



## Artificial Intelligence-Based Prediction of Geotechnical Impacts of Polyethylene Bottles and Polypropylene on Clayey Soil

---

Abolfazl Baghbani, Firas Daghistani, Hasan Baghbani,  
Katayoon Kiany and Jafar Bolouri Bazaz

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

February 19, 2023

# Artificial Intelligence-Based Prediction of Geotechnical Impacts of Polyethylene Bottles and Polypropylene on Clayey Soil

Abolfazl Baghbani<sup>1,a\*</sup>, Firas Daghistani<sup>2,b</sup>, Hasan Baghbani<sup>3,c</sup>

Katayoon Kiany<sup>4,d</sup> and Jafar Bolouri Bazaz<sup>5,e</sup>

<sup>1</sup>School of engineering, Deakin University, Melbourne, VIC, Australia

<sup>2</sup>Department of Civil Engineering, La Trobe University, Melbourne, VIC, Australia; Also with Civil Engineering Department, University of Business and Technology, Jeddah, Saudi Arabia

<sup>3</sup>School of Engineering, Ferdowsi University, Mashhad, Iran

<sup>4</sup>School of Design, University of Melbourne, Melbourne, VIC, Australia

<sup>5</sup>School of Engineering, Ferdowsi University, Mashhad, Iran

<sup>a</sup>abaghbani@deakin.edu.au, <sup>b</sup>f.daghistani@latrobe.edu.au, <sup>c</sup>hasanbaghbani1998@gmail.com,

<sup>d</sup>kiany@student.unimelb.edu.au, <sup>e</sup>bolouri@um.ac.ir

\* corresponding author

**Keywords:** Plastic Waste; Clayey Soil; Artificial intelligence; Soil stabilizer; CRRF; ANN.

**Abstract.** This study aims to investigate the application of artificial intelligence (AI) methods in predicting the resilient modulus of soil mixtures with polyethylene (PE) bottles and polypropylene (PP). The AI methods used in the study are artificial neural network (ANN) and classification and regression random forest (CRRF), and the modeling was conducted using a database of 160 datasets. The study also evaluated the importance of different input parameters on the accuracy of the models. The results show that the CRRF model is more accurate than the ANN model in predicting the effects of materials PE and PP on soil resilient modulus. Additionally, the study found that the number of hidden layers and neurons in the ANN model should be optimized for the best performance and increasing their number does not always lead to increased accuracy. Finally, the study identified the most and least important input parameters for predicting the effect of PE and PP on the resilient modulus of the mixture using both AI models.

## Introduction

Globally, plastic waste has become a significant environmental concern [1]. The disposal of plastic waste has become a significant challenge, resulting in negative environmental impacts, including soil pollution [1-3]. Recently, there has been a growing interest in recycling plastic waste and using it as a partial replacement for traditional construction materials. Plastic waste is one of these materials that can be used in geotechnical applications.

A clayey soil is characterized by high plasticity, high compressibility, and low shear strength. As a result of these characteristics, the soil is prone to instability and erosion. A clayey soil's mechanical behavior is affected by several factors, including its mineral composition, water content, and confining pressure [4-6]. In addition to being non-biodegradable, plastic waste materials can persist in the environment for decades or even centuries. As a result of the accumulation of plastic waste in the soil, soil pollution can occur, and the soil's geotechnical properties can be adversely affected. In order to reduce plastic waste and improve soil mechanical



properties, the use of plastic waste materials in geotechnical applications has been investigated [7-9].

Recently, studies have investigated the effects of plastic waste materials on clayey soil's geotechnical properties. Niyomukiza et al. [10] investigated the effects of polyethylene terephthalate (PET) waste on clayey soil's shear strength and compressibility. According to the results, the addition of PET waste reduced the shear strength of the soil and increased its compressibility. According to Bandyopadhyay and Sharma and Sharma [11], plastic waste affects clayey soil's deformation characteristics and shear strength. As a result of the addition of plastic waste, the soil's shear strength was reduced and its deformation was increased. However, it was observed that plastic waste improved the bearing capacity of the soil and reduced its settlement.

There are several parameters that influence the strength of soil and waste bottles mixtures. Due to the multiplicity and nonlinearity of these parameters, a comprehensive equation for predicting the strength of soil and waste bottles mixtures has not yet been developed. Artificial intelligence has been used as a method of solving this problem. A number of fields, including soil mechanics [12-13], soil dynamics [14-15], soil cracking [16-17], road construction [18-19] have successfully used artificial intelligence methods [20]. In spite of this, no study has yet been conducted to investigate the use of artificial intelligence methods in predicting the resilient modulus of two soil- of polyethylene (PE) bottles and polypropylene (PP) mixtures. To predict the parameters of resilient modulus, two artificial intelligence methods, artificial neural network (ANN) and classification and regression random forest (CRRF), were used in this study for the first time. Modeling was conducted using a database consisting of 160 datasets. The inputs included confining pressure, cyclic stress, constant stress, load cycle number, length of waste materials, fiber content, UCS and CBR of mixtures. The importance of input parameters was evaluated after evaluating the AI models.

## Database Collection and Processing

### Experiment and data collection

In this study, a database consisting of 160 sets with eight inputs and one output was utilized. Collection data for two types of recycled waste bottles, PE and PP, was collected from Hassan et al. [21]. The statistical parameters of this database are shown in Tables 1 and 2 for PE and PP, respectively. According to Tables 1 and 2, the database has a proper distribution.

Table 1. Statistical information of database for PE

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Resilient modulus (Mpa)	160	128.000	188.000	144.833	10.507
Confining pressure (kPa)	160	13.790	41.370	28.442	11.437
Cyclic stress (kPa)	160	12.410	62.050	36.454	17.311
Constant stress (kPa)	160	1.280	6.890	4.046	1.931
Load cycle no.	160	100.000	1000.000	156.250	218.539
length (CM)	160	1.000	2.000	1.500	0.502
Fiber content (%)	160	0.000	4.000	2.000	1.419
UCS (kPa)	160	148.000	291.000	239.500	48.139
CBR (%)	160	4.000	7.200	5.540	1.015

Table 2. Statistical information of database for PP

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Resilient modulus (Mpa)	160	130.000	150.100	137.073	4.091
Confining pressure (kPa)	160	13.790	41.370	28.442	11.437
Cyclic stress (kPa)	160	12.410	62.050	36.454	17.311
Constant stress (kPa)	160	1.280	6.890	4.046	1.931

Load cycle no.	160	100.000	1000.000	156.250	218.539
Length (CM)	160	1.000	2.000	1.500	0.502
Fiber content (%)	160	0.000	4.000	2.000	1.419
UCS (kPa)	160	148.000	256.000	217.400	36.457
CBR (%)	160	4.000	6.000	4.960	0.682

### Preparation of the data for AI modelling

In order to prepare the database for AI modeling, the values of different parameters were linearly normalized using the following equation. Normalizing the database will increase the accuracy of the model since each parameter in the database has a specific unit.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where  $X_{max}$ ,  $X_{min}$ ,  $X$  and  $X_{norm}$  are maximum, minimum, actual, and normalized values, respectively.

As part of the preparation process, the database was divided into two training and testing databases. A total of 20% (26 sets) of the database was used for testing and 80% (134 sets) for training purposes. Tables 3 and 4 illustrate the statistical information for these two databases. To increase the accuracy of the model, it is better to have the statistical information of these two databases close to each other, as is the case in this study.

Table 3. Statistical information of training database

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Resilient modulus (MPa)-PE	134	128.000	188.000	144.619	10.463
Resilient modulus (MPa)-PP	134	130.000	150.100	137.213	4.365
Confining pressure (kPa)	134	13.790	41.370	28.712	11.722
Cyclic stress (kPa)	134	12.410	62.050	37.323	17.048
Constant stress (kPa)	134	1.280	6.890	4.142	1.903
Load cycle no.	134	100.000	1000.000	160.448	226.121
length (CM)	134	1.000	2.000	1.485	0.502
Fiber (%)	134	0.000	4.000	1.993	1.438
UCS-Kpa	134	148.000	291.000	238.507	48.964
CBR (%)	134	4.000	7.200	5.525	1.029

Table 4. Statistical information of testing database

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
Resilient modulus (MPa)-PE	26	132.100	170.000	145.941	10.877
Resilient modulus (MPa)-PP	26	132.100	142.000	136.356	2.107
Confining pressure (kPa)	26	13.790	41.370	27.050	9.929
Cyclic stress (kPa)	26	12.410	62.050	31.980	18.297
Constant stress (kPa)	26	1.280	6.890	3.551	2.036
Load cycle no.	26	100.000	1000.000	134.615	176.505
length (CM)	26	1.000	2.000	1.577	0.504
Fiber (%)	26	0.000	4.000	2.038	1.341
UCS-Kpa	26	148.000	291.000	244.615	44.172
CBR (%)	26	4.000	7.200	5.615	0.959

### Data-driven modeling

#### Artificial neural network (ANN)

Artificial Neural Networks, or ANNs, are a type of machine learning model that has been inspired by the structure and function of the human brain. They have proven to be highly effective at processing complex data sets and making predictions based on them. ANNs consist of multiple interconnected nodes, which process information and adjust the connections between them to learn

from the data. They have been used in a wide variety of applications, ranging from image and speech recognition to natural language processing and predictive modeling. The power of ANNs lies in their ability to learn and generalize from large amounts of data, making them a key technology in the field of artificial intelligence.

### Classification and regression random forest (CRRF)

Random Forest is a machine learning algorithm that can be used for both classification and regression tasks. Random Forest works by constructing multiple decision trees during training and then aggregating the results of each tree to make a final prediction.

For classification tasks, Random Forest constructs a set of decision trees, where each tree predicts the class of a given input based on a set of features. During training, each decision tree is built using a random subset of the training data and a random subset of the features. When making a prediction, the input is passed through each tree, and the predicted class is determined by taking a majority vote of the predictions made by each tree.

For regression tasks, Random Forest works similarly, but instead of predicting a class, it predicts a continuous numerical value. Each decision tree is constructed to predict the output value based on a set of features, and the final prediction is made by averaging the predictions made by each tree.

Random Forest is a popular machine learning algorithm due to its high accuracy and ability to handle large datasets with a large number of features. It also has the advantage of being less prone to overfitting compared to single decision trees, as the aggregation of multiple trees helps to reduce the effects of individual trees that may be overfitting to the data.

## Results

### Artificial neural network (ANN)

The artificial neural network model was used and investigated with a variety of architectures. Several factors influence the architecture of a ANN, including the number of hidden layers and the number of neurons within each layer and type of algorithm. This study examined the number of hidden layers ranging from one layer to five layers. Additionally, two algorithms Bayesian Regularization (BR) and Levenberg-Marquardt (LM) were examined in order to determine their performance. At the end of the process, the number of neurons in each layer was changed in order to achieve the best and most optimal architecture. In Tables 5 and 6, the results of ANN modeling were presented for PE and PP, respectively. For testing database, to predict the effects of adding PE to soil, the average accuracy ( $R^2$ ) of the ANN models for algorithm BR is equal to 0.977, and for algorithm LM is equal to 0.943. Therefore, algorithm BR has performed better than algorithm LM. Based on Table 6, algorithm BR performed better than algorithm LM for predicting the effects of PP.

Table 5. Results of ANN for PE

	The number of hidden layers	R <sup>2</sup> -Test	R <sup>2</sup> -Train	MAE-Test	MAE-Train
Bayesian Regularization	1H	0.954	0.972	1.964	2.431
	2H	0.981	0.974	1.031	1.453
	3H	0.993	0.988	0.743	0.959
	4H	0.983	0.986	0.983	1.432
	5H	0.973	0.987	1.731	1.923
	Average	0.977	0.981	1.290	1.640
Levenberg-Marquardt	1H	0.932	0.943	2.893	2.564
	2H	0.945	0.949	2.230	2.021
	3H	0.949	0.953	1.455	1.421

	<b>4H</b>	0.953	0.955	1.321	1.342
	<b>5H</b>	0.934	0.940	2.102	2.054
	<b>Average</b>	0.943	0.948	2.000	1.880

Table 6. Results of ANN for PP

	The number of hidden layers	R <sup>2</sup> -Test	R <sup>2</sup> -Train	MAE-Test	MAE-Train
<b>Bayesian Regularization</b>	<b>1H</b>	0.864	0.923	1.495	1.453
	<b>2H</b>	0.873	0.938	1.239	1.254
	<b>3H</b>	0.880	0.951	0.742	0.831
	<b>4H</b>	0.881	0.956	0.629	0.735
	<b>5H</b>	0.872	0.938	0.902	0.972
	<b>Average</b>	0.874	0.943	1.001	1.049
<b>Levenberg-Marquardt</b>	<b>1H</b>	0.842	0.905	2.593	2.019
	<b>2H</b>	0.852	0.920	2.442	1.875
	<b>3H</b>	0.874	0.934	2.002	1.549
	<b>4H</b>	0.878	0.949	1.730	1.349
	<b>5H</b>	0.856	0.921	2.021	1.892
	<b>Average</b>	0.860	0.926	2.158	1.737

It has also been determined that the number of 25 neurons for each hidden layer is the optimal number after many trials and errors. It was found that the optimum number of hidden layers for material PE was three, and the optimum number of hidden layers for material PP was four. It appears that increasing the number of hidden layers and neurons will not always increase the accuracy of the ANN model, and there is a point at which there is an optimal number of hidden layers and neurons.

Figs 1 and 2 show diagrams of ANN predicted values versus actual values for PE and PP, respectively. According to the results, the ANN model predicts well the performance of both materials.

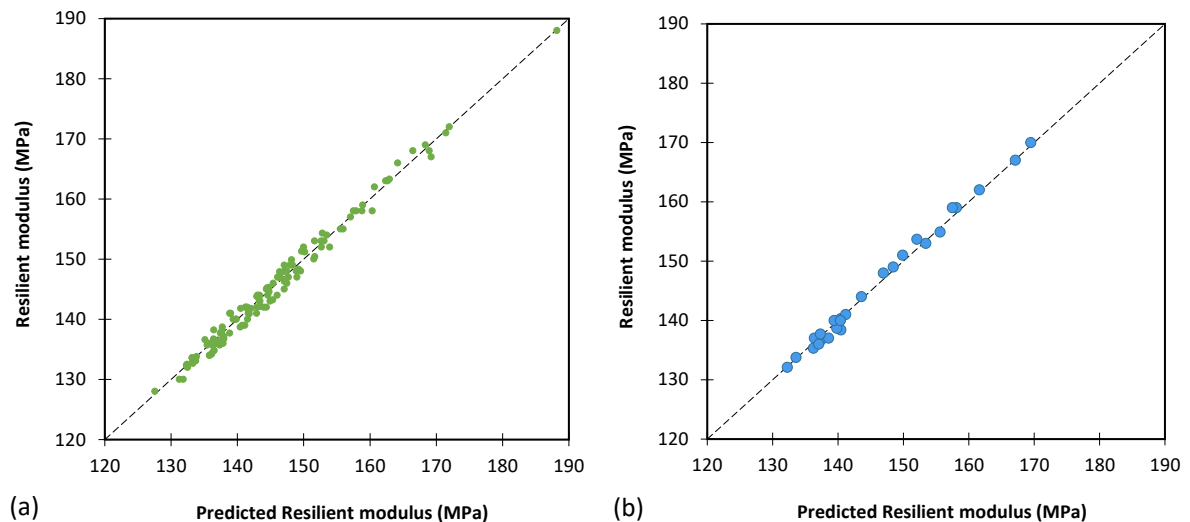


Fig. 1. The results of ANN for predicting resilient modulus for PE

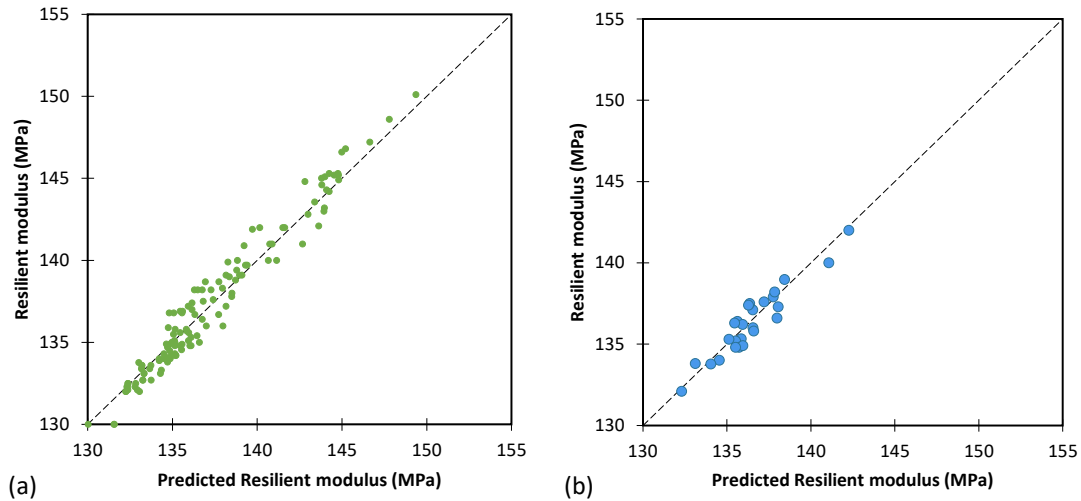


Fig. 2. The results of ANN for predicting resilient modulus for PP

**Classification and regression random forest (CRRF)**

Among the important parameters of the CRRF method are parameters max depth of trees and the number of trees. The effective parameters of the CRRF model were repeatedly changed and a number of trials and errors were conducted in order to obtain the best CRRF model. Table 7 shows the values of the effective parameters in the CRRF model. On the basis of the obtained results, the CRRF method is most effective when the number of trees is equal to 300 and the max depth of trees is 8.

Table 7. The specifications of the best CRRF.

Trees parameters					Forest parameters		
Min. node size	Min. son size	Max depth	Mtry	CP	Sampling	Sample size	Number of trees
2	1	8	2	0.00001	Random with replacement	78	300

Figs 3 and 4 illustrate the values predicted by the CRRF model in comparison with the actual resistance values for materials PE and PP, respectively. The CRRF model is able to predict the effect of both materials on soil resilient modulus based on the results obtained.

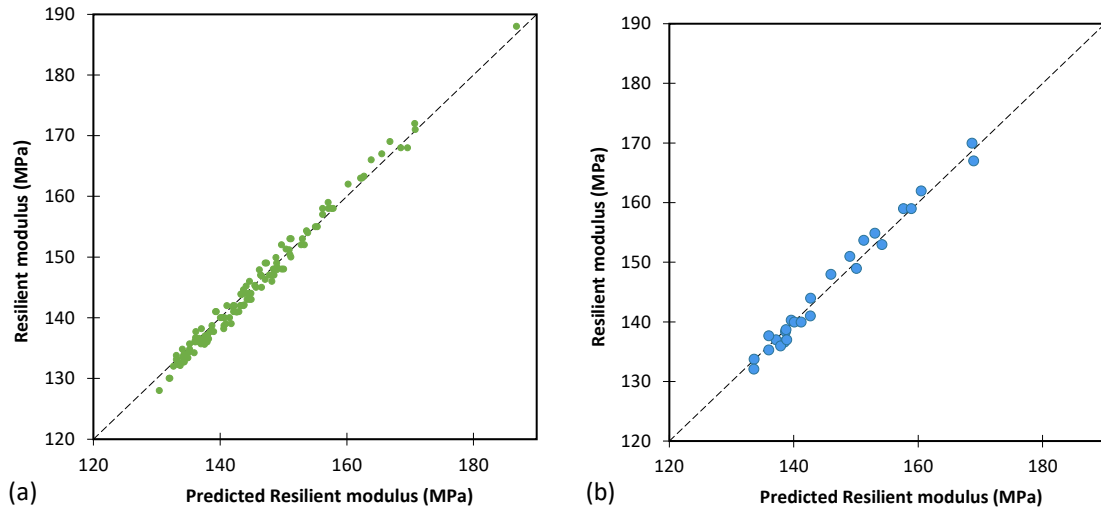


Fig. 3. The results of CRRF for predicting resilient modulus for PE

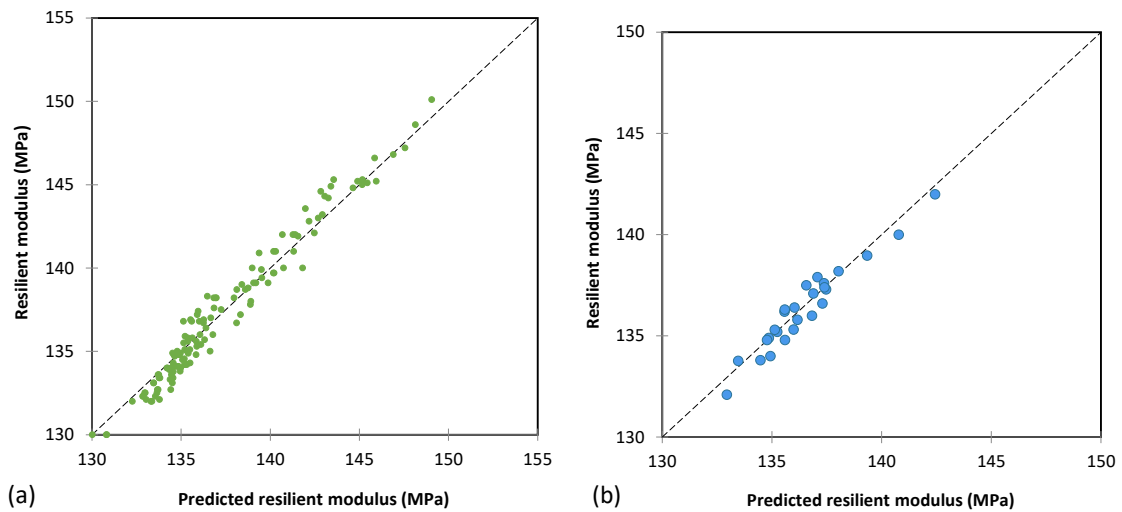


Fig. 4. The results of CRRF for predicting resilient modulus for PP

Table 8 illustrates the performance of the CRRF model for predicting soil resilient modulus after adding materials PE and PP. For predicting the training database for PE, the accuracy ( $R^2$ ) and error (MAE) of the CRRF model are equal to 0.986 and 1.047, respectively, and for predicting the effect of PP, they are equal to 0.961 and 0.705. According to the testing database, the accuracy ( $R^2$ ) of the CRRF model for predicting the effect of PE and PP on soil resilient modulus is 0.961 and 0.926, and its error (MAE) is 0.705 and 0.470, respectively. Based on these results, the CRRF model is capable of predicting the effects of materials PE and PP on soil resilient modulus, and CRRF performed better than ANN method.

Table 8. The performance of CRRF model

Performance metrics	PE		PP	
	Training	Testing	Training	Testing
MAE	1.047	1.237	0.705	0.470
$R^2$	0.986	0.982	0.961	0.926

### The variable importance of input parameters



Figs 5 and 6 show the importance of input parameters on soil resilient modulus prediction for both artificial intelligence models and for PE and PP, respectively. In the ANN model, parameter confining pressure was determined to be the most important parameter to predict the effect of PE on soil resilient modulus, while parameter load cycle number was determined to be the least important parameter. Although, according to Fig. 5b, in the CRRF model, parameters CBR and load cycle number have the highest and lowest importance, respectively in predicting rffect of PE on the resilient modulus of the mixture. According to Fig. 6, in the ANN and CRRF model, the most and least important parameters to predict effect of PP on the resilient modulus of mixture are parameters confining pressure and length, respectively.

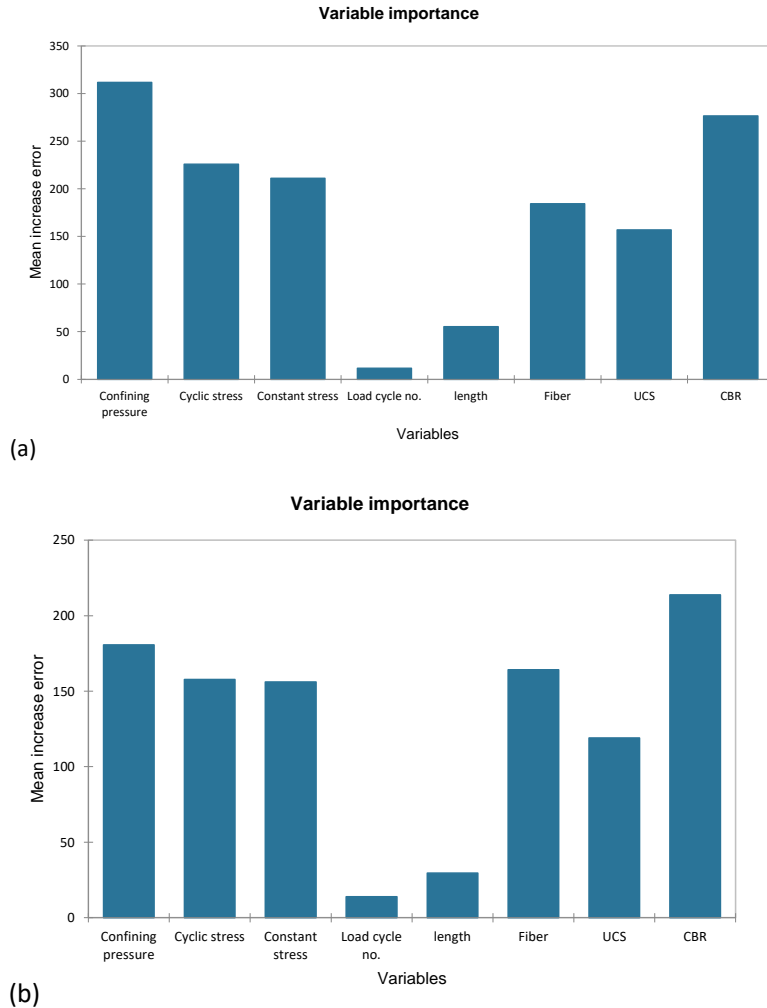
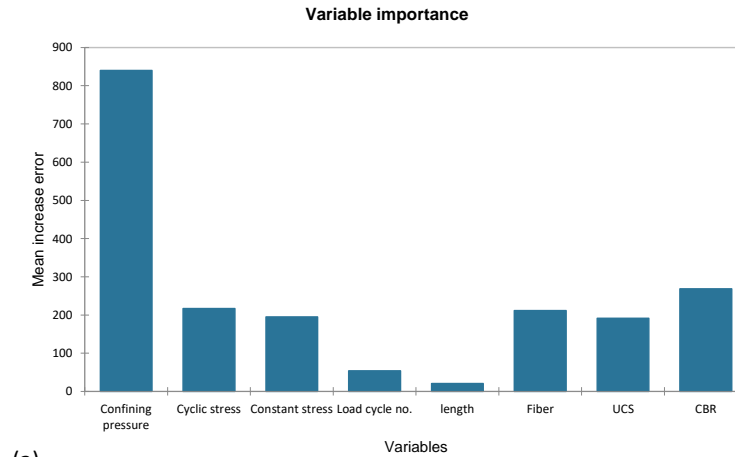
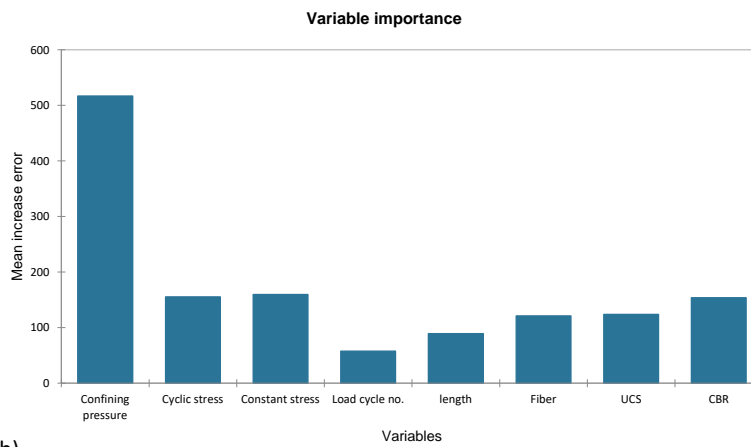


Fig. 5. The importance of parameters to predict resilient modulus for PE, based on (a) ANN and (b) CRRF



(a)



(b)

Fig. 6. The importance of parameters to predict resilient modulus for PP, based on (a) ANN and (b) CRRF

### Conclusion

In this study, artificial intelligence methods were employed to predict the resilient modulus of two soil mixtures containing polyethylene (PE) and polypropylene (PP). For the first time, artificial neural networks (ANNs) and classification and regression random forests (CRRFs) were used to predict resilient modulus parameters. Results indicated that the CRRF model outperformed the ANN model in predicting the effects of PE and PP on soil resilient modulus. In addition, the study found that the number of hidden layers and neurons for the ANN model is optimal and that increasing them beyond this point will not necessarily increase accuracy. It was also found that the ANN model was able to accurately predict the performance of both materials, with the optimal number of hidden layers and neurons being three and 25 for PE, and four and 25 for PP. In summary, the study demonstrated the potential of using artificial intelligence methods to predict soil properties and to assist with the design and construction of soil structures..

### References

- [1] Huang S, Wang H, Ahmad W, Ahmad A, Ivanovich Vatin N, Mohamed AM, Deifalla AF, Mehmood I. Plastic waste management strategies and their environmental aspects: A scientometric analysis and comprehensive review. *International Journal of Environmental Research and Public Health*. 2022 Apr 10;19(8):4556.

- [2] Sahebzadeh S, Heidari A, Kamelnia H, Baghbani A. Sustainability features of Iran's vernacular architecture: A comparative study between the architecture of hot-arid and hot-arid-windy regions. *Sustainability*. 2017 May 4;9(5):749.
- [3] Adeniran AA, Shakantu W. The health and environmental impact of plastic waste disposal in South African Townships: A review. *International Journal of Environmental Research and Public Health*. 2022 Jan 11;19(2):779.
- [4] Baghbani A, Costa S, Lu Y, Soltani A, Abuel-Naga H and Samui P. Effects of Particle Shape on Shear Modulus of Sand Using Dynamic Simple Shear Testing. *Arabian journal of geosciences*. 2023.
- [5] Leng Y, Peng J, Wang S, Lu F. Development of water sensitivity index of loess from its mechanical properties. *Engineering Geology*. 2021 Jan 1;280:105918.
- [6] Baghbani A, Costa S, O'Kelly BC, Soltani A, Barzegar M. Experimental study on cyclic simple shear behaviour of predominantly dilative silica sand. *International Journal of Geotechnical Engineering*. 2022 Oct 23:1-5.
- [7] Abukhattala M, Fall M. Geotechnical characterization of plastic waste materials in pavement subgrade applications. *Transportation Geotechnics*. 2021 Mar 1;27:100472.
- [8] Hassan HJ, Rasul J, Samin M. Effects of plastic waste materials on geotechnical properties of clayey soil. *Transportation Infrastructure Geotechnology*. 2021 Sep;8(3):390-413.
- [9] Baghbani A, Daghistani F, Naga HA, Costa S. Development of a Support Vector Machine (SVM) and a Classification and Regression Tree (CART) to Predict the Shear Strength of Sand Rubber Mixtures. In *Proceedings of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)*, Newcastle, Australia 2022.
- [10] Niyomukiza JB, Bitekateko A, Nsemerirwe J, Kawiso B, Kiwanuka M. Investigating the effect of PET plastic bottle strips on the strength and compressibility properties of clayey soil. In *IOP Conference Series: Earth and Environmental Science 2021 Nov 1 (Vol. 894, No. 1, p. 012021)*. IOP Publishing.
- [11] Sharma A, Sharma RK. Effect of addition of construction-demolition waste on strength characteristics of high plastic clays. *Innovative Infrastructure Solutions*. 2019 Dec;4(1):27.
- [12] Inazumi S, Intui S, Jotisankasa A, Chaiprakaikeow S, Kojima K. Artificial intelligence system for supporting soil classification. *Results in Engineering*. 2020 Dec 1;8:100188.
- [13] Nguyen MD, Baghbani A, Alnedawi A, Ullah S, Kafle B, Thomas M, Moon EM, Milne NA. Experimental Study on the Suitability of Aluminium-Based Water Treatment Sludge as a Next Generation Sustainable Soil Replacement for Road Construction. Available at SSRN 4331275.
- [14] Bao X, Jin Z, Cui H, Chen X, Xie X. Soil liquefaction mitigation in geotechnical engineering: An overview of recently developed methods. *Soil Dynamics and Earthquake Engineering*. 2019 May 1;120:273-91.
- [15] Baghbani A, Choudhury T, Samui P, Costa S. Prediction of secant shear modulus and damping ratio for an extremely dilative silica sand based on machine learning techniques. *Soil Dynamics and Earthquake Engineering*. 2023 Feb 1;165:107708.
- [16] Xu JJ, Zhang H, Tang CS, Cheng Q, Tian BG, Liu B, Shi B. Automatic soil crack recognition under uneven illumination condition with the application of artificial intelligence. *Engineering Geology*. 2022 Jan 1;296:106495.

- [17] Baghbani A, Costa S, Choudhury T, Faradonbeh RS. Prediction of Parallel Desiccation Cracks of Clays Using a Classification and Regression Tree (CART) Technique. In Proceedings of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR), Newcastle, Australia 2022.
- [18] Yaseen ZM, Ali ZH, Salih SQ, Al-Ansari N. Prediction of risk delay in construction projects using a hybrid artificial intelligence model. Sustainability. 2020 Feb 18;12(4):1514.
- [19] Baghbani A, Baghbani H, Shalchiyan MM, Kiany K. Utilizing artificial intelligence and finite element method to simulate the effects of new tunnels on existing tunnel deformation. Journal of Computational and Cognitive Engineering. 2022 Aug 15.
- [20] Baghbani A, Choudhury T, Costa S, Reiner J. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. Earth-Science Reviews. 2022 May 1;228:103991.
- [21] Hassan HJ, Rasul J, Samin M. Effects of plastic waste materials on geotechnical properties of clayey soil. Transportation Infrastructure Geotechnology. 2021 Sep;8(3):390-413.