



# Predicting the Strength of Recycled Glass Powder-Based Geopolymers for Improving Mechanical Behavior of Clay Soils Using Artificial Intelligence

---

Abolfazl Baghbani, Firas Daghistani, Hasan Baghbani and  
Katayoon Kiany

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

February 19, 2023

# Predicting the Strength of Recycled Glass Powder-based Geopolymers for Improving Mechanical Behavior of Clay Soils using Artificial Intelligence

Abolfazl Baghbani<sup>1,a\*</sup>, Firas Daghistani<sup>2,b</sup>, Hasan Baghbani<sup>3,c</sup>  
and Katayoon Kiany<sup>4,d</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>a</sup>abaghbani@deakin.edu.au, <sup>b</sup>f.daghistani@latrobe.edu.au, <sup>c</sup>hasanbaghbani1998@gmail.com,  
<sup>d</sup>kianyk@student.unimelb.edu.au,

\* corresponding author

**Keywords:** Recycled glass powder; Ordinary Portland cement; Clayey Soil; Artificial intelligence; Soil stabilizer; CRRF; SVM.

**Abstract.** The paper investigates the use of artificial intelligence (AI) methods to predict the strength of recycled glass powder (RGP) and soil mixtures based on different input parameters. The study utilized a database of 57 sets with 5 inputs, including RGP percentage, ordinary Portland cement (OPC) percentage, molar concentration, curing temperature and time, and one output, mixed UCS. There were two artificial intelligence models used in this study, a support vector machines (SVM) and classification and regression random forest (CRRF). The results demonstrate the potential of RGP-based geopolymers to improve the mechanical behavior of clay soils, and the use of AI methods to predict the strength of RGP and soil mixtures with high accuracy. Using SVM model, the testing dataset had a mean absolute error (MAE) and  $R^2$  of 0.072 and 0.978, respectively. Also, CRRF had an accurate performance with a MAE of 0.075 and the  $R^2$  of 0.979. These results suggest that the AI models fits well with the data. Also, by analyzing the results of the SVM and CRRF models, it is found that curing time is the most important input parameter, while RGP and OPC are the least significant.

## Introduction

Increasing environmental concerns have led to an interest in reusing waste materials as a partial replacement for traditional construction materials [1-3]. Recycled glass powder (RGP) has been investigated for its potential use in geopolymer applications [4-5]. As an inorganic polymer, geopolymers can be produced from a variety of waste materials, including RGP [6]. A geopolymer based on RGP may improve the mechanical properties of clay soils, as it has been demonstrated that it enhances the strength and durability of other construction materials [4].

A geopolymer is an inorganic material that is formed by the reaction of an aluminosilicate precursor with an alkali activator [7]. A three-dimensional network of Si-O-Al bonds results in a material with unique mechanical and chemical properties [8]. In addition to concrete, ceramics, and composites, geopolymers have been used in a variety of applications. Geopolymers have been



shown to improve the mechanical properties of construction materials, such as their strength, durability, and resistance to corrosion [9].

During the recycling of glass, a waste material called recycled glass powder is generated. In the construction industry, it has shown potential as a partial replacement for cement or as an additive to concrete. As well as reducing the amount of waste sent to landfills, the use of RGP has been shown to reduce the carbon footprint of construction materials. Clay soils have poor mechanical properties, such as low strength and high compressibility [10]. Geopolymer has been investigated as a means of improving the mechanical performance of clay soils. The addition of geopolymer to clay soils has been shown to improve their compressive strength, stiffness, and durability.

Recent studies have investigated the use of RGP-based geopolymers to improve the mechanical behavior of clay soils. Bilondi et al. [4] investigated the effects of RGP-based geopolymer on the mechanical properties of expansive soils. According to the results, the addition of RGP-based geopolymer enhanced the unconfined compressive strength and reduced the compressibility of the soil. In a study by Ashiq et al. [11], RGP-based geopolymer was examined for its effect on the strength and deformation behavior of soft clay. Based on the results, RGP-based geopolymer increased the soil's unconfined compressive strength and reduced its deformation.

It has been shown that various factors, such as the glass content, curing time and temperature, can affect the effect of adding glass. It has not yet been possible to develop a comprehensive model for determining the strength of a mixture of glass and soil. One of the reasons for this problem is the multiplicity of effective factors and the non-linearity of their effects. Artificial intelligence is one method that can be used to solve this problem. Based on AI methods, it is possible to predict the output with high accuracy without knowing the relationship between the parameters in advance [12]. In the last two decades, AI methods were used in geotechnical engineering applications include slope stability [13-14], tunnelling [15-16], road construction [17-18], and soil cracking [19-20], soil dynamics [21-22]. There has not yet been an article published on artificial intelligence methods for determining the strength of RGP and soil mixtures based on different input parameters.

In this study, for the first time, two AI methods, namely, the support vector machines (SVM) and classification and regression random forest (CRRF) method, are used to predict RGP and soil mixtures mixture strength using differing parameters. Input parameters include RGP percentage, ordinary Portland cement (OPC) percentage, molar concentration, curing temperature and time. There are 57 data sets included in the database. After AI modelling, selection of the best artificial intelligence model, sensitivity analysis, and parameter importance have been conducted.

## **Database Collection and Processing**

### **Experiment and data collection**

This study utilized a database of 57 datasets with 5 inputs, including RGP percentage, ordinary Portland cement (OPC) percentage, molar concentration, curing temperature and time, and one output, the strength of RGP and soil mixtures. Database was collected from the study conducted by Bilondi et al. [4]. Table 1 displays descriptive statistics for five variables collected from 57 observations. The variables include:

- UCS: This variable refers to the unconfined compressive strength of a material in mega-pascals (MPa). The observations range from 0.15 MPa to 2.2 MPa, with a mean of 1.003 MPa and a standard deviation of 0.613 MPa.

- Molar Concentration: This variable measures the concentration of a certain compound in solution in moles per liter. The observations range from 0 to 7 moles per liter, with a mean of 2.719 moles per liter and a standard deviation of 1.800 moles per liter.

- RGP: This variable represents the percentage of recycled glass powder used in the material. The observations range from 0% to 25%, with a mean of 8.211% and a standard deviation of 6.681%.
- OPC: This variable refers to the percentage of ordinary Portland cement in the material. The observations range from 0% to 5%, with a mean of 0.263% and a standard deviation of 1.126%.
- Curing Temperature (°C) and Curing Time (Days): These variables represent the curing conditions for the material, with the temperature measured in degrees Celsius and the time measured in days. The observations for curing temperature range from 25 °C to 70 °C, with a mean of 28.158 °C and a standard deviation of 11.597 °C. The observations for curing time range from 7 days to 91 days, with a mean of 32.053 days and a standard deviation of 33.543 days.

Table 1. Statistical information of database

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
UCS (MPa)	57	0.150	2.200	1.003	0.613
Molar Concentration	57	0.000	7.000	2.719	1.800
RGP (%)	57	0.000	25.000	8.211	6.681
OPC (%)	57	0.000	5.000	0.263	1.126
Curing Temperature (°C)	57	25.000	70.000	28.158	11.597
Curing Time (Days)	57	7.000	91.000	32.053	33.543

### Preparation of the data for AI modelling

In the database, the parameters have different units. The accuracy and performance of artificial intelligence models can be adversely affected by this issue. Therefore, linear normalization has been used to normalize the database. Eq. 1 shows the linear normalization equation.

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

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

A linear normalization technique is a commonly used technique in data pre-processing that aims to scale the values of parameters in a database to a common range (in this study from 0 to 1). Data is transformed linearly so that it falls within a specified range using this technique. The units of the parameters are thus standardized, which makes it easier for the AI models to process the data and make accurate predictions.

Additionally, dividing the database into training and testing parts is an important part of the preparation process. As part of this study, randomly 20% (12 datasets) of the total database was used for testing, while 80% (45 datasets) was used for training. Tables 2 and 3 provide statistical information about these two databases. As shown in Tables 2 and 3, the statistical information of the two databases is quite similar, which can lead to more accurate performance of artificial intelligence models. By using similar statistical information in the training and testing data sets, the model is less likely to overfit to the training data and can make better predictions on new, unseen data.

Table 2. Statistical information of training database

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
UCS (MPa)	45	0.150	2.200	1.026	0.617
Molar Concentration	45	0.000	7.000	2.733	1.851
RGP (%)	45	0.000	25.000	8.800	7.086

OPC (%)	45	0.000	5.000	0.222	1.042
Curing Temperature (°C)	45	25.000	70.000	28.000	11.352
Curing Time (Days)	45	7.000	91.000	32.667	32.867

Table 3. Statistical information of testing database

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
UCS (MPa)	12	0.250	2.200	0.918	0.615
Molar Concentration	12	0.000	6.000	2.667	1.670
RGP (%)	12	0.000	12.000	6.000	4.431
OPC (%)	12	0.000	5.000	0.417	1.443
Curing Temperature (°C)	12	25.000	70.000	28.750	12.990
Curing Time (Days)	12	7.000	91.000	29.750	37.414

## Data-driven modeling

### Support vector machine (SVM)

Support vector machine (SVM) is a powerful and popular machine learning algorithm used in classification and regression analyses. A group of scientists led by Vladimir Vapnik developed the technique in the 1990s. A SVM is particularly useful when dealing with complex, high-dimensional, or nonlinear data [23]. This technique is widely used in many fields, including image and text classification, bioinformatics, and many others. In SVM, the basic idea is to find a hyperplane that divides the data into two classes with the greatest margin. Margin refers to the distance between the hyperplane and the nearest data points in each class. It is the hyperplane that maximizes the margin that is most robust to new data points and has the best generalization capability.

In SVM, the data is transformed into a high-dimensional feature space, where it is easier to find separate hyperplanes. During the transformation, a kernel function measures the similarity between pairs of data points in the original space. There are several types of kernel functions, including linear, polynomial, and radial basis functions. In the case of nonlinearly separable data, SVM uses a technique known as the kernel trick. By using this trick, the algorithm is able to find a hyperplane in a high-dimensional feature space without having to calculate the coordinates of the data. As a result, SVM is computationally efficient and scalable.

In addition to binary classifications, SVMs can also be used to analyze regression data and perform multi-class classifications. SVM uses several binary classifiers to separate each pair of classes in the case of multi-class classification. These classifiers determine the final decision. SVMs have several advantages over other machine learning algorithms, such as decision trees and artificial neural networks. As a result, it is less prone to overfitting, requires fewer data preprocessing, and can be used with both numerical and categorical data. There are, however, some limitations associated with SVM, including the selection of the kernel function and the difficulty of interpreting the results.

### Classification and regression random forest (CRRF)

Classification and Regression Random Forest (CRRF) is a powerful machine learning algorithm that combines decision trees and random forests to perform both classification and regression tasks. To create a robust and accurate model, random forests are ensembles of decision trees trained on different subsets of data and feature sets. The CRRF algorithm can be applied to both classification and regression problems, making it a versatile algorithm with a wide range of applications. Based on different features, the CRRF algorithm divides the data into smaller and

smaller subsets using decision trees. Each decision tree divides the data into subsets based on different rules, and each subset corresponds to a particular decision. To predict the output of the classification or regression problem, the algorithm uses an ensemble of decision trees.

In the case of classification, CRRF uses a combination of decision trees to classify data points into different categories. A random subset of the data and a random subset of the features are used to train each decision tree. In this way, the model is able to reduce overfitting and improve accuracy. As a result of the ensemble of decision trees, CRRF predicts the class label of a new data point based on the results from the ensemble of decision trees. The class with the highest number of votes is considered to be the final prediction. As for regression, CRRF uses a similar approach to predict continuous values rather than class labels. A combination of decision trees is used to predict the output value for a given set of input features. Based on a random subset of the data and a random subset of the features, each decision tree is trained, and the final prediction is derived from the average of all the predictions.

In comparison with other machine learning algorithms, CRRF has several advantages, including its ability to handle both classification and regression tasks, its accuracy and robustness, and its ability to handle missing or noisy data. In addition, CRRF is less likely to overfit than other models, making it a good choice for complex datasets.

## Results

### Support vector machine (SVM)

By trial and error, different models with opposite values of effective parameters were tested in order to determine the most optimal SVM model. In Fig. 1, the predicted UCS values are compared with the actual UCS values. Using the obtained results, SVM model has successfully determined the UCS values for the RGP and soil mixture.

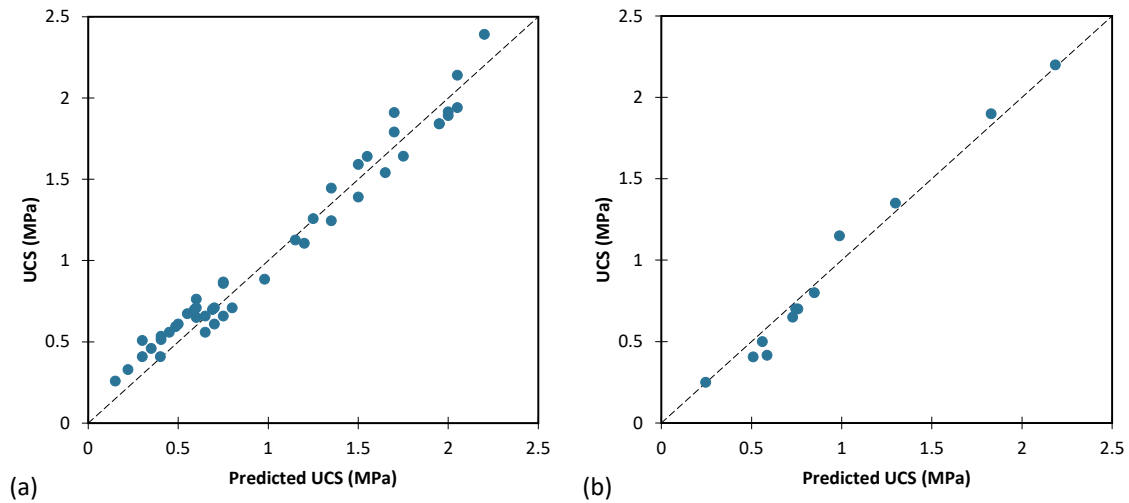


Fig. 1. Results of SVM

The performance metrics for SVM model are discussed in Table 4, including the mean absolute error (MAE) and R-squared ( $R^2$ ) values for both the training and testing datasets. The MAE for the training dataset is 0.098, whereas the MAE for the testing dataset is 0.072. Having a lower MAE value indicates that the model performs better on the testing dataset. For the training dataset, the  $R^2$  value is 0.969, while for the testing dataset, the  $R^2$  value is 0.978. This indicates that the model fits well with the data, particularly with the testing data.

Table 4. The performance of SVM model

Performance metrics	Training	Testing
---------------------	----------	---------

MAE	0.098	0.072
R <sup>2</sup>	0.969	0.978

### Classification and regression random forest (CRRF)

To find the most optimal CRRF model, various CRRF models were constructed by changing the effective parameters. This process involved testing different combinations of parameter values to find the combination that results in the best CRRF model performance. The results of these experiments are presented in Table 5, which shows the specifications of the best CRRF model. The parameters that were optimized for the best model include the minimum node size, minimum son size, maximum depth, Mtry, CP, sampling method, sample size, and number of trees. These parameters were selected based on their impact on the model performance, with the goal of maximizing accuracy while minimizing computational resources.

Table 5. 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	7	2	0.00001	Random with replacement	44	200

Fig. 2 shows the predicted values of UCS compared to their actual values for both the training and test databases. According to the figure, the CRRF model is performing well, since the predicted values are close to the actual values.

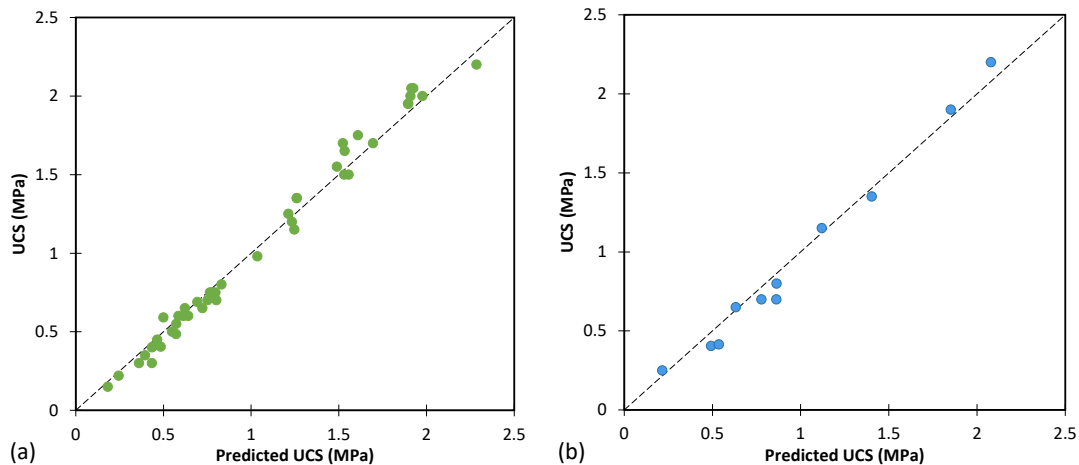


Fig. 2. The results of CRRF for predicting UCS

Table 6 provides additional information regarding the accuracy and error values of the CRRF model for both the training and test databases. According to the table, the R<sup>2</sup> value for predicting the training database is 0.986, indicating a good fit between the model and the data. MAE for the training database is 0.060, indicating that the predicted values are generally accurate.

Also, Table 6 shows the results for the test database. R<sup>2</sup> for predicting the UCS in the test database is 0.979, which indicates that the model is also well suited to the test data. Nevertheless, the MAE value for the test database is slightly higher than that of the training database, indicating that the magnitude of errors in the predicted values is greater for the test dataset.

Table 6. The performance of CRRF model

Performance metrics	Training	Testing
MAE	0.060	0.075
R <sup>2</sup>	0.986	0.979

### The variable importance of input parameters

It is important to investigate the sensitivity of artificial intelligence models to their input parameters in order to evaluate their importance. The error was calculated by varying one input parameter at a time from -100% to +100% while keeping the other parameters constant. The purpose of this analysis is to identify the input parameters that are most important for the accuracy of the AI model. Fig. 3 shows the results of this analysis for two different models. According to the results of the SVM model, curing time is the most important input parameter, while RGP is the least significant. As a result, changes in curing time have a greater impact on the accuracy of the SVM model than changes in RGB. Similarly, curing time is identified as the most significant input parameter in the CRRF model, while OPC is identified as the least significant.

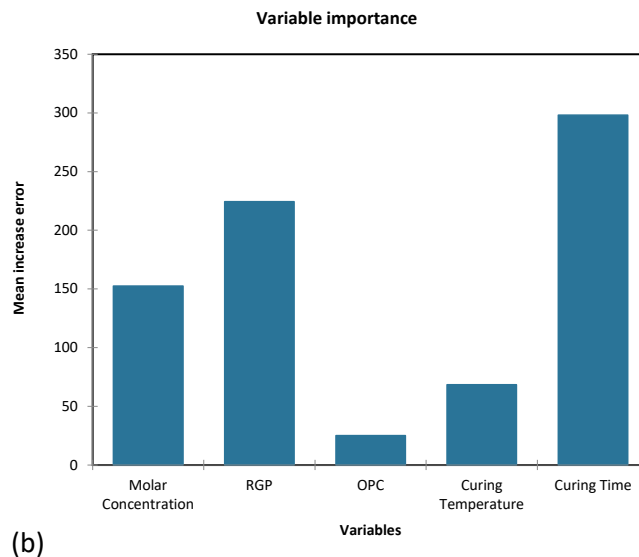
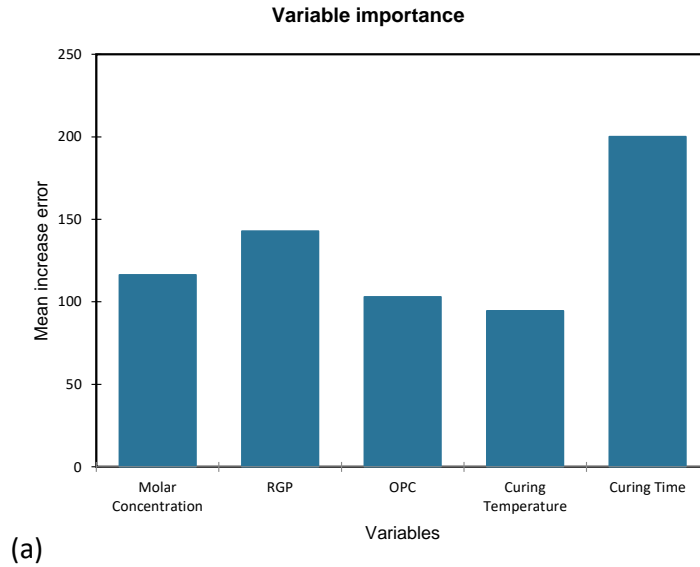




Fig. 3. The importance of parameters to predict UCS, based on (a) SVM and (b) CRRF

### Conclusion

The use of recycled glass powder in geotechnical applications has become increasingly important due to the significant benefits it provides. As a byproduct of the glass manufacturing process, glass powder can be recycled to reduce landfill waste, conserve natural resources, and reduce greenhouse gas emissions. The study aimed to predict soil and glass mixture strength using two AI methods, namely the artificial intelligence method and the vector machine method, using different input parameters. The database used in the study contained 57 items with 5 inputs, including density, and one output, mixed resistance. The input parameters used in the study were RGP percentage, ordinary Portland cement (OPC) percentage, molar concentration, curing temperature, and time.

The SVM model was successful in determining the UCS values for the soil and glass mixture. According to the results, the MAE for the training dataset was 0.098, while the MAE for the testing dataset was 0.072. In the training dataset, the R<sup>2</sup> value was 0.969, while in the testing dataset, the R<sup>2</sup> value was 0.978, indicating that the model is well suited to the data.

The results showed that the model, called Classification and Regression Random Forest (CRRF), performed better than SVM model in predicting the UCS, with a high R<sup>2</sup> value and low mean absolute error (MAE) for both the training and test databases. The study also conducted a sensitivity analysis to identify the most important input parameters for the accuracy of the CRRF model. The analysis showed that curing time was the most significant input parameter for both CRRF and Support Vector Machine (SVM) models, while RGP and OPG were the least significant.

### References

- [1] AlBiajawi, M.I., Embong, R. and Muthusamy, K., 2022. An overview of the utilization and method for improving pozzolanic performance of agricultural and industrial wastes in concrete. *Materials Today: Proceedings*, 48, pp.778-783.
- [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] 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.
- [4] Bilondi, M.P., Toufigh, M.M. and Toufigh, V., 2018. Experimental investigation of using a recycled glass powder-based geopolymer to improve the mechanical behavior of clay soils. *Construction and Building Materials*, 170, pp.302-313.
- [5] 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.
- [6] Komnitsas, K., Zaharaki, D. and Perdikatsis, V., 2009. Effect of synthesis parameters on the compressive strength of low-calcium ferronickel slag inorganic polymers. *Journal of Hazardous Materials*, 161(2-3), pp.760-768.
- [7] Abdullah, M.M.A., Hussin, K., Bnhussain, M., Ismail, K.N. and Ibrahim, W.M.W., 2011. Mechanism and chemical reaction of fly ash geopolymer cement-a review. *Int. J. Pure Appl. Sci. Technol*, 6(1), pp.35-44.

- [8] Wu, Y., Lu, B., Bai, T., Wang, H., Du, F., Zhang, Y., Cai, L., Jiang, C. and Wang, W., 2019. Geopolymer, green alkali activated cementitious material: Synthesis, applications and challenges. *Construction and Building Materials*, 224, pp.930-949.
- [9] Koushkbaghi, M., Alipour, P., Tahmouresi, B., Mohseni, E., Saradar, A. and Sarker, P.K., 2019. Influence of different monomer ratios and recycled concrete aggregate on mechanical properties and durability of geopolymer concretes. *Construction and Building Materials*, 205, pp.519-528.
- [10] Tatlisoz, N., Benson, C.H. and Edil, T.B., 1997. Effect of fines on mechanical properties of soil-tire chip mixtures. *ASTM Special Technical Publication*, 1275, pp.93-124.
- [11] Ashiq, S.Z., Akbar, A., Farooq, K. and Mujtaba, H., 2022. Sustainable improvement in engineering behavior of Siwalik Clay using industrial waste glass powder as additive. *Case Studies in Construction Materials*, 16, p.e00883.
- [12] 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.
- [13] Bardhan A, Samui P. Application of artificial intelligence techniques in slope stability analysis: a short review and future prospects. *International Journal of Geotechnical Earthquake Engineering (IJGEE)*. 2022 Jan 1;13(1):1-22.
- [14] 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.
- [15] Zhang X, Wu X, Park Y, Zhang T, Huang X, Xiao F, Usmani A. Perspectives of big experimental database and artificial intelligence in tunnel fire research. *Tunnelling and Underground Space Technology*. 2021 Feb 1;108:103691.
- [16] 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.
- [17] Sharma S, Ahmed S, Naseem M, Alnumay WS, Singh S, Cho GH. A survey on applications of artificial intelligence for pre-parametric project cost and soil shear-strength estimation in construction and geotechnical engineering. *Sensors*. 2021 Jan 11;21(2):463.
- [18] 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.
- [19] 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.
- [20] 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.

[21] Njock PG, Shen SL, Zhou A, Lyu HM. Evaluation of soil liquefaction using AI technology incorporating a coupled ENN/t-SNE model. *Soil Dynamics and Earthquake Engineering*. 2020 Mar 1;130:105988.

[22] 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.

[23] Vapnik V. *The nature of statistical learning theory*. Springer science & business media; 1999 Nov 19.