Article

Improving Soil Stability with Alum Sludge: An AI-Enabled Approach for Accurate Prediction of California Bearing Ratio

Abolfazl Baghbani [1,*] **, Minh Duc Nguyen** [1] **, Ali Alnedawi** [1,†] **, Nick Milne** [1] **, Thomas Baumgartl** [2]
and Hossam Abuel-Naga [3,*]

1   School of Engineering, Deakin University, Geelong, VIC 3216, Australia; minhd@deakin.edu.au (M.D.N.);
     alnedawi.ali.m@gmail.com (A.A.); n.milne@deakin.edu.au (N.M.)
2   Future Regions Research Centre, Federation University Australia, Churchill, Gippsland, VIC 3841, Australia;
     t.baumgartl@federation.edu.au
3   Department of Civil Engineering, La Trobe University, Melbourne, VIC 3086, Australia
*   Correspondence: abaghbani@deakin.edu.au (A.B.); h.aboel-naga@latrobe.edu.au (H.A.-N.)
†   Current address: Investment Department, Diwan Endowment, Baghdad, Iraq.

Abstract: Alum sludge is a byproduct of water treatment plants, and its use as a soil stabilizer has gained increasing attention due to its economic and environmental benefits. Its application has been shown to improve the strength and stability of soil, making it suitable for various engineering applications. However, to go beyond just measuring the effects of alum sludge as a soil stabilizer, this study investigates the potential of artificial intelligence (AI) methods for predicting the California bearing ratio (CBR) of soils stabilized with alum sludge. Three AI methods, including two black box methods (artificial neural network and support vector machines) and one grey box method (genetic programming), were used to predict CBR, based on a database with nine input parameters. The results demonstrate the effectiveness of AI methods in predicting CBR with good accuracy (R^2 values ranging from 0.94 to 0.99 and MAE values ranging from 0.30 to 0.51). Moreover, a novel approach, using genetic programming, produced an equation that accurately estimated CBR, incorporating seven inputs. The analysis of parameter sensitivity and importance, revealed that the number of hammer blows for compaction was the most important parameter, while the parameters for maximum dry density of soil and mixture were the least important. This study highlights the potential of AI methods as a useful tool for predicting the performance of alum sludge as a soil stabilizer.

Keywords: alum sludge; soil stabilization; artificial intelligence; California bearing ratio; genetic programming

1. Introduction

Alum sludge (also known as red mud or bauxite tailings) is a byproduct of the alum used in the drinking water treatment process [1]. In recent decades, several studies have examined the use of alum sludge as a soil stabilizer [2]. As a soil stabilizer, alum sludge has the main advantage of improving the soil's physical properties, such as strength and stability [3]. The reason for this, is that alum sludge contains high levels of alum, which reacts with the soil to form a solid and stable structure [4]. As a result of this reaction, soil density and compaction increase, improving soil stability and reducing erosion sensitivity [4]. In addition, for pure aluminum sludge, the stability gained by adapted scheduling of bauxite tailings, is critical to warrant safe tailings deposition [5].

Alum sludge, a byproduct of the water treatment process, can be used as a soil stabilizer, providing several environmental benefits. Firstly, it can reduce soil erosion. The application of alum sludge to soil can help to reduce soil erosion by increasing its stability and reducing its susceptibility to erosion from wind or water [6]. Secondly, by improving soil structure, alum sludge can also help to reduce runoff [7], which can carry pollutants and sediment into nearby waterways, damaging aquatic ecosystems. Thirdly,

alum sludge can reduce water usage in agriculture [8], by improving soil moisture retention, reducing the need for irrigation and saving water resources. Moreover, by improving soil structure and nutrient availability, alum sludge can reduce the use of fertilizers [9], which can contribute to water pollution and environmental degradation. Furthermore, using alum sludge can reduce greenhouse gas emissions [10] by reducing the emissions associated with the production and transportation of synthetic fertilizers and irrigation. Alum sludge offers another advantage from an environmental standpoint, which is a reduction in waste in landfills [11]. Alum sludge can be diverted from landfills and used as a soil stabilizer, reducing the amount of waste going into landfills and its associated environmental impacts. Finally, the utilization of alum sludge as a soil stabilizer promotes sustainable agriculture [12], by improving soil health and productivity while reducing environmental impacts.

According to Nguyen et al. [4], alum sludge as a soil stabilizer reduces the soil's flexibility and makes it less susceptible to deformation. This reduction in flexibility results in improved soil structure and increased stability [4]. Moreover, on the geotechnical side, Nguyen et al. [4] showed that, the addition of sludge to the soil, increases the undrained compression strength (UCS) of the soil. In addition, several studies have demonstrated that alum sludge can be an effective alternative to traditional soil stabilizing agents, such as cement and lime [13–15].

Alum sludge as a soil stabilizer can have several environmental advantages [16]. The first benefit, is that it reduces the amount of waste produced by the alum production process, by recycling the sludge and reusing it for soil stabilization [16]. Furthermore, alum sludge reduces the need for conventional soil stabilizers, which are often derived from non-renewable sources and may emit large amounts of carbon dioxide [17]. In addition to improving soil stability, alum sludge can also improve soil chemical properties [16]. Studies have shown, that alum sludge can neutralize soil acidity and improve soil pH [18,19]. The importance of this, is that soil acidity can inhibit plant growth and reduce crop yields [18]. As a result, better water infiltration and retention occurs in the soil, which is essential for the growth and development of plants. Additionally, improving soil permeability can reduce soil erosion and contribute to improving soil health [20]. Alum sludge is cost-effective when used as a soil stabilizer, compared to traditional soil stabilizers such as cement, coal, and fly ash [21].

As a soil stabilizer, alum sludge has been shown to have varying efficiency, depending on the type of soil and the conditions of its application. As an example, clay soils are naturally active, so the use of alum sludge may greatly benefit them [22]. Similarly, the effectiveness of alum sludge as a soil stabilizer can be affected by environmental factors such as temperature, humidity, and rainfall [23]. Studies have shown that it is important to consider the specific soil and environmental conditions when using alum sludge as a soil stabilizer [23,24].

Various factors have been identified as influencing the behavior of sludge. There are several parameters to consider, including the soil and sludge density, the specific gravity of the soil, the liquid limit and plasticity of the soil, the sludge content, etc. With such a multiplicity of parameters and nonlinearity, it would be necessary to predict performance using an equation with a large number of parameters, that would require quantification, which does not exist. In the past two decades, artificial intelligence (AI) methods have been used to resolve this issue. Using artificial intelligence techniques, it is possible to determine the relationship between different parameters with a high degree of accuracy, without prior knowledge. Various topics in geotechnical engineering, such as slope stability [25–27], tunneling [28–30], pavement and road construction [31,32], soil cracking [33–35], rock mechanics [36,37], soil dynamics [38–41], and soil stabilizers [42–44] have been addressed using artificial intelligence methods [45]. Nevertheless, only two studies have used artificial intelligence to predict the properties arising from mixing sludge with soil [46,47]. Aamir et al. [46] used a small database of 18 datasets to predict the CBR parameter, using an artificial neural network (ANN). A high R^2 of 0.99 was observed, as well as a small RMSE of

0.057. Similarly, Shah et al. [47] used a 21-set database and artificial neural network (ANN) method to predict the CBR of mixtures of sludge and soil. The results of the study were very promising, with an R^2 of 0.97 and an RMSE of 0.58. However, both studies used only one classic artificial intelligence method, ANN, and grey box methods were not employed. Grey box methods are those methods that have an output such as a tree or an equation. In addition, the used databases were relatively small, which may limit the range of values of the parameters.

In order to fill these gaps, this study uses two black-box AI methods, namely an artificial neural network (ANN) and support vector machines (SVM), as well as a grey-box AI method, namely genetic programming (GP), in order to predict CBR. By using this approach, both black-box and grey-box models will be examined. The ANN method that is currently used as standard, will be compared with another black-box method, SVM. SVM and GP have been widely used in the field of civil engineering for various applications. For example, SVM has been used to predict the remaining service life of aging infrastructure, such as bridges and buildings [48–50]. SVM has also been used for damage detection and identification in structures, using vibration data [51,52]. SVM has found application in predicting the degradation of pavements and estimating their remaining service life [53,54]. Additionally, SVM has been utilized for classifying the types of distresses on pavement surfaces and forecasting their friction coefficient [55]. GP has been utilized for optimizing the operation and design of various systems in civil engineering, including water supply [56] and stormwater management systems [57], and for analyzing flood forecasting and dam failure [58]. Furthermore, GP has been applied to optimize the design of various structures, including trusses [59], frames [60], and bridges [61], as well as for the placement of sensors to monitor the structural health of buildings [62].

In this study, a database of 27 CBR test results, on a variety of soil types, is used for this purpose. Initially, there are nine input parameters from laboratory tests, whose values are discussed in detail. The sensitivity of the AI models, as well as the importance of the input parameters, is evaluated.

2. Data-Driven Modeling

After analyzing the collected database, the CBR was predicted using three artificial intelligence methods, including artificial neural network (ANN), genetic programming (GP), and support vector machine (SVM). The hypothesis of the performed study, is that artificial intelligence (AI) methods can be utilized to predict the California bearing ratio (CBR), which is a crucial parameter in geotechnical engineering, that measures the load-bearing capacity of soil, of soil stabilized with alum sludge, a byproduct of water treatment plants.

The California bearing ratio (CBR) is a penetration test performed on a compacted sample of soil or aggregate, to evaluate its strength and bearing capacity. It is a standardized test method used to determine the relative strength of a soil sample by measuring the pressure required to penetrate the sample with a plunger of standard area at a standardized rate. The CBR value is expressed as a percentage of the pressure required to penetrate a standard material, typically crushed rock or limestone, at the same penetration depth and under the same conditions. A higher CBR value indicates greater strength and load-bearing capacity of the soil or aggregate sample. The CBR test is widely used in civil engineering and construction to determine the design parameters of roads, airfields, and other infrastructure projects.

The use of AI algorithms for material characterization and design, has been met with skepticism, due to concerns about the reliability of their complex models. The lack of transparency and knowledge extraction processes in AI-based models, is a major challenge. Mathematical modeling techniques can be categorized as white-box, black-box, and grey-box, depending on their level of transparency (Figure 1). While white-box models are based on first principles and provide an explanation of the underlying physical relationships of a system, black-box models do not provide any feasible structure of the model. Grey-box models, on the other hand, identify the patterns between the data, and provide a

mathematical structure of the model. ANN is a popular black-box modeling technique widely used in engineering, but its weights and bias representation does not provide details about the derived relationships. GP is a newer grey-box modeling technique, that uses an evolutionary process to develop explicit prediction functions, making it more transparent than other ML methods, especially black-box methods such as ANN and SVM. The mathematical structures derived by GP can be used to gain important information about the system's performance.
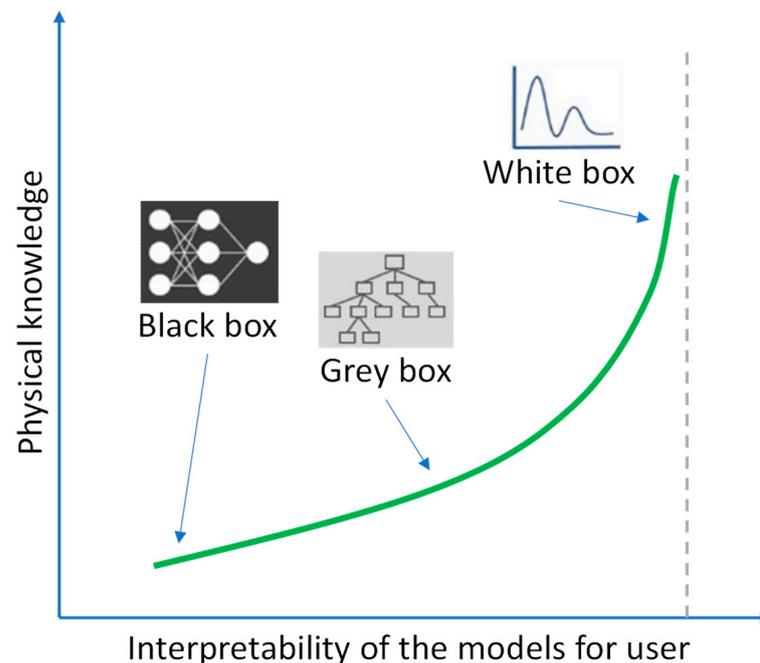
Figure 1. Classification of the AI modeling techniques (adapted from Giustolisi et al. [63] and Zhang et al. [64]).

In civil engineering, SVM has been applied for predicting soil properties [65], classifying building materials [66], and detecting structural damage [67]. For instance, SVM has been used for predicting the shear strength of soil, by using the geotechnical parameters as input variables [68]. SVM has also been employed for detecting structural damage, by using the vibration signals from a structure as input variables [69].

In civil engineering, GP has been used for optimization problems, such as the optimal design of structures and construction scheduling [70,71]. For example, GP has been applied for optimizing the design of steel frames in high-rise buildings, by evolving the design variables such as beam size, column size, and connection details [72].

2.1. Support Vector Machine (SVM)

Support vector machine (SVM) is one of supervised learning models. SVM initially developed by Vapnik [73], weights are calculated based on input data, known as training data, in order to learn the governing function from a set of inputs. SVM selects a limited number of input sample vectors, which are always a fraction of the total number of samples. These input vectors are referred to as support vectors. Using these input vectors, the parameter values that minimize a cost function are calculated. As a result, SVM requires much less data than similar methods such as ANN, and as a result, it takes much less time and has a lower cost.

The goal of this method, is to find a classifier that separates the data and maximizes the distance between these two classes. Figure 2 shows a plane cloud, that separates two sets of points.
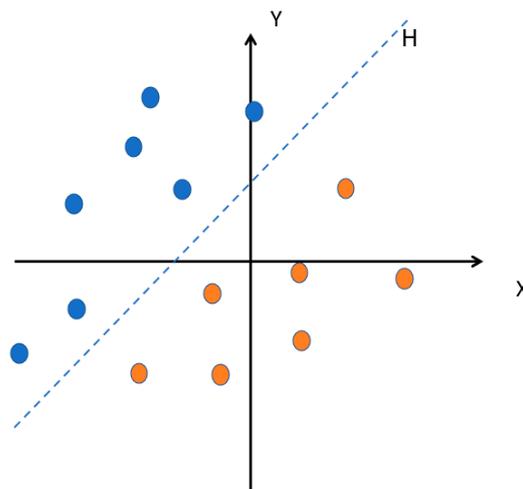
Figure 2. The hyperplane H, that separates the two groups of points.

According to Figure 3, the closest points used to determine the hyperplane, are called support vectors. As is clear, there are many hypermaps that can divide samples into two categories. The principle of SVM, is to choose something that maximizes the minimum distance between the hypermap and the training samples (that is, the distance between the hypermap and the support vectors), this distance is called the margin.
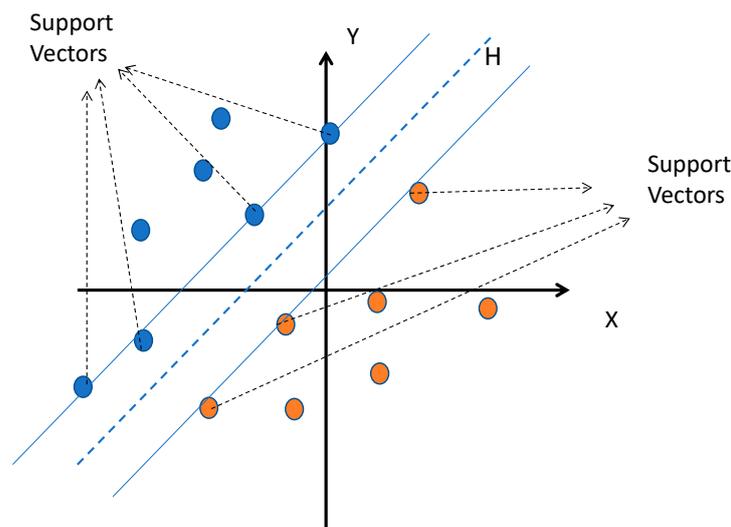
Figure 3. Support vectors.

The kernel function, as an important part of the SVM method, receives data as input and transforms them into the required form. Various functions are provided for this purpose. Some of these functions, including linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid, are given in Equations (1)–(5). According to Equations (1)–(5), the kernel function is a series of mathematical functions that provide a window for data manipulation. With the help of this transformation, a complex and nonlinear level of decision making becomes a linear equation, but in a larger number of dimensional spaces.

$$\text{Polynomial kernel}: \ k = k(x_i, x_j) = (x_i.x_j + 1)^d \tag{1}$$

$$\text{Gaussian kernel}: \ k = k(x_i, x_j) = exp\left(-\frac{\left\| x_i - x_j \right\|^2}{2\sigma^2}\right) \tag{2}$$

$$\text{Radial basis function (RBF)}: \ k = k(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|^2\right) \qquad (3)$$

$$\text{Hyperbolic } \tan \text{gent kernel}: \ k = k(x_i, x_j) = tanh(k\prime x_i.x_j + c) \qquad (4)$$

$$\text{Sigmoid kernel}: \ k = k(x_i, x_j) = tanh\left(\alpha x^T y + c\right) \qquad (5)$$

where x_i and x_j are vectors in the input space, the γ and σ parameters define the distance influence of a single sample, d is the degree of the polynomial, $c \in R$ is the relative position to the origin, $\kappa\prime$ is a real-valued positive type function/kernel, and α, T, and c are certain values.

In this study, two important functions, i.e., the radial basis function (RBF) and Gaussian kernel, which have been successful in past research, were evaluated, along with two other functions, i.e., the sigmoid kernel and polynomial kernel functions.

2.2. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) have a history that dates back to the mid-20th century, when the mathematical model for the perceptron was first introduced by Rosenblatt [74]. The development of the backpropagation algorithm in the 1980s, by Rumelhart et al. [75], improved the training of multi-layer ANNs and led to their widespread use in various fields. Since then, ANNs have continued to evolve, with the development of new architectures and learning algorithms [76].

An artificial neural network (ANN) is a machine learning model that is based on the structure and function of the human brain [76]. Artificial neural networks are composed of interconnected nodes, or artificial neurons, that process information and make predictions in response to inputs. By combining the inputs and the weights assigned to each connection, the interconnected nodes produce an output. There are a number of applications for artificial neural networks, including image classification, speech recognition, and natural language processing, among others. In the field of machine learning and artificial intelligence, they have proven to be effective at solving complex problems.

Due to their ability to learn from data and make accurate predictions or classifications based on that data, artificial neural networks (ANNs) have become one of the most powerful tools in the field of machine learning and artificial intelligence [77]. As a result, they simulate the structure and function of the human brain, which is composed of many neurons connected by synapses.

Each node or artificial neuron in an ANN, performs simple computations based on the input it receives and the weights assigned to the connections [45]. One neuron's output is then passed on as input to other neurons in the network, allowing the information to be processed and transformed multiple times before it is finally produced. Feedforward, refers to this process of processing and transforming information through multiple layers of neurons.

There are various types of artificial neural networks, each designed for a particular task or application. The most popular types of ANNs include feedforward networks, recurrent networks, and convolutional neural networks [78]. An ANN can be used for a variety of purposes, and each type has its own strengths and weaknesses.

As the most basic type of ANN, feedforward networks are used for simple tasks, such as binary classification and linear regression. On the other hand, recurrent networks are used for sequences of data, such as speech or text, and are capable of capturing temporal relationships. Image or video analysis can be performed using convolutional neural networks, which are specifically designed to capture the spatial relations between pixels in an image or video.

Artificial neural networks (ANNs) can be used to model complex relationships with no established mathematical relationships [79]. The ANN method consists of artificial neurons, and is commonly used to fit nonlinear statistical data. For this purpose, ANN identifies the

relationships between input and output parameters based on the strength of the connection between the two neurons, known as the weight. Next, the network seeks to optimize the matrix of weights, which is achieved through practice and adjustment, which is called the paradigm [80]. One of the strongest paradigms, is the backpropagation paradigm. In this study, two backpropagation algorithms are used: (i) the Levenberg–Marquardt (LM) algorithm [81], and (ii) the Bayesian regularization (BR) [82] algorithm.

Input, hidden, and output layers are the three main components of ANN architecture. In this study, for each algorithm, i.e., LM and BR, one, two, three, four, and five hidden layers are considered. Log-sig and tan-sig are transfer functions within each neuron. In this study, five different architectures of ANN, with one to five hidden layers, were modelled for each algorithm. Furthermore, by trial and error, 45 neurons were considered for each hidden layer and each network was re-trained three times.

2.3. Genetic Programming (GP)

John Koza, a computer scientist and researcher at Stanford University, developed the first genetic programming system in the early 1990s. Using this system, he evolved computer programs capable of solving mathematical problems and controlling robots. Several genetic programming frameworks and toolkits were developed as a result of this development.

As a subfield of artificial intelligence and evolutionary computation, genetic programming (GP) is a method for solving complex problems, using a process inspired by natural selection and genetics. In this approach, potential solutions are represented as trees of operations (also known as computer programs) and then improved using genetic algorithms.

The GP algorithm generates an initial population of computer programs at random and evaluates their fitness in accordance with a problem-specific objective function. Through genetic operators such as crossover and mutation, the best programs are selected, to produce the next generation. The process continues until a satisfactory solution is found or a stopping criterion is met.

A wide range of problems can be solved using GP, including function approximation, symbolic regression, and even game play. There are several strengths of GP, including its ability to find complex solutions that are difficult or impossible to discover manually, as well as its ability to handle high levels of uncertainty and noise in the data. It can, however, be computationally expensive, and may have difficulty finding solutions in a reasonable amount of time, for very complex problems.

3. Database Collection and Processing

3.1. Experiment and Data Collection

The database was prepared based on the results of two studies [4,47]. A 3D representation of the collected database for CBR is shown in Figure 4. The input parameters for this study were carefully selected, to ensure the accurate prediction of the California Bearing Ratio (CBR) value for alum sludge used in road construction. The database consisted of nine parameters, including the liquid limit (LL) and plasticity index (PI) of soil, the specific gravity (G_s) of soil, the number of compaction blows, the optimum moisture content of soil, the maximum dry density of soil, the optimum moisture content of the mixture, the maximum dry density of the mixture, and the sludge content. These parameters were chosen, to consider different basic geotechnical tests, including compaction, Atterberg limits, and sludge content. The main goal of this study, was to predict the CBR value, which is crucial in the design of pavement and road construction. By selecting these input parameters, the researchers ensured that their predictions would be accurate and reliable, which is essential for the success of any road construction project.
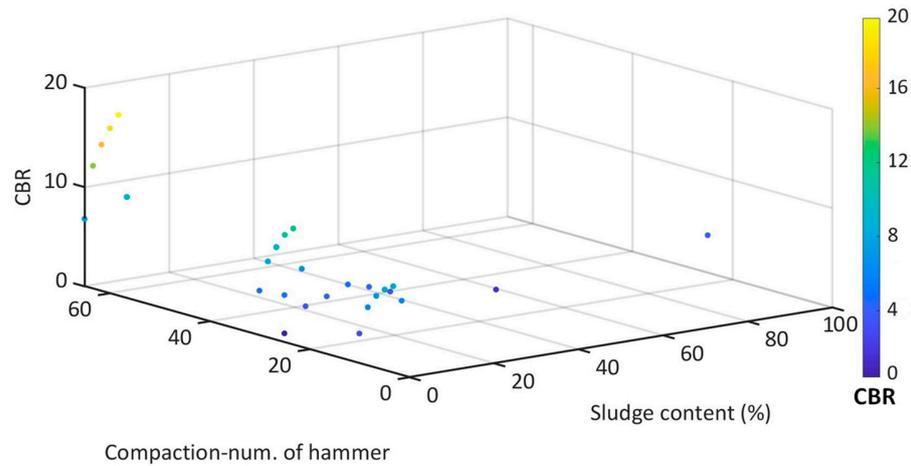
Figure 4. Database used for modeling based on the CBR.

Figure 5 shows the effect of different input parameters, such as number of compaction blows, sludge content, maximum dry density, and optimum water content, on the CBR. According to the results in Figure 5a, with an increase in the number of compression hammer blows, the CBR increases dramatically. Figure 5b illustrates the effect of sludge content on CBR. It can be seen from Figure 5b that as the sludge content increases, CBR increases significantly at first. When the sludge content reaches around 8%, CBR begins to decrease as the sludge percentage increases. Figure 5c shows the effect of increasing the maximum dry density (MDD) of the soil on the CBR of the soil. According to the results, it can be concluded that the CBR increases as the maximum dry density of the soil increases. Figure 5d shows the effect of optimum moisture content (OMC) on CBR. The results show that the CBR decreases as the optimum water content increases.
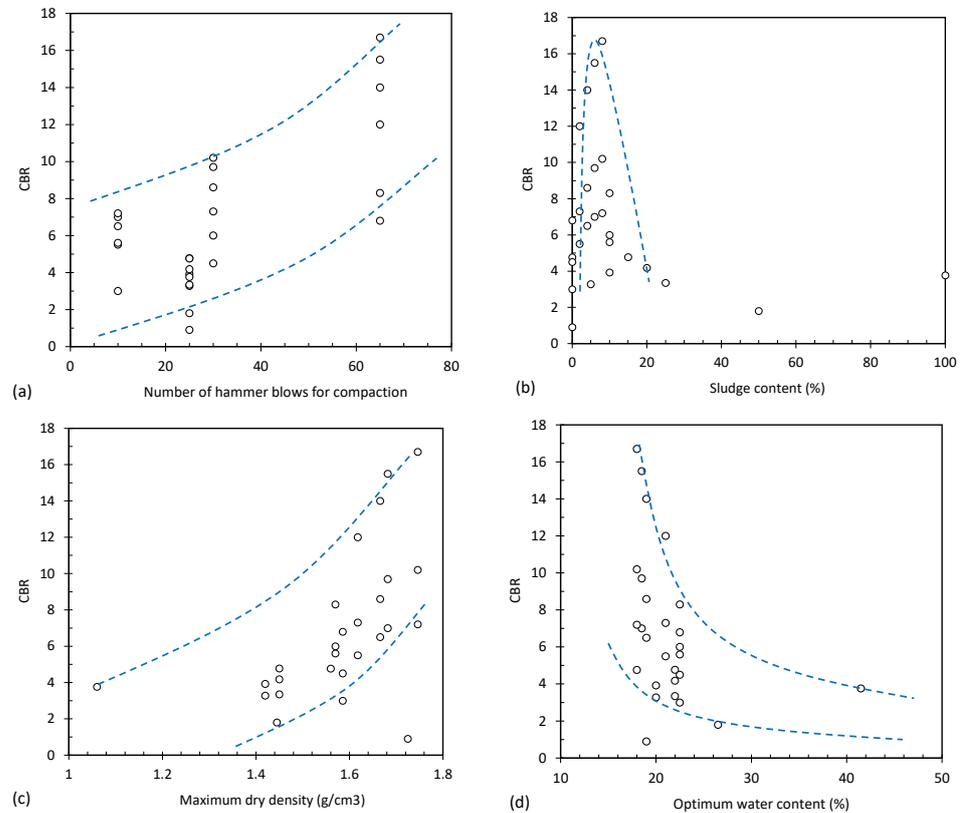

Figure 5. Effect of (**a**) number of compaction blows, (**b**) sludge content, (**c**) MDD, and (**d**) OMC on the CBR in the collected database.

In order to provide a more detailed analysis of the database, Table 1 shows the statistical information for each variable, including the minimum, maximum, mean, and standard deviation.

Table 1. The statistical information of the variables in the database.

Variable	Observations	Minimum	Maximum	Mean	Std. Deviation
CBR	27	0.900	16.700	6.856	4.013
LL (soil)	27	26.120	55.000	46.391	12.660
PI (soil)	27	8.830	34.000	26.239	11.399
% Sludge	27	0.000	100.000	11.667	20.438
Compaction (number of blows)	27	10.000	65.000	31.667	19.513
OMC (soil)	27	18.000	22.500	21.111	2.021
MDD (soil)	27	1.560	1.725	1.596	0.048
OMC (mixture)	27	18.000	41.500	21.315	4.545
MDD (mixture)	27	1.060	1.746	1.577	0.146
G_s (soil)	27	2.170	2.750	2.616	0.243

### *3.2. Preparation of the Data for AI Modeling*

3.2.1. Normalization

In a database, each input or output variable has a specific unit. The data normalization function can eliminate the weight of the units, reduce network errors, and increase training speed, by normalizing the data to a value between zero and one. The linear normalization function utilized in this study is outlined below.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{6}$$

The four terms in this equation are X_{max}, X_{min}, X, and X_{norm}, which correspond to the maximum, minimum, actual, and normalized values, respectively.

3.2.2. Testing and Training Databases

There are two types of databases required for the implementation of mathematical models, namely, training databases and testing databases. For all three mathematical models, training and testing were randomly divided. Figure 6 shows the main division of the database, where 80% of the main database was used for training and 20% for testing.
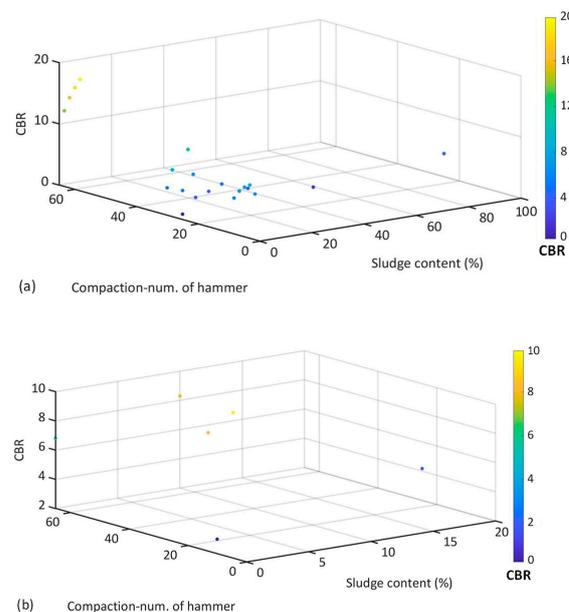
Figure 6. Distribution of (**a**) training and (**b**) testing databases.

A random database division resulted in two test and experiment databases that were considered fixed, and were used in all three mathematical models. The statistical parameters for the training and test databases, are shown in Tables 2 and 3, respectively. The statistical information, including the minimum, maximum, mean, and standard deviation, is very similar for both databases. This can increase the accuracy of the network in predicting the output.

Table 2. The statistical information of the training database.

Variable	Observations	Minimum	Maximum	Mean	Std. Deviation
CBR	21	0.900	16.700	6.883	4.379
LL (soil)	21	26.120	55.000	45.252	13.046
PI (soil)	21	8.830	34.000	25.220	11.731
% Sludge	21	0.000	100.000	13.095	22.784
Compaction (number of blows)	21	10.000	65.000	30.000	18.841
OMC (soil)	21	18.000	22.500	20.929	2.075
MDD (soil)	21	1.560	1.725	1.600	0.054
OMC (mixture)	21	18.000	41.500	21.357	5.094
MDD (mixture)	21	1.060	1.746	1.573	0.161
G_s (soil)	21	2.170	2.750	2.605	0.250

Table 3. The statistical information of the testing database.

Variable	Observations	Minimum	Maximum	Mean	Std. Deviation
CBR	6	3.000	9.700	6.763	2.654
LL (soil)	6	27.280	55.000	50.380	11.317
PI (soil)	6	8.830	34.000	29.805	10.276
% Sludge	6	0.000	20.000	6.667	7.554
Compaction (number of blows)	6	10.000	65.000	37.500	22.528
OMC (soil)	6	18.000	22.500	21.750	1.837
MDD (soil)	6	1.560	1.586	1.582	0.011
OMC (mixture)	6	18.500	22.500	21.167	1.889
MDD (mixture)	6	1.450	1.682	1.590	0.083
G_s (soil)	6	2.170	2.750	2.653	0.237

3.2.3. Statistical Parameters

The performance of a network can be evaluated through a variety of parameters, such as coefficient of determination (R^2) and mean absolute error (MAE). Equations (7)–(12) show the definitions of mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean squared logarithmic error (MSLE), root mean squared logarithmic error (RMSLE), and coefficient of determination (R^2).

$$\text{MAE} = \frac{\sum_N (X_m - X_p)}{N} \tag{7}$$

$$\text{MSE} = \frac{\sum_N (X_m - X_p)^2}{N} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_N (X_m - X_p)^2}{N}} \tag{9}$$

$$\text{MSLE} = \frac{\sum_N (\log (X_m + 1) - \log (X_p + 1))^2}{N} \tag{10}$$

$$\text{RMSLE} = \sqrt{\frac{\sum_N \left(\log\left(X_m + 1\right) - \log\left(X_p + 1\right)\right)^2}{N}} \tag{11}$$

$$\text{R}^2 = \left[\frac{\sum_{i=1}^{N}\left(X_m - \bar{X_m}\right)\left(X_p - \bar{X_p}\right)}{\sum_{i=1}^{N}\left(X_m - \bar{X_m}\right)^2 \sum_{i=1}^{n}\left(X_p - \bar{X_p}\right)^2}\right]^2 \tag{12}$$

where N is the number of datasets, X_m and X_p are the actual and predicted values, respectively, and $\bar{X}_m$ and $\bar{X}_p$ are the averages of the actual and predicted values, respectively. Ideally, the model should have an R^2 value of 1 and MAE, MSE, RMSE, MSLE, and RMSLE values of 0.

4. Results

4.1. Support Vector Machine (SVM)

In order to derive an optimal SVM model, multiple parameters were adjusted, and the resulting model with the most favorable performance is delineated in this section. Table 4 presents the specifications of the best support vector machine (SVM) model derived through parameter tuning. The model was constructed using the sequential minimal optimization (SMO) algorithm, with a penalty parameter of $C = 2$, and a tolerance value of 0.001. The epsilon parameter, which controls the width of the margin, was set to 0.5. The input data was pre-processed using standardization, to ensure that each feature had a mean of zero and a standard deviation of one. The kernel function used in this model was the radial basis function (RBF), which is a popular kernel function for SVMs due to its flexibility in modeling complex, nonlinear relationships between input variables. The gamma parameter of the RBF kernel was set to 0.5, which controls the smoothness of the decision boundary. A lower value of gamma, results in a smoother decision boundary, while a higher value of gamma, leads to a more complex boundary, that can overfit the data. Overall, these specifications provide a well-performing SVM model, that can accurately classify data while avoiding overfitting. These parameters can be used as a starting point for future SVM modeling tasks or as a benchmark for comparing the performance of other SVM models.

**Table 4.** The specifications of the best performing SVM.

SMO Parameters				Kernel Parameters	
C	Tolerance	Epsilon	Pre-processing	Type of Kernel	Gamma
2	0.001	0.5	Standardization	Radial basis function (RBF)	0.5

The predicted values of CBR against the values obtained from the tests in the laboratory for the training and testing database, are shown in Figure 7. The results show that the SVM model has been able to predict the CBR values with a good accuracy.

Table 5 presents the performance metrics of the SVM used to predict the California bearing ratio (CBR) values for mixtures of alum sludge and soil. Both training and testing datasets were used to evaluate the performance of the SVM model.

Table 5. Results of SVM predictions of CBR for mixtures of alum sludge and soil.

Performance Metrics	Training Database	Testing Database
MAE	0.497	0.512
MSE	0.409	0.357
RMSE	0.640	0.598
MSLE	0.017	0.011
RMSLE	0.129	0.103
R^2	0.978	0.939

(a)

(b)
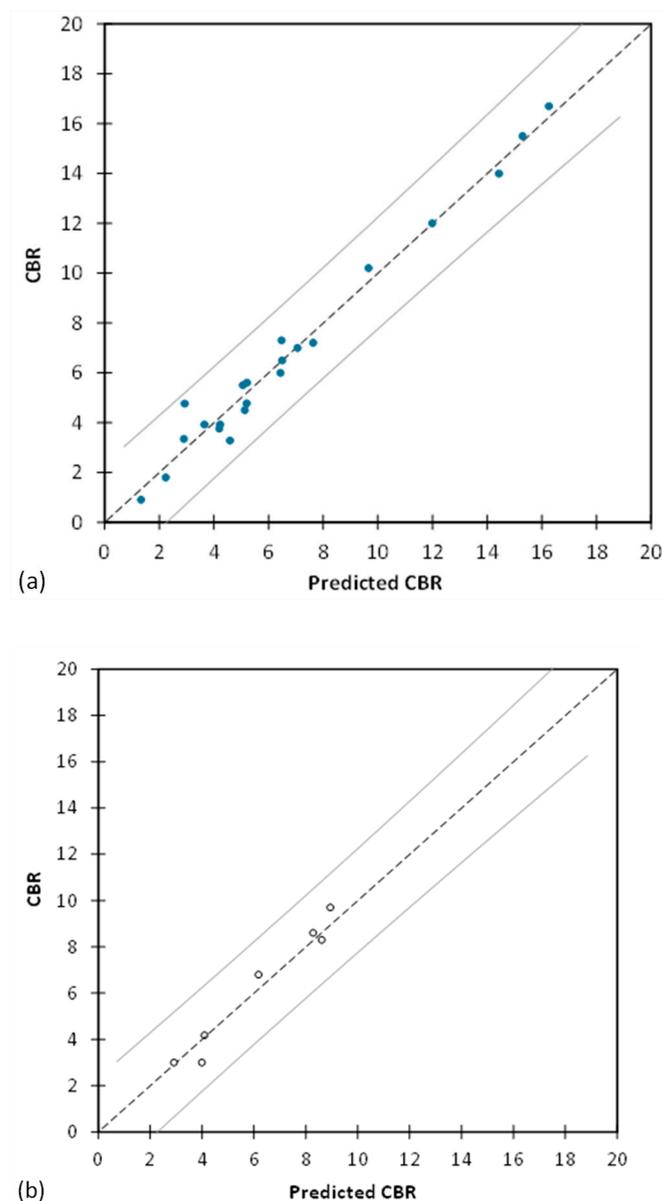
Figure 7. The results of the best SVM model predicting CBR values for the (**a**) training and (**b**) testing databases.

The first performance metric is the mean absolute error (MAE), which measures the average absolute difference between the predicted and actual CBR values. In the training and testing sets, the values are 0.497 and 0.512, respectively, indicating that the model has an average error of approximately 0.5. A second performance metric is the mean squared error (MSE), which measures the average squared difference between the predicted and actual CBR values. The training and testing sets have values of 0.409 and 0.357, respectively, indicating that the model has a relatively low overall error in predicting the CBR values. The third performance metric, is the root mean squared error (RMSE), which is the square root of the mean square error (MSE). The training and testing sets have values of 0.640 and 0.598, respectively, indicating that the model has a relatively low overall error in predicting the CBR values. The fourth performance metric is the mean squared log error (MSLE). The MSLE measures the difference between the logarithm of the predicted CBR values and the logarithm of the actual CBR values. For the training and testing sets, the values are 0.017 and 0.011, respectively, indicating that the model has a relatively low overall error when predicting the CBR values. The fifth performance metric is the root mean squared log

error (RMSLE), which is the square root of the MSLE. In the training and testing sets, the values are 0.129 and 0.103, respectively, which indicates that the model has a relatively low overall error in predicting the CBR values, on a logarithmic scale. The final performance metric is the R^2, which measures the amount of variance in the CBR values that can be explained by the model. As can be seen from the values for the training and testing sets, the model has a high degree of correlation between the predicted and actual CBR values. The results suggest that the SVM model is capable of predicting CBR values for mixtures of alum sludge and soil, as indicated by the low error values and high R-squared values, for both the training and testing sets.

4.2. Artificial Neural Network (ANN)

In this study, all artificial neural network (ANN) modeling was performed using MATLAB (R2020a: The Math Works Inc., Natick, MA, USA). Table 6 displays the results of the Bayesian regularization (BR) and Levenberg–Marquardt (LM) algorithms for all networks. The table contains the R^2 and mean absolute error (MAE) values for both test and training datasets. The results are based on the best-performing network among all 45 neurons and three training repetitions, as determined by the highest R^2 and lowest MAE values.
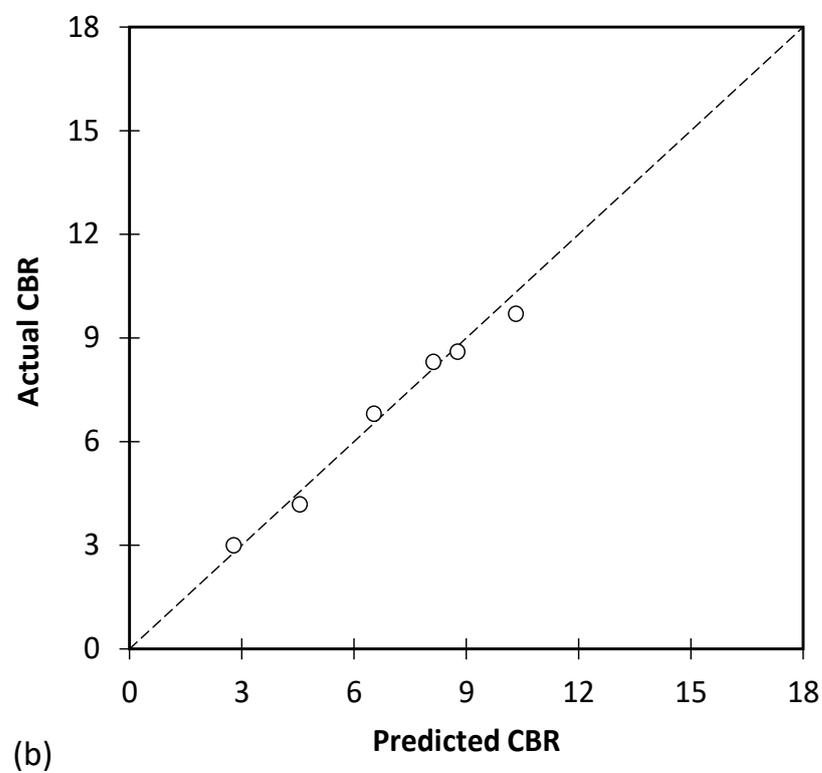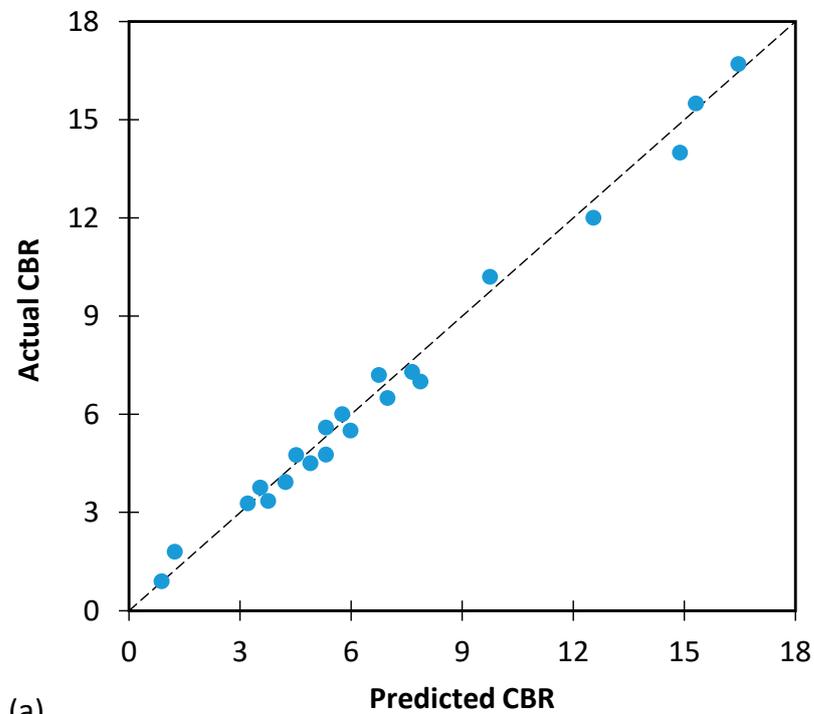
Table 6. The results of ANN modeling.

	Number of Hidden Layers	R-Test	R-Train	MAE-Test	MAE-Train
Bayesian Regularization	1H	0.943	0.955	0.563	0.598
	2H	0.980	0.989	0.302	0.392
	3H	0.972	0.977	0.405	0.431
	4H	0.970	0.976	0.463	0.452
	5H	0.963	0.971	0.490	0.523
	Average	0.965	0.973	0.444	0.479
Levenberg–Marquardt	1H	0.913	0.928	0.953	0.835
	2H	0.956	0.969	0.512	0.463
	3H	0.942	0.958	0.673	0.574
	4H	0.933	0.951	0.753	0.682
	5H	0.918	0.937	0.841	0.797
	Average	0.932	0.948	0.746	0.670

The findings indicate, that the network with two hidden layers achieved the highest performance for both the BR and LM algorithms. Moreover, the average accuracy of the BR algorithm was superior to that of the LM algorithm.

Figure 8 depicts the comparison between the predicted values of the artificial neural network (ANN) model and the actual values of laboratory experiments. The results indicate that the ANN model exhibits a very high level of precision in predicting the California bearing ratio (CBR) values for both the training and testing databases.

Table 7 displays the best outcomes of utilizing the ANN modeling method to predict the California bearing ratio (CBR) for the mixtures of alum sludge and soil. The performance indicators of the ANN model are exhibited for both the training and testing datasets. The ANN model yielded a mean absolute error (MAE) of 0.392 and 0.303, for the training and testing datasets, respectively; the mean squared error (MSE) values for the training and testing datasets were 0.200 and 0.116, respectively; the root mean squared error (RMSE) values obtained for the training and testing datasets were 0.447 and 0.341, respectively; the mean squared log error (MSLE) values were 0.005 and 0.002, for the training and testing datasets, respectively; while the root mean squared log error (RMSLE) values were 0.074 and 0.046, for the training and testing datasets, respectively. The ANN model yielded

a high coefficient of determination (R^2), of 0.989 and 0.980, for the training and testing datasets, respectively, implying that the model is highly precise in predicting the CBR values for mixtures of alum sludge and soil.


(a)

(b)

Figure 8. The results of the best ANN model predicting the CBR values for the (**a**) training and (**b**) testing databases.

Table 7. Results of ANN predictions of the CBR for mixtures of alum sludge and soil.

Performance Metrics	Training Database	Testing Database
MAE	0.392	0.302
MSE	0.200	0.116
RMSE	0.447	0.341
MSLE	0.005	0.002
RMSLE	0.074	0.046
R^2	0.989	0.980

One of the key determinants of the precision and complexity of an artificial intelligence (AI) model, is the number of neurons present in each hidden layer. To find the optimal number of neurons, a group of analyses is conducted. Figure 9a shows the accuracy of the ANN model, based on the number of neurons present in each hidden layer. According to the results, the accuracy of the ANN model does not improve significantly beyond a certain number of neurons, which in this case is 10 neurons. Therefore, it is essential to determine the optimal number of neurons, to achieve the best accuracy while minimizing the complexity of the model. Figure 9b displays the model's error rate for each neuron. The results show that the error rate remains almost constant after the 10th neuron. Thus, the optimal number of neurons for the ANN model was determined to be 10. This information is crucial for researchers and developers of AI models, as it allows them to minimize the complexity while achieving the best accuracy. By optimizing the number of neurons, the performance of the AI model can be significantly enhanced.
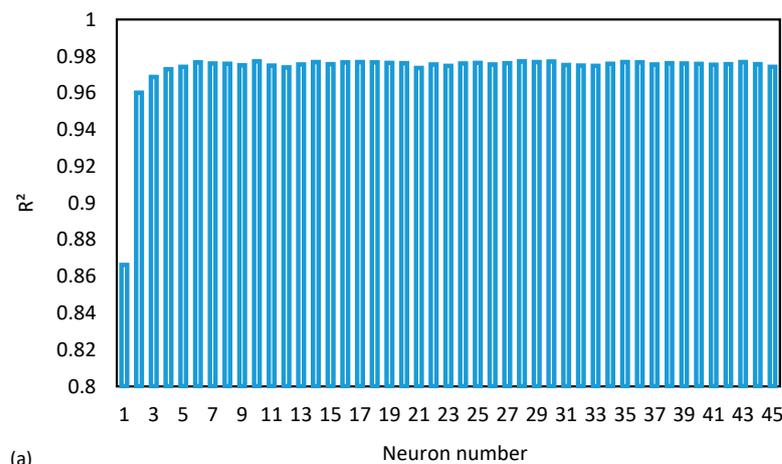

(a)
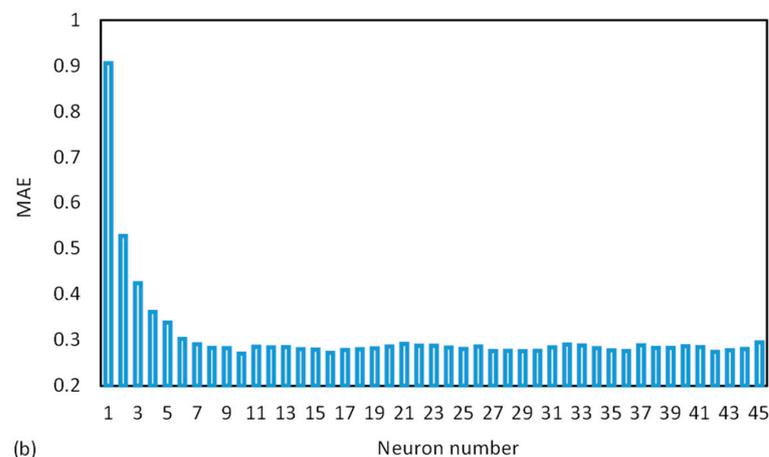
(b)

Figure 9. The (**a**) accuracy (R^2) and (**b**) error (MAE) of different numbers of neurons for the training database in the ANN model.
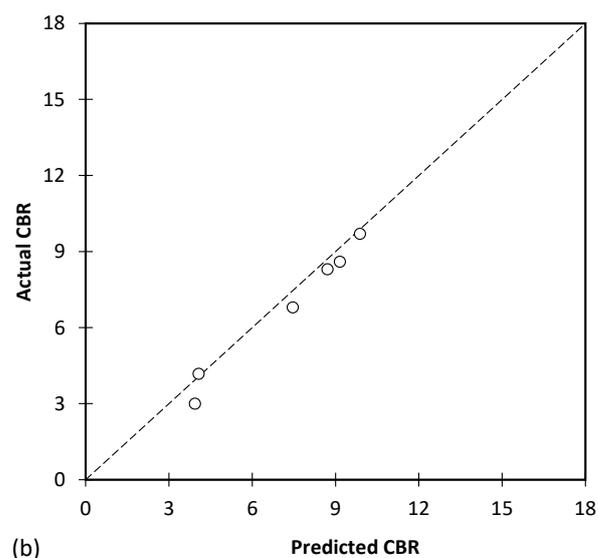
4.3. Genetic Programming (GP)

The genetics of programming involves applying evolutionary algorithms to generate computer programs through the use of genetic operators, such as crossover and mutation. Important parameters that affect the effectiveness of genetic programming, include population size, mutation rate, and fitness evaluation function. In this study, after a series of analyses, the best GP model was selected, based on its ability to produce high-quality programs for a given problem domain.

Figure 10 shows the projected values of CBR by the best GP model, against the actual values obtained in the experiments, for both the training and testing sets. According to the findings presented in Figure 10, the best genetic programming (GP) model was capable of accurately predicting the values of the California bearing ratio (CBR) in the experiments, as indicated by the close match between the model's projected values and the experimental values. These results suggest that the GP model has a high level of accuracy in predicting CBR values.
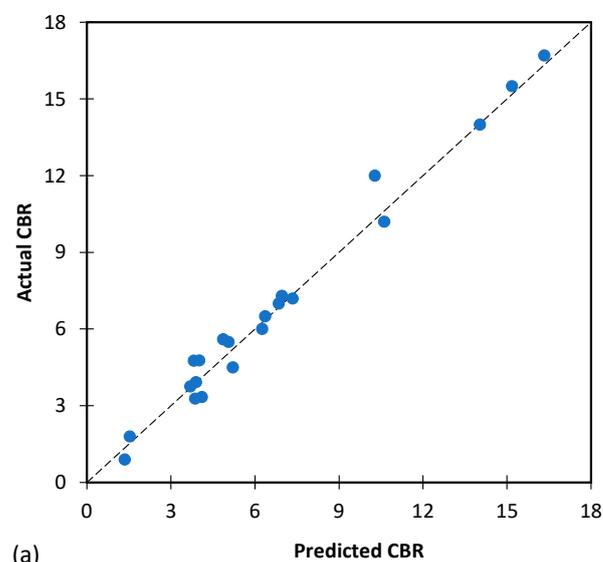
(a)

(b)

Figure 10. The results of the best GP model predicting CBR values for the (**a**) training and (**b**) testing databases.

A significant distinction between genetic programming (GP) and black-box models, is that the GP produces an equation as an outcome, which can be directly utilized by the reader. In this study, Equation (13) represents the GP model outcome for predicting California bearing ratio (CBR) values. The reason for the length of this equation, is that it incorporates seven distinct inputs, namely liquid limit, plasticity index, alum sludge content, number of compaction blows, optimum moisture content of soil and mixture, and G_s of soil. It is worth noting, that traditional methods are incapable of producing an equation such as Equation (13), with the consideration of all seven influential inputs. Consequently, this study represents the first successful attempt to formulate an equation for predicting CBR numbers based on these seven important inputs. It is important to note, that the parameters in Equation (6) are normalized values, and values should be normalized based on Table 1 before being used in Equation (13).

$$
\begin{aligned}
\text{CBR} = &\;((((0.234 + (((X_2 - 0.339) - 0.185)^2)) - ((((X_7{}^2) - X_3) - ((0.339 - X_5)^3)) \times X_6)) \times ((X_4 + \\
&((X_5{}^2) \times (0.339 + X_2)) \times ((X_3 \times X_5) \times (X_1 + 0.339)))) - ((((X_4 + 0.185) - X_7) \times ((X_6 \times X_5) - 0.234)) \\
&- ((X_5{}^3) - ((0.234^2) \times (0.234 + X_1)))))))) \qquad (13)
\end{aligned}
$$

where X_1, X_2, X_3, X_4, X_5, X_6, and X_7 are LL, PI, AS content, number of compaction blows, OMC of soil, OMC of mixture, and G_s of soil, respectively.

Table 8 presents the results of the GP model in predicting the California bearing ratio (CBR) for the mixtures of alum sludge and soil. The GP model's performance was evaluated using various performance metrics on both the training and testing datasets. The mean absolute error (MAE) of the model on the training database was 0.459, indicating that, on average, the model's predictions were off by 0.459 CBR units. The MAE on the testing database was slightly higher, at 0.475, indicating that the GP model's generalization ability was slightly worse than its performance on the training database. The mean squared error (MSE) on the training database was 0.358, indicating that the GP model's predictions had a higher variance than its performance on the testing database (0.304). The root mean squared error (RMSE) on the training database was 0.598, indicating that the GP model's predictions had a higher magnitude of error than its performance on the testing database (0.552). The mean squared logarithmic error (MSLE) on the training set was 0.010, and the corresponding value on the testing set was slightly lower, at 0.009, indicating that the GP model's predictions were more accurate on the testing set. The root mean squared logarithmic error (RMSLE) on the training set was 0.100, and the corresponding value on the testing set was slightly lower, at 0.097. Finally, the coefficient of determination (R^2) was high for both the training set (0.980) and the testing set (0.948), indicating that the model explained most of the variance in the data, and that it had a good fit to the data. In conclusion, the results suggest that the GP model can be used to accurately predict CBR values for mixtures of alum sludge and soil, and that it generalizes well to unseen data. However, the model's predictions on the training set had a higher variance and magnitude of error, compared to the testing set.

Table 8. Results of GP to predict CBR for mixtures of alum sludge and soil.

Performance Metrics	Training Database	Testing Database
MAE	0.459	0.475
MSE	0.358	0.304
RMSE	0.598	0.552
MSLE	0.010	0.009
RMSLE	0.100	0.097
R^2	0.980	0.948

One benefit of using an equation generated by genetic programming, is that it can potentially provide a more accurate prediction of the CBR value, compared to traditional methods. The equation is generated based on the data and relationships observed in the inputs and outputs, allowing for a more customized and precise model. Another benefit,

is that the generated equation can be easily implemented and applied to new data. Once the equation is generated, it can be used to predict the CBR values of new soil samples, without requiring additional testing or experimentation. A further benefit of the generated equation, is that it can reduce testing time and costs. Traditional methods of predicting CBR values require extensive testing and experimentation. Using a generated equation, can significantly reduce the time and cost involved in testing, as the equation can provide accurate predictions without the need for additional experimentation.

5. Discussion

5.1. Comparison of Different Models

Table 9 compares the results of three artificial intelligence (AI) models, namely artificial neural network (ANN), support vector machine (SVM), and genetic programming (GP), in predicting the California bearing ratio (CBR), using different performance metrics, for both the training and testing databases. The performance metrics used to evaluate the models' performance include, mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean squared logarithmic error (MSLE), root mean squared logarithmic error (RMSLE), and coefficient of determination (R^2).

Table 9. Results of all three AI models in predicting CBR for both training and testing databases.

Performance Metrics	Training			Testing		
	SVM	ANN	GP	SVM	ANN	GP
MAE	0.497	0.392	0.459	0.512	0.302	0.475
MSE	0.409	0.200	0.358	0.357	0.116	0.304
RMSE	0.640	0.447	0.598	0.598	0.341	0.552
MSLE	0.017	0.005	0.010	0.011	0.002	0.009
RMSLE	0.129	0.074	0.100	0.103	0.046	0.097
R^2	0.978	0.989	0.980	0.939	0.980	0.948

In addition, Table 10 shows the ranking of the different AI models. The SVM model was the first AI model examined in this study, and is known as a black-box model. The ANN method, another black-box model, was subsequently employed, to explore another black-box model's performance. The GP method was utilized to examine grey-box AI methods.

Table 10. Overall rank analysis of different performance parameters for different machine learning techniques.

Performance Parameters	SVM		ANN		GP	
	TR	TS	TR	TS	TR	TS
MAE	3	3	1	1	2	2
MSE	3	3	1	1	2	2
RMSE	3	3	1	1	2	2
MSLE	3	3	1	1	2	2
RMSLE	3	3	1	1	2	2
R^2	3	3	1	1	2	2
Subtotal	18	18	6	6	12	12
Total score	36		12		24	
Overall rank	3		1		2	

The results demonstrate, that the SVM model exhibits the lowest accuracy for both the training and test datasets, but still, it shows a good performance. In contrast, the ANN method performs significantly better, and predicts CBR values with high accuracy, achieving an R^2 of 0.989 and 0.980 for the training and test databases, respectively. The

GP method, a grey-box model, achieves slightly lower accuracy than the ANN method but still performs well, with an R^2 of 0.98 and 0.948 for the training and testing databases, respectively. Notably, the GP method offers the benefit of producing an equation as output, which can be used for other datasets.

The reason for the observed results, is that the SVM and GP models are known to be better suited for problems with a small number of inputs and a large number of data points, whereas ANN models are better suited for problems with a large number of inputs and a small number of data points. These findings contribute to a better understanding of the performance of different AI models for predicting CBR, which may have implications for future research and practical applications in the field of civil engineering and transportation planning.
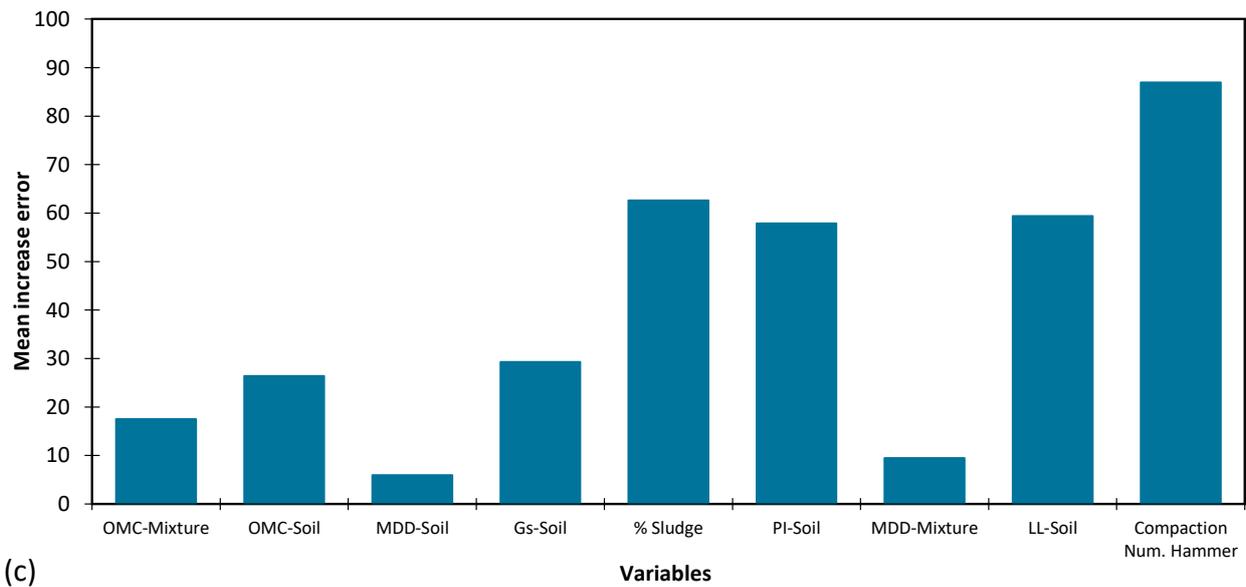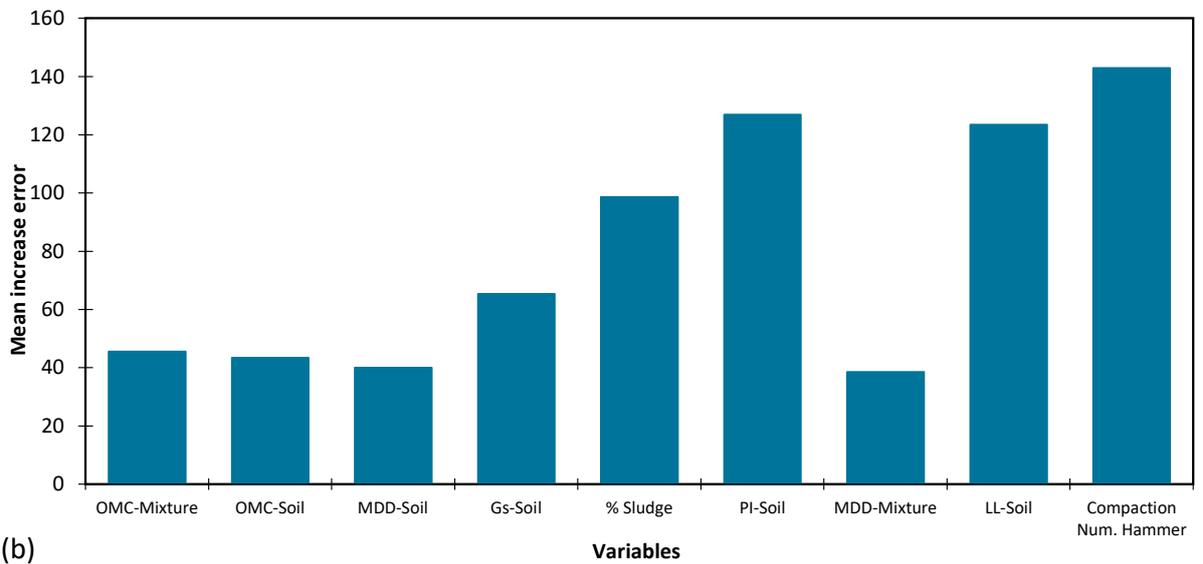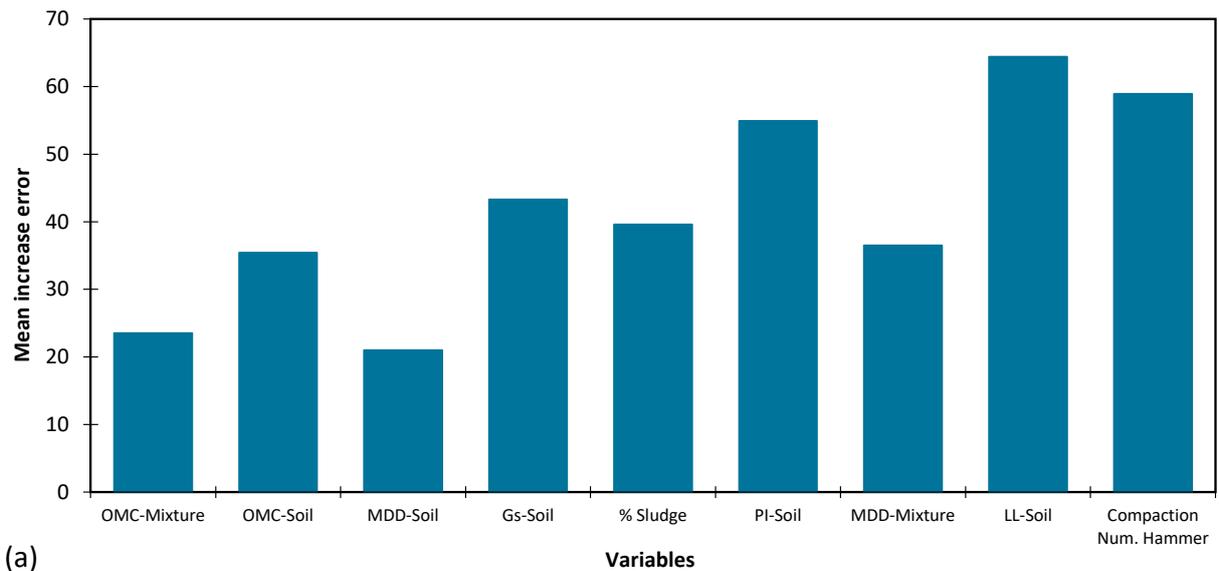
5.2. The Variable Importance of Input Parameters

Investigating the significance of the input parameters, is a crucial aspect of artificial intelligence modeling. In this study, the impact of the individual input parameters on network error was examined, by altering each parameter by 100%, while maintaining all other inputs at their actual values. The resulting network errors were recorded for each alteration, and are presented in Figure 11 for each of the three AI models tested, namely artificial neural network (ANN), support vector machines (SVM), and genetic programming (GP). A greater network error resulting from a given parameter alteration, indicates that the network exhibits increased sensitivity to that particular parameter. The variable importance ranking was evaluated based on the input parameters, including liquid limit of soil (LL-Soil), plasticity index of soil (PI-Soil), G_s of soil, number of compaction blows (Compaction Num. Hammer), optimum moisture content of soil (OMC-Soil), maximum dry density of soil (MDD-Soil), optimum moisture content of mixture (OMC-mixture), maximum dry density of mixture (MDD-Mixture), and sludge content (% Sludge). Table 11 summarizes the ranking of inputs based on the different AI models.

Table 11. The importance of different input parameters for all AI models.

Ranking	Input Parameters								
	LL-Soil	**PI-Soil**	**G$_s$-Soil**	**Compaction Num. Hammer**	**OMC-Soil**	**MDD-Soil**	**OMC-Mixture**	**MDD-Mixture**	**% Sludge**
SVM	1	3	4	2	7	9	8	6	5
ANN	3	2	5	1	7	8	6	9	4
GP	3	4	5	1	6	9	7	8	2
Total	7	9	14	4	20	26	21	23	11
Ranking	2	3	5	1	6	9	7	8	4

The results reveal, that the SVM model ranked the LL of soil as the most important input parameter, followed by the number of compaction blows and the PI of soil. While the MDD of soil was the least important parameter. The ANN ranked the number of compaction blows as the most important input parameter, followed by the PI and LL of soil. While the MDD of the mixture was the least important parameter. The GP ranked the number of compaction blows as the most important input parameter, followed by sludge content and LL of soil. For all three models, and according to Table 11, it can be concluded that the number of compaction blows and LL of soil parameters, are the most important input parameters, while the MDD of soil and mixture, are the least important.

Figure 11. The importance of input parameters on the MAE of the best (**a**) SVM, (**b**) ANN, and (**c**) GP models.

The explanation for the variation in the importance ranking of input parameters across AI models, could be the differences in the underlying algorithms and architectures. SVM is a kernel-based method, that tries to find a hyperplane that maximally separates data points. ANN is a feedforward neural network with hidden layers, and it is trained using a backpropagation algorithm. GP is an evolutionary algorithm, that evolves a population of candidate solutions using genetic operators, such as mutation and crossover. Therefore, these algorithms have different ways of handling the input parameters and constructing models.

Another explanation for the variation in importance ranking, could be the nature of the input parameters themselves. Some parameters may have stronger correlations with the target variable, which is CBR, in certain datasets or applications, while others may be less relevant. Moreover, different input parameters may interact with each other in complex ways, making it challenging to determine their individual contributions to the model's performance.

The California bearing ratio (CBR) is a measure of the soil's load-bearing capacity and is widely used in geotechnical engineering to assess the suitability of soils for construction. The CBR test involves measuring the resistance of a soil sample to penetration by a standard plunger, under controlled conditions of moisture and compaction. The stiffness and load-bearing capacity of a soil depend on various factors, including its texture, structure, moisture content, density, and compaction effort. The liquid limit of soil is an important input parameter for predicting CBR, because it reflects the soil's ability to resist deformation and support loads. A soil with a high liquid limit is more plastic and less stable, which results in lower CBR values. In contrast, a soil with a low liquid limit is more rigid and stable, leading to higher CBR values. The number of compaction blows is another critical input parameter for predicting CBR, because it represents the compaction effort applied to the soil during construction. The more compaction blows applied, the higher the soil density and stiffness, which result in higher CBR values. Conversely, inadequate compaction effort leads to low soil density and stiffness, resulting in lower CBR values.

Regarding the maximum dry density, it is an indicator of the soil's compactability and weight. However, it does not directly relate to the soil's stiffness and load-bearing capacity, which are the primary factors affecting CBR. Therefore, while it may affect CBR values to some extent, its influence is relatively weak compared to other parameters, such as liquid limit and compaction effort.

5.3. Sustainability

The performed study is highly significant in terms of sustainability. The use of alum sludge as a soil stabilizer can provide an alternative to traditional stabilizers, which can be expensive and have negative environmental impacts. By utilizing alum sludge, this study offers a sustainable solution for improving soil stability in various engineering applications. Additionally, the application of artificial intelligence (AI) methods to predict the California bearing ratio (CBR) of soils stabilized with alum sludge, provides a more efficient and accurate way to determine the effectiveness of this stabilizer. This can reduce the need for extensive and costly field testing, leading to a more sustainable and cost-effective approach to engineering projects. The results of this study also provide valuable insights into the sensitivity and importance of input parameters, which can guide future research and development in the field of soil stabilization.

5.4. Limitations and Scope for Future Works

Although this study provides valuable insights into the use of AI methods for predicting the CBR of soil–alum sludge mixtures, there are limitations that need to be addressed in future work. Firstly, the study only considered nine input parameters, and there may be other important parameters that were not included. Therefore, future research could explore the effect of additional parameters, such as the pH value of the sludge and the particle size distribution of the soil, on the CBR prediction accuracy.

Secondly, the study only used a limited number of CBR test results, which may not be representative of all possible soil–alum sludge mixtures. Thus, future work could expand the database to include more CBR test results, from a wider range of soil and sludge combinations, to further evaluate the accuracy and robustness of the AI models.

Thirdly, while the study evaluated the sensitivity and importance of the input parameters, it did not consider the interaction effects between the parameters. Therefore, future work could investigate the interaction effects, to provide a more comprehensive understanding of the CBR prediction process.

Fourthly, while the AI models used in this study can accurately predict CBR values, they do not provide physical insights into the underlying mechanisms. Understanding the physical processes that govern the strength and stability of soil–alum sludge mixtures, could help to improve the accuracy and reliability of the AI models. Therefore, future work could combine AI methods with physical modeling approaches, to gain a deeper understanding of the system's behavior.

Finally, the study focused on predicting the CBR of soil–alum sludge mixtures for geotechnical engineering applications. However, the use of alum sludge has potential applications in other fields, such as agriculture and water treatment. Thus, future work could explore the use of AI methods to predict other parameters relevant to these fields, such as soil pH and nutrient availability.

6. Conclusions

The use of alum sludge in geotechnical engineering, has gained importance due to its cost-effectiveness and environmental benefits, as it has been found to improve the strength and stability of soils. The CBR test is a crucial parameter in geotechnical engineering, which measures the load-bearing capacity of soil. However, predicting the CBR of soil–alum sludge mixtures can be challenging, due to the large number of input variables. This highlights the significance of utilizing artificial intelligence (AI) methods to overcome this challenge, as this approach has not been widely used previously in this field.

This study found that artificial intelligence (AI) methods can be utilized to predict the CBR, which is a crucial parameter in geotechnical engineering, that measures the load-bearing capacity of soil, of soil–alum sludge mixtures. The use of AI methods can overcome the challenge of predicting CBR, due to the large number of input variables. The study compared the performance of three different AI models (two black-box models and one grey-box model), using a database of 27 CBR test results, and found that the artificial neural network (ANN) and genetic programming (GP) methods outperformed the support vector machines (SVM) method. The database consisted of nine parameters, including liquid limit of soil, plasticity index of soil, specific gravity of soil, number of compaction blows, optimum moisture content of soil, maximum dry density of soil, optimum moisture content of mixture, maximum dry density of mixture, and sludge content.

The ANN method achieved high accuracy, with R^2 of 0.989 and 0.980 for the training and test databases, respectively. The GP method had slightly lower accuracy, but had the advantage of producing an equation as output, that can be used for other datasets. The study also determined the optimal number of neurons for the ANN model, which is 10 neurons. The examination of parameter sensitivity and importance, indicated that the number of compaction hammer blows, and the soil's liquid limit were the most significant parameters, while the maximum dry density parameters for soil and mixture were the least significant. This information is important for engineers and researchers, to optimize their soil stabilization process. Overall, the study provides valuable insights into the performance of different AI models for predicting CBR and can be useful for future research and practical applications.

Author Contributions: Conceptualization, M.D.N.; Methodology, A.B.; Software, A.B.; Validation, A.B.; Writing—original draft, A.B.; Writing—review & editing, M.D.N., A.A., N.M., T.B. and H.A.-N.; Supervision, H.A.-N. All authors have read and agreed to the published version of the manuscript.

References

1. Nguyen, M.D.; Adhikari, S.; Mallya, D.S.; Thomas, M.; Surapaneni, A.; Moon, E.M.; Milne, N.A. Reuse of alumin-ium-based water treatment sludge for phosphorus adsorption: Evaluating the factors affecting and correlation between ad-sorption and sludge properties. *Environ. Technol. Innov.* **2022**, *27*, 102717. [CrossRef]
2. Nguyen, M.D.; Thomas, M.; Surapaneni, A.; Moon, E.M.; Milne, N.A. Beneficial reuse of water treatment sludge in the context of circular economy. *Environ. Technol. Innov.* **2022**, *28*, 102651. [CrossRef]
3. Fiore, F.A.; Rodgher, S.; Ito, C.Y.K.; Bardini, V.S.D.S.; Klinsky, L.M.G. Water sludge reuse as a geotechnical component in road construction: Experimental study. *Clean. Eng. Technol.* **2022**, *9*, 100512. [CrossRef]
4. Nguyen, M.D.; Baghbani, A.; Alnedawi, A.; Ullah, S.; Kafle, B.; Thomas, M.; Moon, E.M.; Milne, N.A. Experimental Study on the Suitability of Aluminium-Based Water Treatment Sludge as a Next Generation Sustainable Soil Replacement for Road Construction. 2023. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331275 (accessed on 20 January 2023).
5. Shaygan, M.; Usher, B.; Baumgartl, T. Modelling Hydrological Performance of a Bauxite Residue Profile for Deposition Manage-ment of a Storage Facility. *Water* **2020**, *12*, 1988. [CrossRef]
6. Osman, K.T. *Soil Degradation, Conservation and Remediation*; Springer: Dordrecht, The Netherlands, 2014.
7. Babatunde, A.; Zhao, Y. A novel alum-sludge based Constructed Wetland system to reduce pollution effects of agricultural run-off: First results. *Int. J. Water* **2007**, *3*, 207. [CrossRef]
8. Dassanayake, K.; Jayasinghe, G.; Surapaneni, A.; Hetherington, C. A review on alum sludge reuse with special reference to agricultural applications and future challenges. *Waste Manag.* **2015**, *38*, 321–335. [CrossRef]
9. Lucas, J.B.; Dillaha, T.A.; Reneau, R.B.; Novak, J.T.; Knocke, W.R. Alum sludge land application and its effect on plant growth. *J. Am. Water Work. Assoc.* **1994**, *86*, 75–83. [CrossRef]
10. Luo, L.; Liu, Y.; Zhuge, Y.; Chow, C.W.; Clos, I.; Rameezdeen, R. A multi-objective optimization approach for supply chain design of alum sludge-derived supplementary cementitious material. *Case Stud. Constr. Mater.* **2022**, *17*, e01156. [CrossRef]
11. Zhao, Y.Q.; Babatunde, A.O.; Hu, Y.S.; Kumar, J.L.; Zhao, X.H. Pilot field-scale demonstration of a novel alum sludge-based con-structed wetland system for enhanced wastewater treatment. *Process Biochem.* **2011**, *46*, 278–283. [CrossRef]
12. Odimegwu, T.C.; Zakaria, I.; Abood, M.M.; Nketsiah, C.B.K.; Ahmad, M. Review on Different Beneficial Ways of Applying Alum Sludge in a Sustainable Disposal Manner. *Civ. Eng. J.* **2018**, *4*, 2230. [CrossRef]
13. James, J.; Pandian, P.K. Industrial Wastes as Auxiliary Additives to Cement/Lime Stabilization of Soils. *Adv. Civ. Eng.* **2016**, *2016*, 1267391. [CrossRef]
14. Gonzalez, J.; Sargent, P.; Ennis, C. Sewage treatment sludge biochar activated blast furnace slag as a low carbon binder for soft soil stabilisation. *J. Clean. Prod.* **2021**, *311*, 127553. [CrossRef]
15. 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**, *9*, 749. [CrossRef]
16. Tony, M.A. Valorization of undervalued aluminum-based waterworks sludge waste for the science of "The 5 Rs' criteria". *Appl. Water Sci.* **2022**, *12*, 20. [CrossRef]
17. Mohamad, N.; Muthusamy, K.; Embong, R.; Kusbiantoro, A.; Hashim, M.H. Environmental impact of cement production and Solutions: A review. *Mater. Today Proc.* **2021**, *48*, 741–746. [CrossRef]
18. Jiu-Yu, L.I.; Ning, W.A.N.G.; Ren-Kou, X.U.; Tiwari, D. Potential of industrial byproducts in ameliorating acidity and alumi-num toxicity of soils under tea plantation. *Pedosphere* **2010**, *20*, 645–654.
19. Shetty, R.; Vidya, C.S.-N.; Prakash, N.B.; Lux, A.; Vaculík, M. Aluminum toxicity in plants and its possible mitigation in acid soils by biochar: A review. *Sci. Total Environ.* **2020**, *765*, 142744. [CrossRef]
20. Luo, H.L.; Hsiao, D.H.; Lin, D.F.; Lin, C.K. Cohesive soil stabilized using sewage sludge ash/cement and nano aluminum ox-ide. *Int. J. Transp. Sci. Technol.* **2012**, *1*, 83–99. [CrossRef]
21. Baghbani, A.; Baumgartl, T.; Filipovic, V. Effects of Wetting and Drying Cycles on Strength of Latrobe Valley Brown Coal (No. EGU23-4804). 2023, Copernicus Meetings. Available online: https://meetingorganizer.copernicus.org/EGU23/EGU23-4804.html (accessed on 22 February 2023).
22. Batley, G.E.; Kirby, J.K.; McLaughlin, M.J. Fate and Risks of Nanomaterials in Aquatic and Terrestrial Environments. *Accounts Chem. Res.* **2012**, *46*, 854–862. [CrossRef]
23. Huang, J.; Kogbara, R.B.; Hariharan, N.; Masad, E.A.; Little, D.N. A state-of-the-art review of polymers used in soil stabiliza-tion. *Constr. Build. Mater.* **2021**, *305*, 124685. [CrossRef]
24. Jadhav, P.; Sakpal, S.; Khedekar, H.; Pawar, P.; Malipatil, M. Experimental Investigation of Soil Stabilization by Using Alum Sludge. *Int. J. Res. Appl. Sci. Eng. Technol.* **2022**, *10*, 598–602. [CrossRef]
25. Omar, M.B.H.C.; Che Mamat, R.; Abdul Rasam, A.R.; Ramli, A.; Samad, A. Artificial intelligence application for predict-ing slope stability on soft ground: A comparative study. *Int. J. Adv. Technol. Eng. Explor.* **2021**, *8*, 362–370. [CrossRef]

26. Suman, S.; Khan, S.Z.; Das, S.K.; Chand, S.K. Slope stability analysis using artificial intelligence techniques. *Nat. Hazards* **2016**, *84*, 727–748. [CrossRef]

27. Baghbani, A.; Daghistani, F.; Naga, H.A.; 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, 14–16 December 2022.

28. Lin, S.S.; Shen, S.L.; Zhang, N.; Zhou, A. Modelling the performance of EPB shield tunnelling using machine and deep learn-ing algorithms. *Geosci. Front.* **2021**, *12*, 101177. [CrossRef]

29. Baghbani, A.; Baghbani, H.; Shalchiyan, M.M.; Kiany, K. Utilizing artificial intelligence and finite element method to simulate the effects of new tunnels on existing tunnel deformation. *J. Comput. Cogn. Eng.* **2022**. [CrossRef]

30. Ayawah, P.E.; Sebbeh-Newton, S.; Azure, J.W.; Kaba, A.G.; Anani, A.; Bansah, S.; Zabidi, H. A review and case study of Ar-tificial intelligence and Machine learning methods used for ground condition prediction ahead of tunnel boring Machines. *Tunn. Undergr. Space Technol.* **2023**, *125*, 104497. [CrossRef]

31. Nguyen, M.D.; Baghbani, A.; Alnedawi, A.; Ullah, S.; Kafle, B.; Thomas, M.; Moon, E.M.; Milne, N.A. Investigation on the suitability of aluminium-based water treatment sludge as a sustainable soil replacement for road construction. *Transp. Eng.* **2023**, 100175. [CrossRef]

32. Anysz, H.; Foremny, A.; Kulejewski, J. February. Comparison of ANN classifier to the neuro-fuzzy system for collusion detec-tion in the tender procedures of road construction sector. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2019; Volume 471, p. 112064.

33. Baghbani, A.; Baghbani, H. Improving Travel Time Relative Formula Using Calories Burned in Activities. In Proceedings of the 2nd International Congress on Structure, Architecture and Urban Development, Tabriz, Iran, 16–18 December 2014.

34. Baghbani, A.; Costa, S.; Choundhury, T.; Faradonbeh, R.S. Prediction of Parallel Desiccation Cracks of Clays Using a Classi-fication and Regression Tree (CART) Technique. In Proceedings of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR), Newcastle, Australia, 14–16 December 2022.

35. Onyelowe, K.C.; Aneke, F.I.; Onyia, M.E.; Ebid, A.M.; Usungedo, T. AI (ANN, GP, and EPR)-based predictive models of bulk density, linear-volumetric shrinkage & desiccation cracking of HSDA-treated black cotton soil for sustainable subgrade. *Géoméch. Geoengin.* **2022**, 1–20. [CrossRef]

36. Baghbani, A.; Abuel-Naga, H.; Shirani Faradonbeh, R.; Costa, S.; Almasoudi, R. Ultrasonic Characterization of Compacted Salty Kaolin–Sand Mixtures Under Nearly Zero Vertical Stress Using Experimental Study and Machine Learning. *Geotech. Geol. Eng.* **2023**. [CrossRef]

37. Baghbani, A.; Costa, S.; Faradonbeh, R.S.; Soltani, A.; Baghbani, H. Modeling the Effects of Particle Shape on Damping Ratio of Dry Sand by Simple Shear Testing and Artificial Intelligence. *Appl. Sci.* **2023**, *13*, 4363. [CrossRef]

38. Njock, P.G.A.; Shen, S.-L.; Zhou, A.; Lyu, H.-M. Evaluation of soil liquefaction using AI technology incorporating a coupled ENN/t-SNE model. *Soil Dyn. Earthq. Eng.* **2019**, *130*, 105988. [CrossRef]

39. 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 Dyn. Earthq. Eng.* **2023**, *165*, 107708. [CrossRef]

40. Baghbani, A.; Costa, S.; O'Kelly, B.C.; Soltani, A.; Barzegar, M. Experimental study on cyclic simple shear behaviour of pre-dominantly dilative silica sand. *Int. J. Geotech. Eng.* **2022**, *17*, 91–105. [CrossRef]

41. Baghbani, A.; Costa, S.; Faradonbeh, R.S.; Soltani, A.; Baghbani, H. Experimental-AI Investigation of the Effect of Particle Shape on the Damping Ratio of Dry Sand under Simple Shear Test Loading. 2023; *Preprint*. [CrossRef]

42. Baghbani, A.; Daghistani, F.; Baghbani, H.; Kiany, K. *Predicting the Strength of Recycled Glass Powder-Based Geopolymers for Improving Mechanical Behavior of Clay Soils Using Artificial Intelligence (No. 9741)*; EasyChair: Manchester, UK, 2023.

43. Baghbani, A.; Costa, S.; Choudhury, T. Developing Mathematical Models for Predicting Cracks and Shrinkage Intensity Factor During Clay Soil Desiccation. 2023. [CrossRef]

44. Baghbani, A.; Daghistani, F.; Kiany, K.; Shalchiyan, M.M. *AI-Based Prediction of Strength and Tensile Properties of Expansive Soil Stabilized with Recycled Ash and Natural Fibers (No. 9743)*; EasyChair: Manchester, UK, 2023.

45. Baghbani, A.; Choudhury, T.; Costa, S.; Reiner, J. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Sci. Rev.* **2022**, *228*, 103991. [CrossRef]

46. Aamir, M.; Mahmood, Z.; Nisar, A.; Farid, A.; Ahmed Khan, T.; Abbas, M.; Ismaeel, M.; Shah, S.A.R.; Waseem, M. Perfor-mance evaluation of sustainable soil stabilization process using waste materials. *Processes* **2019**, *7*, 378. [CrossRef]

47. Shah, S.A.R.; Mahmood, Z.; Nisar, A.; Aamir, M.; Farid, A.; Waseem, M. Compaction performance analysis of alum sludge waste modified soil. *Constr. Build. Mater.* **2020**, *230*, 116953. [CrossRef]

48. Taffese, W.Z.; Sistonen, E. Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions. *Autom. Constr.* **2017**, *77*, 1–14. [CrossRef]

49. Flah, M.; Nunez, I.; Ben Chaabene, W.; Nehdi, M.L. Machine learning algorithms in civil structural health monitoring: A systemat-ic review. *Arch. Comput. Methods Eng.* **2021**, *28*, 2621–2643. [CrossRef]

50. Prakash, G.; Narasimhan, S.; Al-Hammoud, R. A two-phase model to predict the remaining useful life of corroded reinforced concrete beams. *J. Civ. Struct. Health Monit.* **2019**, *9*, 183–199. [CrossRef]

51. Farooq, M.; Zheng, H.; Nagabhushana, A.; Roy, S.; Burkett, S.; Barkey, M.; Kotru, S.; Sazonov, E. Damage detection and identification in smart structures using SVM and ANN. In *Smart Sensor Phenomena, Technology, Networks, and Systems Integration*; SPIE: Bellingham, WA, USA, 2012; Volume 8346, p. 83461O. [CrossRef]

52. Huo, L.S.; Li, X.; Yang, Y.B.; Li, H.N. Damage detection of structures for ambient loading based on cross correlation function am-plitude and SVM. *Shock. Vib.* **2016**, *2016*, 3989743.

53. Karballaeezadeh, N.; Mohammadzadeh, S.D.; Shamshirband, S.; Hajikhodaverdikhan, P.; Mosavi, A.; Chau, K.W. Prediction of re-maining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road). *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 188–198.

54. Wu, L.; Fu, X.; Guan, Y. Review of the remaining useful life prognostics of vehicle lithium-ion batteries using data-driven meth-odologies. *Appl. Sci.* **2016**, *6*, 166. [CrossRef]

55. Upadhya, A.; Thakur, M.S.; Sihag, P.; Kumar, R.; Kumar, S.; Afeeza, A.; Afzal, A.; Saleel, C.A. Modelling and prediction of binder con-tent using latest intelligent machine learning algorithms in carbon fiber reinforced asphalt concrete. *Alex. Eng. J.* **2023**, *65*, 131–149. [CrossRef]

56. Fallah-Mehdipour, E.; Haddad, O.B.; Mariño, M.A. Extraction of Optimal Operation Rules in an Aquifer-Dam System: Genetic Programming Approach. *J. Irrig. Drain. Eng.* **2013**, *139*, 872–879. [CrossRef]

57. Tayfur, G. Modern Optimization Methods in Water Resources Planning, Engineering and Management. *Water Resour. Manag.* **2017**, *31*, 3205–3233. [CrossRef]

58. Mohammad-Azari, S.; Bozorg-Haddad, O.; Loáiciga, H.A. State-of-art of genetic programming applications in water-resources systems analysis. *Environ. Monit. Assess.* **2020**, *192*, 73. [CrossRef]

59. Fenton, M.; McNally, C.; Byrne, J.; Hemberg, E.; McDermott, J.; O'Neill, M. Automatic innovative truss design using grammatical evolution. *Autom. Constr.* **2014**, *39*, 59–69. [CrossRef]

60. Baldock, R.; Shea, K. Structural Topology Optimization of Braced Steel Frameworks Using Genetic Programming. In Proceedings of the Intelligent Computing in Engineering and Architecture: 13th EG-ICE Workshop 2006, Ascona, Switzerland, 25–30 June 2006; pp. 54–61. [CrossRef]

61. Hasançebi, O. Optimization of truss bridges within a specified design domain using evolution strategies. *Eng. Optim.* **2007**, *39*, 737–756. [CrossRef]

62. Azimi, M.; Eslamlou, A.; Pekcan, G. Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review. *Sensors* **2020**, *20*, 2778. [CrossRef] [PubMed]

63. Giustolisi, O.; Doglioni, A.; Savic, D.; Webb, B. A multi-model approach to analysis of environmental phenomena. *Environ. Model. Softw.* **2007**, *22*, 674–682. [CrossRef]

64. Zhang, Q.; Barri, K.; Jiao, P.; Salehi, H.; Alavi, A.H. Genetic programming in civil engineering: Advent, applications and future trends. *Artif. Intell. Rev.* **2020**, *54*, 1863–1885. [CrossRef]

65. Deiss, L.; Margenot, A.J.; Culman, S.W.; Demyan, M.S. Tuning support vector machines regression models improves prediction ac-curacy of soil properties in MIR spectroscopy. *Geoderma* **2020**, *365*, 114227. [CrossRef]

66. Abd, A.M.; Abd, S.M. Modelling the strength of lightweight foamed concrete using support vector machine (SVM). *Case Stud. Constr. Mater.* **2017**, *6*, 8–15. [CrossRef]

67. Kourehli, S.S. Prediction of unmeasured mode shapes and structural damage detection using least squares support vector ma-chine. *Struct. Monit. Maint.* **2018**, *5*, 379–390.

68. Ly, H.B.; Pham, B.T. Prediction of shear strength of soil using direct shear test and support vector machine model. *Open Constr. Build. Technol. J.* **2020**, *14*, 268–277. [CrossRef]

69. Ghiasi, R.; Noori, M.; Altabey, W.; Silik, A.; Wang, T.; Wu, Z. Uncertainty Handling in Structural Damage Detection via Non-Probabilistic Meta-Models and Interval Mathematics, a Data-Analytics Approach. *Appl. Sci.* **2021**, *11*, 770. [CrossRef]

70. Zhou, Z.; Zhang, J.; Liu, P.; Li, Z.; Georgiadis, M.C.; Pistikopoulos, E.N. A two-stage stochastic programming model for the optimal design of distributed energy systems. *Appl. Energy* **2012**, *103*, 135–144. [CrossRef]

71. Beghini, L.L.; Beghini, A.; Katz, N.; Baker, W.F.; Paulino, G.H. Connecting architecture and engineering through structural topology optimization. *Eng. Struct.* **2014**, *59*, 716–726. [CrossRef]

72. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [CrossRef] [PubMed]

73. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef]

74. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]

75. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

76. Shanmuganathan, S. *Artificial Neural Network Modelling: An Introduction*; Springer International Publishing: New York, NY, USA, 2016; pp. 1–14. [CrossRef]

77. Dongare, A.D.; Kharde, R.R.; Kachare, A.D. Introduction to artificial neural network. *Int. J. Eng. Innov. Technol.* **2012**, *2*, 189–194.

78. Macukow, B. Neural networks–state of art, brief history, basic models and architecture. In Proceedings of the Computer Information Systems and Industrial Management: 15th IFIP TC8 International Conference, CISIM 2016, Vilnius, Lithuania, 14–16 September 2016; Springer International Publishing: New York, NY, USA, 2016; pp. 3–14.

79. Sun, Y.; Zeng, W.; Zhao, Y.; Qi, Y.; Ma, X.; Han, Y. Development of constitutive relationship model of Ti600 alloy using artificial neural network. *Comput. Mater. Sci.* **2010**, *48*, 686–691. [CrossRef]

80. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [CrossRef]

81. Marquardt, D.W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [CrossRef]

82. Burden, F.; Winkler, D. Bayesian regularization of neural networks. *Artif. Neural Netw. Methods Appl.* **2009**, 23–42. [CrossRef]