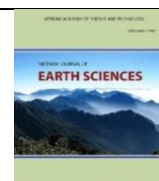




Vietnam Academy of Science and Technology

Vietnam Journal of Earth Sciences

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



Accuracy assessment of extreme learning machine in predicting soil compression coefficient

Hai-Bang Ly¹, Panagiotis G. Asteris², Binh Thai Pham^{1*}

¹*University of Transport Technology, Hanoi 100000, Vietnam*

²*Computational Mechanics Laboratory, School of Pedagogical and Technological Education, Heraklion, GR 14121, Athens, Greece*

Received 7 February 2020; Received in revised form 18 April 2020; Accepted 25 June 2020

ABSTRACT

The compression coefficient (C_c) is an important soil mechanical parameter that represents soil compressibility in the process of consolidation. In this study, a machine learning derived model, namely extreme learning algorithm (ELM), was used to predict the C_c of soil. A total of 189 experimental results were used and randomly divided to construct the training and testing parts for the development and validation of ELM. Monte Carlo approach was applied to take into account the random sampling of samples constituting the training dataset. A number of 13 input parameters reflecting the experiment were used as the input variables to predict the output C_c . Several statistical criteria, such as mean absolute error (MAE), root mean square error (RMSE), correlation coefficient (R) and the Monte Carlo convergence estimator were used to assess the performance of ELM in predicting the C_c of soil. The results showed that ELM had a strong capacity to predict the C_c of soil, with the R value > 0.95 . The convergence of results, as well as the capability of ELM were fully investigated to understand the advantage of using ELM as a predictor.

Keywords: Compression coefficient; extreme machine learning; Monte Carlo simulations.

©2020 Vietnam Academy of Science and Technology

1. Introduction

Machine learning, which is a part of the artificial intelligence approach, has been effectively applied in various fields to solve a lot of problems in the real world (Armaghani et al., 2016; Collins and Moons, 2019; Ly et al., 2019c; Michie et al., 1994; Mohamad et al., 2018, 2016; Thanh et al., 2020). This approach is based on the computational

mathematic algorithm to analyze the relationship between input and output variables; therefore, it can solve a complex problem with large data (Dao et al., 2020a; Khandelwal et al., 2018; Khosravi et al., 2019; Le et al., 2020; Ly et al., 2019b; Pham et al., 2020b, 2020a). Compared with traditional approaches like linear regression or weighted methods, machine learning is known as more subjective and effective approaches for better performance (Chen et al., 2019; Ly et al., 2019e).

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

In geotechnical engineering, machine learning has also been applied effectively in solving both classification and regression problems. Dao et al. (Dao et al., 2020b) developed and compared deep learning and various machine learning models (quadratic discriminant analysis, Fisher's linear discriminant analysis, and multi-layer perceptron neural network) to solve a classification problem of spatial landslide prediction, and stated that machine learning and deep learning are potential approaches for quick and accurate prediction of landslides. In another work, machine learning models have been developed and applied to predict important properties of soil materials as regression problems such as soil compression coefficient using HHO-ANN and GOA-ANN ensembles (Moayedi et al., 2020a) and LCA-ANFIS hybrid model (Moayedi et al., 2020b), soil aggregate stability (Rivera and Bonilla, 2020), soil shear strength (Bui et al., 2019), soil consolidation coefficient (Pham et al., 2019a). In general, these mentioned studies show a great potential application of machine learning models for the prediction of geotechnical problems, including the prediction of soil properties.

In this study, the main aim is to use Extreme Learning Machine (ELM) for the prediction of an important property of soil, namely compression coefficient (C_c), which can help in quick and accurate determination of this parameter for reducing the cost and time for construction. ELM is a well-known machine learning model that has been applied effectively in solving real-world problems (Huang et al., 2006; Nizar et al., 2008), but its application is still limited in geotechnical problems. For this, a total of 189 experimental results were used and randomly divided to construct the datasets for the training and testing of the model. Monte Carlo approach was applied to evaluate the effect of the random sampling strategy of samples on the

performance of the prediction model. Various statistical criteria, such as mean absolute error, root mean square error, the correlation coefficient (R) and the Monte Carlo convergence estimator were used to validate the predictive capability of the model. Matlab software and application was used for data processing and modeling.

2. Data collection and preparation

In this study, a number of 189 experimental data were collected from the national highway project in Hai Phong and Ninh Binh (Pham et al., 2019b). The samples were composed of soft clayed soil, and the tests were performed in the laboratory to determine 13 input variables and one output variable. The input variables contained the depth of sample, clay content in soil, moisture content, bulk, and dry soil densities, specific gravity, void ratio, porosity, saturation degree, liquid and plastic limits, plasticity, and liquidity indexes. The target variable was the Compression Coefficient of soil (denoted as C_c), which reflects the compressibility of soil in the consolidation process. It is considered as one of the important soil physicommechanical parameters of soil usually utilized in the determination of the primary consolidation settlement (Das and Sobhan, 2013). For the description of the input and output variables, the readers could refer to the work of Das and Sobhan (Das and Sobhan, 2013). The statistical analysis of the dataset used in this study can be found in our previous work (Pham et al., 2019b).

In this work, the dataset was divided into two parts: the training and testing one. The training dataset contained 132 samples (70%) and used for the training phase of ELM algorithm, whereas the testing dataset contained 57 samples (30%), served to verify the performance of ELM model. In addition, the whole dataset was scaled in the range of [0, 1], as suggested in many machine learning

problems to reduce numerical bias. Eq. (1) was used to scale the dataset:

$$\chi_{scaled} = \frac{2(\chi - \lambda)}{\mu - \lambda} - 1 \quad (1)$$

where λ and μ are the minimum and maximum values of given variables, and χ is the value of the variable to be scaled.

3. Methods used

3.1. Extreme learning machine (ELM)

Extreme learning machine (ELM) algorithm is basically a feedforward neural network and often used for classification, regression, clustering, sparse approximation, compression, and feature learning with a single layer or multiple layers of hidden nodes. The basis of ELM lied in the generation of hidden nodes, which can be randomly assigned, and no update is performed during the simulation process. Otherwise, the parameters of hidden nodes (weights that connect input variables to the hidden nodes and their biases) were not tuned or can be inherited from the ancestors in several cases. According to the creators, ELM can produce a good performance and the ability to learn much faster than backpropagation networks. For the detailed development of ELM algorithm, the readers are referred to the literature (Huang et al., 2011, 2006, 2004). To validate the performance of ELM model, various statistical indexes such as mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R) were used. Detail description of these indexes is presented in previously published works (Ly et al., 2019a; Mohamad et al., 2012; Momeni et al., 2015; Nguyen et al., 2019; Pham et al., 2019b).

3.2. Monte Carlo approach

Monte Carlo (MC) method is widely applied in problems related to civil engineering to

take into account the variability of parameters (Dao et al., 2020a; Ly et al., 2019d; Qi et al., 2020). The main idea of MC lies in the computation of the target using as many realizations as possible, which repeats the random sampling of input variables (Dao et al., 2020c; Nguyen et al., 2020; Pham et al., 2020c). The concept of using the MC method consists in propagating the variability of input parameters on the outputs using a predefined model and using statistical analysis of the output results to investigate the sensitivity or robustness of such a predefined model. By doing so, the input space reflects the random combinations of variables, where each variable obeys a probability density function, represented by its own variability. With the variability of the input space, the obtained outputs possess a distribution of response and the corresponding statistical properties. Statistical analysis needs to be performed to assess the accuracy of the model or the sensitivity of inputs. The higher the number of combinations in MC simulation, the more reliable the outputs. Nevertheless, to achieve a reasonable time consuming and reach the appropriate number of MC simulations, statistical convergence analysis of the mean value of a random variable is defined by the following equation:

$$NMC \mapsto \text{Conv.}(R|NMC) = \frac{1}{\bar{R}} \frac{1}{NMC} \sum_{j=1}^{NMC} R_j \quad (2)$$

where \bar{R} is the mean value of the random variable R , NMC is the number of Monte Carlo simulations, R_j is the value of the j^{th} observation of R .

4. Results and Discussion

4.1. Convergence of results

The prediction convergence results in the function of RMSE, MAE, and R are plotted in Fig. 1. A fluctuation up to 15% around the average values of RMSE and MAE was

observed, whereas that of the case of R was smaller (i.e., 5%). The predicted output stabilized in the 5% range after about 20 Monte Carlo simulations and smaller than 3% when NMC was higher than 300. Considering the R criterion, the NMC required for the stabilization of results below 1% was 20, whereas to obtain smaller fluctuation (i.e.,

<0.5%), a number of 300 NMC was needed. This result showed that ELM somehow outperformed support vector machine (SVM) algorithm in predicting the Cc of soft soil, as the NMC required to achieve a 1% converged result was 20, compared with 50 in the literature (Pham et al., 2019b).

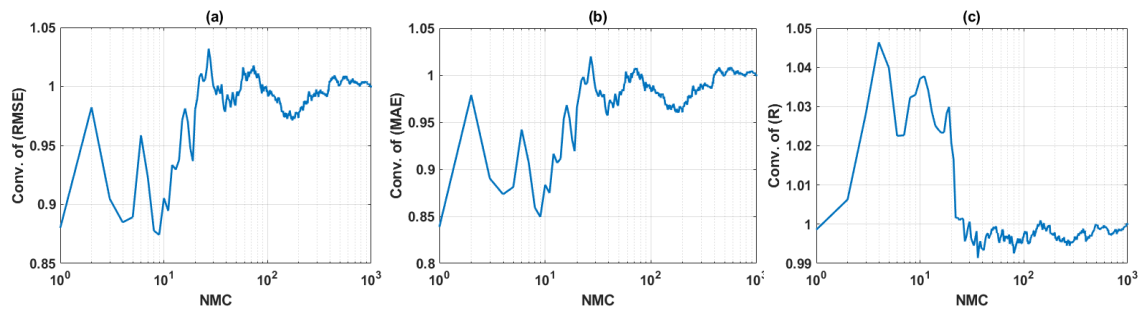


Figure 1. Normalized statistical convergence over 1000 NMC using ELM for: (a) RMSE; (b) MAE; (c) R

4.2. Prediction capability

The probability density distributions of RMSE, MAE, and R values are presented in Fig. 2. It is observed that all distributions were asymmetric. The highest probability values of RMSE, MAE, and R were observed at 0.1467, 0.1166, and 0.8873, respectively. With respect to R criterion, values between R = 0.4 to R = 0.7 were noticed (Fig. 3c), showing that ELM exhibited a strong fluctuation in predicting Cc of soft soil. Similar observations were noticed

with RMSE and MAE (Fig. 3a, b). The average values of RMSE, MAE, and R over 1000 simulations were 0.1743, 0.1450, and 0.8568, respectively. In conclusion, from overall statistical analysis, ELM algorithm was an excellent predictor to predict Cc. As in ELM, the weights and biases were randomly generated, an optimization algorithm might be needed to stabilize and enhance the prediction performance.

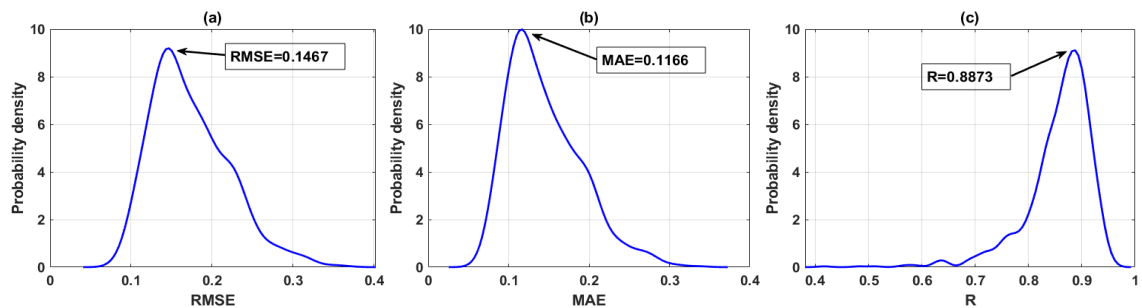


Figure 2. Probability distribution over 1000 NMC using ELM for the prediction of Cc for: (a) RMSE; (b) MAE; (c) R

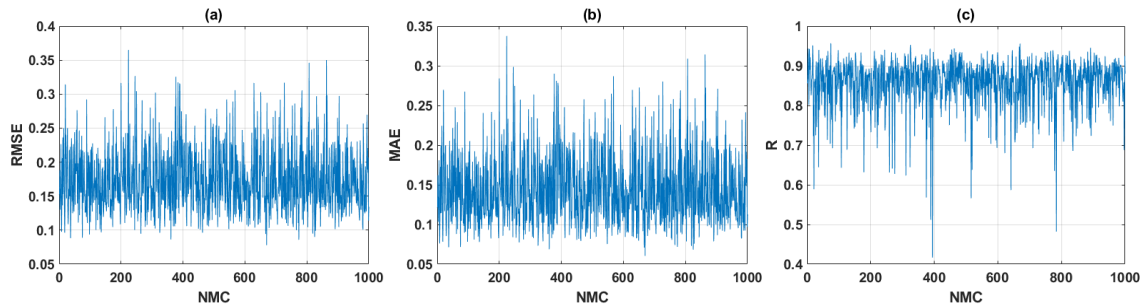


Figure 3. Performance results of ELM over 1000 NMC for: (a) RMSE; (b) MAE; (c) R

4.3. Typical prediction results

In this section, a typical result of the prediction of C_c using ELM is presented. Obviously, the experimental and predicted C_c were highly correlated for both the training part (Fig. 4a) and the testing dataset (Fig. 4b). Regarding the testing data, lower values of error were found (error mean = -0.0517, error StD = 0.0728), whereas the training part exhibited errors of mean = -0.0725 and StD = 0.0982. The distribution of error with respect to the training and testing datasets was shown in Fig. 5. The maximum error was found at 20% in the testing set and 30% in the training set.

Validation results by the linear fit lines, its equations, and the R values are given in Fig. 6 for the training and testing datasets. Two

equations relating the predicted and actual values of C_c were obtained as $y = 1.1 \cdot x + 0.034$ for the training part and $y = x + 0.0520$ for the testing one. The values of R were computed as 0.9158 and 0.9566 for the training, testing datasets, respectively, showing an excellent accuracy of ELM in predicting C_c of soft soil.

The performance of ELM used in this study was compared with previously published work presented in Pham et al. (Pham et al., 2019b) and observed that the performance of ELM ($R^2 = 0.915$) is slightly lower than the performance of Artificial Neural Network (ANN) ($R^2 = 0.9499$), Support Vector Machines (SVM) ($R^2 = 0.9841$), and Adaptive Network-based Fuzzy Inference System (ANFIS) ($R^2 = 0.9906$).

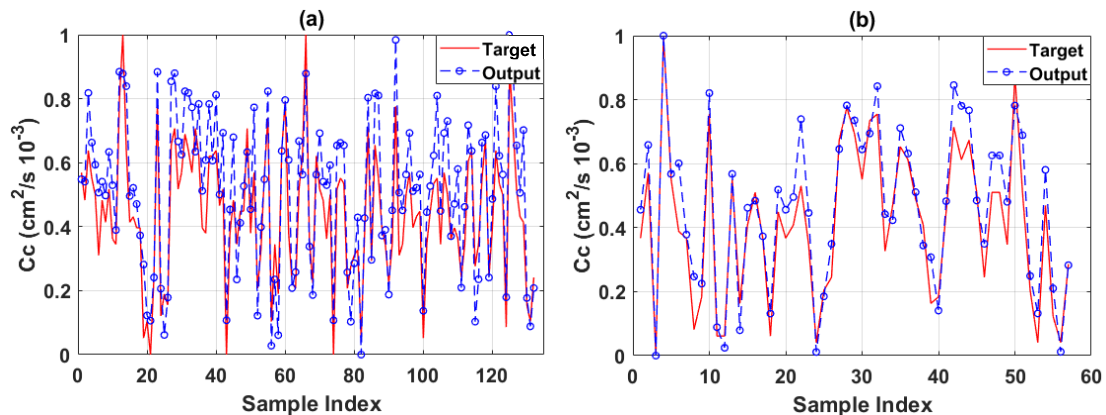


Figure 4. Probability distribution over 1000 NMC using ELM for the prediction of C_c for (a) RMSE; (b) MAE; (c) R

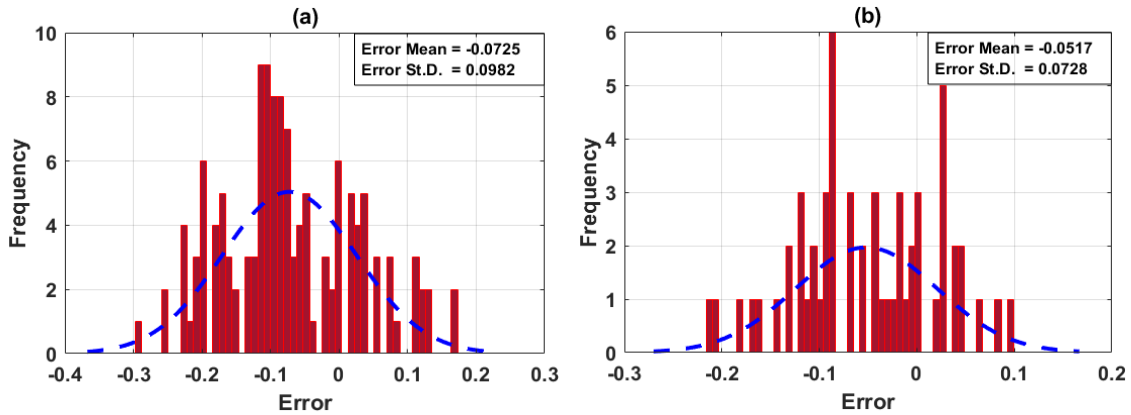


Figure 5. Histograms and error distribution of ELM algorithm in predicting Cc of soft soil for (a) training dataset; (b) testing dataset

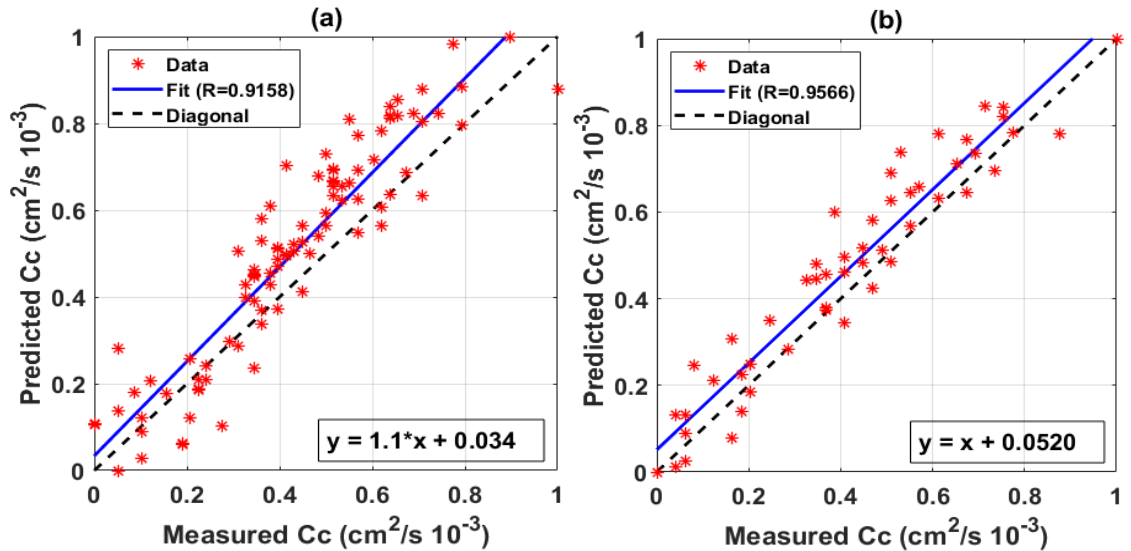


Figure 6. Linear fit lines, fit equations, and correlation graphs of ELM in predicting Cc of soft soil for (a) training dataset; (b) testing dataset

5. Conclusions

In this study, a popular machine learning model, namely ELM was developed and applied to predict the compression coefficient (Cc)-a critical soil mechanical parameter. A total of 13 input parameters and one output parameter extracted from 189 experimental soil results were used to generate the training and testing parts for building and validating the model. Monte Carlo approach and other statistical indexes such as RMSE, MAE, and

R were applied to validate the performance of the models.

The analysis results showed that ELM showed a strong capacity to predict the Cc of soil, with the value of R up to 0.9158 (training dataset) and 0.9566 (testing dataset). Monte Carlo results also proved that the ELM model has a good convergence under the effect of the random sampling strategy. However, it is also noticed that the ELM was applied in a small dataset (189 samples) collected in the local

site of Vietnam; thus, a bigger dataset with a diversity of soil properties should be used to validate the further application of ELM. The results of this study might help in quick and accurate prediction of the C_c and other vital parameters for reducing cost and time for practical construction.

Acknowledgments

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.08-2019.03.

References

- Armaghani D.J., Mohamad E.T., Hajihassani M., Yagiz S., Motaghedi H., 2016. Application of several non-linear prediction tools for estimating uniaxial compressive strength of granitic rocks and comparison of their performances. *Engineering with Computers*, 32, 189–206.
- Bui D.T., Hoang N.-D., Nhu V.-H., 2019. 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*, 35, 955–965.
- Chen J., de Hoogh K., Gulliver J., Hoffmann B., Hertel O., Ketzler M., Bauwelinck M., van Donkelaar A., Hvidtfeldt U.A., Katsouyanni K., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environment international*, 130, 104934.
- Collins G.S., Moons K.G., 2019. Reporting of artificial intelligence prediction models. *The Lancet*, 393, 1577–1579.
- 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., Jaafari A., Bayat M., Mafi-Gholami D., Qi C., Moayedhi H., Phong T.V., Ly H.-B., Le T.-T., Trinh P.T., Luu C., Quoc N.K., Thanh B.N., Pham B.T., 2020b. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *CATENA*, 188, 104451. <https://doi.org/10.1016/j.catena.2019.104451>.
- Dao D.V., Ly H.-B., Vu H.-L.T., Le T.-T., Pham B.T., 2020c. Investigation and Optimization of the C-ANN Structure in Predicting the Compressive Strength of Foamed Concrete. *Materials*, 13, 1072.
- Das B.M., Sobhan K., 2013. *Principles of geotechnical engineering*. Cengage learning.
- Huang G.-B., Zhou H., Ding X., Zhang R., 2011. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42, 513–529.
- Huang G.-B., Zhu Q.-Y., Siew C.-K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks, in: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). IEEE, 985–990.
- Huang G.-B., Zhu Q.-Y., Siew C.-K., 2006. Extreme learning machine: theory and applications. *Neurocomputing*, 70, 489–501.
- Khandelwal M., Marto A., Fatemi S.A., Ghoroghi 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.
- Khosravi K., Daggupati P., Alami M.T., Awadh S.M., Ghareb M.I., Panahi M., Pham B.T., Rezaie F., Qi C., Yaseen Z.M., 2019. Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Computers and Electronics in Agriculture*, 167, 105041. <https://doi.org/10.1016/j.compag.2019.105041>.
- 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. https://doi.org/10.1007/978-981-15-0802-8_191.

- Ly H.-B., Le L.M., Duong H.T., Nguyen T.C., Pham T.A., Le T.-T., Le V.M., Nguyen-Ngoc L., Pham B.T., 2019a. Hybrid Artificial Intelligence Approaches for Predicting Critical Buckling Load of Structural Members under Compression Considering the Influence of Initial Geometric Imperfections. *Applied Sciences*, 9, 2258. <https://doi.org/10.3390/app9112258>.
- Ly H.-B., Le L.M., Phi L.V., Phan V.-H., Tran V.Q., Pham B.T., Le T.-T., Derrible S., 2019b. 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., 2019c. 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>.
- Ly H.-B., Monteiro E., Le T.-T., Le V.M., Dal M., Regnier G., Pham B.T., 2019d. Prediction and Sensitivity Analysis of Bubble Dissolution Time in 3D Selective Laser Sintering Using Ensemble Decision Trees. *Materials*, 12, 1544.
- Ly H.-B., Pham B.T., Dao D.V., Le V.M., Le L.M., Le T.-T., 2019e. Improvement of ANFIS Model for Prediction of Compressive Strength of Manufactured Sand Concrete. *Applied Sciences*, 9, 3841. <https://doi.org/10.3390/app9183841>.
- Michie D., Spiegelhalter D.J., Taylor C., 1994. Machine learning. *Neural and Statistical Classification*, 13, 1–298.
- Moayedi H., Gör M., Lyu Z., Bui D.T., 2020a. Herding Behaviors of grasshopper and Harris hawk for hybridizing the neural network in predicting the soil compression coefficient. *Measurement*, 152, 107389.
- Moayedi H., Tien Bui D., Dounis A., Ngo P.T.T., 2020b. A Novel Application of League Championship Optimization (LCA): Hybridizing Fuzzy Logic for Soil Compression Coefficient Analysis. *Applied Sciences*, 10, 67.
- Mohamad E.T., Armaghani D.J., Hasanipanah M., Murlidhar B.R., Alel M.N.A., 2016. Estimation of air-overpressure produced by blasting operation through a neuro-genetic technique. *Environmental Earth Sciences*, 75, 174.
- Mohamad E.T., Armaghani D.J., Momeni E., Yazdavar A.H., Ebrahimi M., 2018. Rock strength estimation: a PSO-based BP approach. *Neural Computing and Applications*, 30, 1635–1646.
- Mohamad E.T., Hajihassani M., Armaghani D.J., Marto A., 2012. Simulation of blasting-induced air overpressure by means of artificial neural networks. *Int. Rev. Model. Simul*, 5, 2501–2506.
- Momeni E., Nazir R., Armaghani D.J., Maizir H., 2015. Application of Artificial Neural Network for Predicting Shaft and Tip Resistances of Concrete Piles. *Earth Sciences Research Journal*, 19, 85–93. <https://doi.org/10.15446/esrj.v19n1.38712>.
- Nguyen H.Q., Ly H.-B., Tran V.Q., Nguyen T.-A., Le T.-T., Pham B.T., 2020. Optimization of Artificial Intelligence System by Evolutionary Algorithm for Prediction of Axial Capacity of Rectangular Concrete Filled Steel Tubes under Compression. *Materials*, 13, 1205.
- Nguyen M.D., Pham B.T., Tuyen T.T., Hai Yen H.P., Prakash I., Vu T.T., Chapi K., Shirzadi A., Shahabi H., Dou J., Quoc N.K., Bui D.T., 2019. Development of an Artificial Intelligence Approach for Prediction of Consolidation Coefficient of Soft Soil: A Sensitivity Analysis. *The Open Construction and Building Technology Journal*, 13. <https://doi.org/10.2174/1874836801913010178>.
- Nizar A., Dong Z., Wang Y., 2008. Power utility nontechnical loss analysis with extreme learning machine method. *IEEE Transactions on Power Systems*, 23, 946–955.
- 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., Bui K.-T.T., Prakash I., Chapi K., Bui D.T., 2019a. A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. *CATENA*, 173, 302–311. <https://doi.org/10.1016/j.catena.2018.10.004>.

- Pham B.T., Nguyen M.D., Dao D.V., Prakash I., Ly H.-B., Le T.-T., Ho L.S., Nguyen K.T., Ngo T.Q., Hoang V., Son L.H., Ngo H.T.T., Tran H.T., Do N.M., Van Le H., Ho H.L., Tien Bui D., 2019b. 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*, 679, 172–184. <https://doi.org/10.1016/j.scitotenv.2019.05.061>.
- 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. https://doi.org/10.1007/978-981-15-0802-8_187.
- 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.
- Qi C., Ly H.-B., Chen Q., Le T.-T., Le V.M., Pham B.T., 2020. Flocculation-dewatering prediction of fine mineral tailings using a hybrid machine learning approach. *Chemosphere*, 244, 125450. <https://doi.org/10.1016/j.chemosphere.2019.125450>.
- Rivera J.I., Bonilla C.A., 2020. Predicting soil aggregate stability using readily available soil properties and machine learning techniques. *Catena*, 187, 104408.
- Thanh T.T.M., Ly H.-B., Pham B.T., 2020. A Possibility of AI Application on Mode-choice Prediction of Transport Users in Hanoi, 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*. Springer Singapore, Singapore, 1179–1184. https://doi.org/10.1007/978-981-15-0802-8_189.