) liquid limit, (d) moisture content, (e) plastic limit, and (f) specific gravity

4. Conclusions

Machine learning is known as an advanced soft computing technique used to solve effectively many real-world problems. In this study, one of the popular machine learning techniques, namely ANN, was applied to predict one of the essential geotechnical parameters, namely the q_u. In this study, laboratory results of the six physico-mechanical properties (clay content, void ratio, liquid limit, moisture content, plastic limit, and specific gravity) of 118 soil samples collected from the Long Phu 1 power plant project, Vietnam were considered as input variables in ANN model study and the q_u as an output variable parameter. Various quantitative statistical indices, namely RMSE and R were used in the validation of the model's performance.

Validation results showed that the ANN performed well for the prediction of q_u (RMSE = 0.442, and R = 0.861). In addition, the PDP analysis was applied to validate the importance of the input variables, and the results showed that the liquid limit is the most crucial factor for the prediction of q_u.

It can be concluded that the ANN is a promising tool that can be used for quick and accurate prediction of the q_u instead of a number of laboratory tests for the designing of the civil engineering structures. However, its application should be validated by conducting a few more tests at different sites to confirm the accuracy of observed and estimated values (q_u) and for creating a larger database of physico-mechanical properties for further use in the model studies at other sites in Vietnam as well as other areas of the world.

Acknowledgments

This research is funded by Ministry of Transport, project title "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 number DT 203029.

References

Abad S.V.A.N.K., Yilmaz M., Armaghani D.J., Tugrul A., 2018. Prediction of the durability of limestone aggregates using computational techniques. Neural Computing and Applications, 29, 423–433.

Ahmadlou M., Karimi M., Alizadeh S., Shirzadi A., Parvinnejhad D., Shahabi H., Panahi M., 2019. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). Geocarto International, 34, 1252–1272.

Armaghani D.J., Koopialipoor M., Marto A., Yagiz S., 2019. Application of several optimization techniques for estimating TBM advance rate in granitic rocks. Journal of Rock Mechanics and Geotechnical Engineering, 11, 779–789.

Bejarbaneh B.Y., Bejarbaneh E.Y., Fahimifar A., Armaghani D.J., Abd Majid M.Z., 2018. Intelligent modelling of sandstone deformation behaviour using fuzzy logic and neural network systems. Bulletin of Engineering Geology and the Environment, 77, 345–361.

Bondi G., Creamer R., Ferrari A., Fenton O., Wall D., 2018. Using machine learning to predict soil bulk density on the basis of visual parameters: Tools for in-field and post-field evaluation. Geoderma, 318, 137–147.

Busscher W.J., Spivey L.D., Campbell R.B., 1987. Estimation of soil strength properties for critical rooting conditions. Soil and Tillage Research, 9, 377–386.
https://doi.org/10.1016/0167-1987(87)90062-6.

Cokca E., Erol O., Armangil F., 2004. Effects of compaction moisture content on the shear strength of an unsaturated clay. Geotechnical and Geological Engineering, 22, 285.
https://doi.org/10.1023/B:GEGE.0000018349.40866.3e.

Dao D.V., Adeli H., Ly H.-B., Le L.M., Le V.M., Le T.-T., Pham B.T., 2020a. A Sensitivity and Robustness Analysis of GPR and ANN for High-Performance Concrete Compressive Strength Prediction Using a Monte Carlo Simulation. Sustainability, 12, 830. https://doi.org/10.3390/su12030830.

Dao D.V., Ly H.-B., Vu H.-L.T., Le T.-T., Pham B.T., 2020b. Investigation and Optimization of the C-ANN Structure in Predicting the Compressive Strength of Foamed Concrete. Materials, 13, 1072. https://doi.org/10.3390/ma13051072.

Das B.M., Sobhan K., 2013. Principles of geotechnical engineering. Cengage learning.

Das S.K., Samui P., Sabat A.K., 2011. Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil. Geotechnical and Geological Engineering, 29, 329–342.

Dou J., Yunus A.P., Merghadi A., Shirzadi A., Nguyen H., Hussain Y., Avtar R., Chen Y., Pham B.T., Yamagishi H., 2020. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. Science of the total environment, 720, 137320.

Du Y., Chen Z., Zhang C., Cao X., 2017. Research on axial bearing capacity of rectangular concrete-filled steel tubular columns based on artificial neural networks. Frontiers of Computer Science, 11, 863–873. https://doi.org/10.1007/s11704-016-5113-6.

Friedman J., Hastie T., Tibshirani R., 2001. The elements of statistical learning. Springer series in statistics New York.

Kalkan E., Akbulut S., Tortum A., Celik S., 2009. Prediction of the unconfined compressive strength of compacted granular soils by using inference systems. Environmental Geology, 58, 1429–1440.

Khandelwal M., Marto A., Fatemi S.A., Ghoroqi M., Armaghani D.J., Singh T., Tabrizi O., 2018. Implementing an ANN model optimized by genetic algorithm for estimating cohesion of limestone samples. Engineering with Computers, 34, 307–317.

Kirts S., Panagopoulos O.P., Xanthopoulos P., Nam B.H., 2018. Soil-compressibility prediction models using machine learning. Journal of Computing in Civil Engineering, 32, 04017067.

Le T.-T., Pham B.T., Ly H.-B., Shirzadi A., Le L.M., 2020. Development of 48-hour Precipitation Forecasting Model using Nonlinear Autoregressive Neural Network, in: Ha-Minh C., Dao D.V., Benboudjema F., Derrible S., Huynh D.V.K., Tang A.M. (Eds.), CIGOS 2019. Innovation for Sustainable Infrastructure, Lecture Notes in Civil Engineering. Springer Singapore, 1191–1196.

Ly H.-B., Le L.M., Phi L.V., Phan V.-H., Tran V.Q., Pham B.T., Le T.-T., Derrible S., 2019a. Development of an AI Model to Measure Traffic Air Pollution from Multisensor and Weather Data. Sensors, 19, 4941. https://doi.org/10.3390/s19224941.

Ly H.-B., Le T.-T., Le L.M., Tran V.Q., Le V.M., Vu H.-L.T., Nguyen Q.H., Pham B.T., 2019b. Development of Hybrid Machine Learning Models for Predicting the Critical Buckling Load of I-Shaped Cellular Beams. Applied Sciences, 9, 5458. https://doi.org/10.3390/app9245458.

Moayedi H., Tien Bui D., Dounis A., Kok Foong L., Kalantar B., 2019. Novel nature-inspired hybrids of neural computing for estimating soil shear strength. Applied Sciences, 9, 4643.

Mozumder R.A., Laskar A.I., 2015. Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using artificial neural network. Computers and Geotechnics, 69, 291–300.

Narendra B., Sivapullaiah P., Suresh S., Omkar S., 2006. Prediction of unconfined compressive strength of soft grounds using computational intelligence techniques: A comparative study. Computers and Geotechnics, 33, 196–208.

Nguyen H.-L., Le T.-H., Pham C.-T., Le T.-T., Ho L.S., Le V.M., Pham B.T., Ly H.-B., 2019. Development of Hybrid Artificial Intelligence Approaches and a Support Vector Machine Algorithm for Predicting the Marshall Parameters of Stone Matrix Asphalt. Applied Sciences, 9, 3172. https://doi.org/10.3390/app9153172.

Pham B.T., Hoang T.-A., Nguyen D.-M., Bui D.T., 2018. Prediction of shear strength of soft soil using machine learning methods. Catena, 166, 181–191.

Pham B.T., Le L.M., Le T.-T., Bui K.-T.T., Le V.M., Ly H.-B., Prakash I., 2020a. Development of advanced artificial intelligence models for daily rainfall prediction. Atmospheric Research, 237, 104845. https://doi.org/10.1016/j.atmosres.2020.104845.

Pham B.T., Nguyen M.D., Ly H.-B., Pham T.A., Hoang V., Van Le H., Le T.-T., Nguyen H.Q., Bui G.L., 2020b. Development of Artificial Neural Networks for Prediction of Compression Coefficient of Soft Soil, in: Ha-Minh C., Dao D.V., Benboudjema F., Derrible S., Huynh D.V.K., Tang A.M. (Eds.), CIGOS 2019. Innovation for Sustainable Infrastructure, Lecture Notes in Civil Engineering. Springer Singapore, 1167–1172.

Pham B.T., Nguyen-Thoi T., Ly H.-B., Nguyen M.D., Al-Ansari N., Tran V.-Q., Le T.-T., 2020c. Extreme Learning Machine Based Prediction of Soil Shear Strength: A Sensitivity Analysis Using Monte Carlo Simulations and Feature Backward Elimination. Sustainability, 12, 2339.
https://doi.org/10.3390/su12062339.

Phong T.V., Phan T.T., Prakash I., Singh S.K., Shirzadi A., Chapi K., Ly H.-B., Ho L.S., Quoc N.K., Pham B.T., 2019. Landslide susceptibility modeling using different artificial intelligence methods: a case study at Muong Lay district, Vietnam. Geocarto International, 1–24. https://doi.org/10.1080/10106049.2019.1665715.

Pradhan B., Lee S., 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environmental Modelling & Software, 25, 747–759.

Singh S., 2012. Backpropagation Learning Algorithm Based on Levenberg Marquardt Algorithm, Computer Science & Information Technology. https://doi.org/10.5121/csit.2012.2438.

Spoor G., Godwin R.J., 1979. Soil Deformation and Shear Strength Characteristics of Some Clay Soils at Different Moisture Contents. Journal of Soil Science, 30, 483–498. https://doi.org/10.1111/j.1365-2389.1979.tb01003.x.

Yılmaz I., 2000. Evaluation of shear strength of clayey soils by using their liquidity index. Bulletin of engineering Geology and the Environment, 59, 227–229.