LJMU Research Online

Trong, DK, Pham, BT, Jalal, FE, Iqbal, M, Roussis, PC, Mamou, A, Ferentinou, M, Vu, DQ, Duc Dam, N, Tran, QA and Asteris, PG

 On Random Subspace Optimization-Based Hybrid Computing Models Predicting the California Bearing Ratio of Soils

http://researchonline.ljmu.ac.uk/id/eprint/15784/

Article

Article

On Random Subspace Optimization-Based Hybrid Computing Models Predicting the California Bearing Ratio of Soils

Duong Kien Trong [1], Binh Thai Pham [1,*], Fazal E. Jalal [2], Mudassir Iqbal [2,3], Panayiotis C. Roussis [4], Anna Mamou [5], Maria Ferentinou [6], Dung Quang Vu [1], Nguyen Duc Dam [1], Quoc Anh Tran [7] and Panagiotis G. Asteris [5,*]

[1] Faculty of Civil Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi 100000, Vietnam; duongtk@utt.edu.vn (D.K.T.); dungvq@utt.edu.vn (D.Q.V.); damnd@utt.edu.vn (N.D.D.)

[2] State Key Laboratory of Ocean Engineering, Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; jalal2412@sjtu.edu.cn (F.E.J.); mudassiriqbal29@sjtu.edu.cn (M.I.)

[3] Department of Civil Engineering, University of Engineering and Technology, Peshawar 25000, Pakistan

[4] Department of Civil and Environmental Engineering, University of Cyprus, Nicosia 1678, Cyprus; roussis@ucy.ac.cy

[5] Computational Mechanics Laboratory, School of Pedagogical and Technological Education, Heraklion, 14121 Athens, Greece; a.p.mamou@gmail.com

[6] School of Civil Engineering and Built Environment, Liverpool John Moores University, Liverpool L3 3AF, UK; M.Ferentinou@ljmu.ac.uk

[7] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway; quoc.a.tran@ntnu.no

* Correspondence: binhpt@utt.edu.vn (B.T.P.); panagiotisasteris@gmail.com (P.G.A.)

Abstract: The California Bearing Ratio (CBR) is an important index for evaluating the bearing capacity of pavement subgrade materials. In this research, random subspace optimization-based hybrid computing models were trained and developed for the prediction of the CBR of soil. Three models were developed, namely reduced error pruning trees (REPTs), random subsurface-based REPT (RSS-REPT), and RSS-based extra tree (RSS-ET). An experimental database was compiled from a total of 214 soil samples, which were classified according to AASHTO M 145, and included 26 samples of A-2-6 (clayey gravel and sand soil), 3 samples of A-4 (silty soil), 89 samples of A-6 (clayey soil), and 96 samples of A-7-6 (clayey soil). All CBR tests were performed in soaked conditions. The input parameters of the models included the particle size distribution, gravel content (G), coarse sand content (CS), fine sand content (FS), silt clay content (SC), organic content (O), liquid limit (LL), plastic limit (PL), plasticity index (PI), optimum moisture content (OMC), and maximum dry density (MDD). The accuracy of the developed models was assessed using numerous performance indexes, such as the coefficient of determination, relative error, MAE, and RMSE. The results show that the highest prediction accuracy was obtained using the RSS-based extra tree optimization technique.

Keywords: California Bearing Ratio; modulus of subgrade reaction; elastic modulus; metaheuristic algorithms

1. Introduction

Accurate prediction of the mechanical index of geomaterials is critical for robust pavement design [1,2]. The strength of the subgrade soil is routinely assessed in terms of its California Bearing Ratio (CBR). The California Bearing Ratio (CBR) of soil is a static strength and bearing capacity index, which may be obtained from either laboratory or in situ measurements [3,4]. The CBR is an important input parameter predicting the stiffness modulus of the soil subgrade, which is a key pavement design index considering the effect of cyclic loading on the soil's stiffness [5–7]. The CBR value is used to indirectly estimate the thickness of the subgrade materials in major infrastructure projects. Therefore, fast

and reliable estimation of this parameter is significant to the design process and relevant construction time.

The CBR test was originally introduced by the California Highway Department, during World War-II, and was subsequently adopted widely as a standard method for soil strength and bearing capacity evaluation [8], obtained using either ASTM Standard D-1883-05 or BS 1377 [9]. Laboratory tests are performed on compacted soil samples with OMC in un-soaked and soaked conditions, and they can also be carried out on natural soils. According to current AASHTO 2003 standards, the laboratory CBR test involves soil mass penetration using a circular 50mm diameter plunger applied at a 1.25 mm/min rate [10] into a compacted soil specimen at the optimum moisture content. The in situ CBR tests are conducted at the natural ground surface level, prepared subgrade level, or on a level surface of the test pit at the construction site. The applied plunger pressure is then divided by the pressure that is required to incur the same penetration of standard crushed rock [11]. The CBR value of soil is influenced by a number of parameters, namely the particle size, soil fabric, plasticity index, moisture content, suction, and dry density [11,12].

Direct laboratory or in situ measurement of the CBR index of soil is a lengthy process, often yielding inaccurate results due to the disturbance of samples, negligence during testing, and poor testing facilities. In this context, developing artificial intelligence (AI) models for the prediction of CBR may be a viable alternative [1], if CBR could be obtained through a cost-effective method resulting in less construction time. Artificial intelligence models can simulate highly non-linear associations between numerous input and output parameters and can therefore provide more accurate predictions than those obtained using simple and multiple regression analysis [13–15]. During the last decade numerous artificial intelligence models techniques, including artificial neural network (ANN) [16], adaptive neuro-fuzzy inference system (ANFIS) [17], gene and multi expression programming [18–20], ensemble framework techniques (for instance, bagging, rotation forest, and random subspace (RSS)) [21], decision tree (DT) [22], and support vector machine (SVM) [23], have been used in engineering and numerous other disciplines [13,24]. Hybrid ensemble strategies including bagging, RSS, and boosting with ensemble pruning are particularly suitable for extracting deep features from multivariate data [25].

The aim of this research was to develop subspace optimization-based hybrid computing models for the prediction of CBR using 10 input variables: gravel percentage (G), coarse sand % (CS), fine sand % (FS), fine material (silt and clay %—passing sieve No. 200) (SC), organic matter content (O), liquid limit (LL), plastic limit (PL), plasticity index (PI), optimum moisture content (OMC), and maximum dry density (MDD). To this end, three hybrid ensemble models were developed, i.e., REPT (reduced error pruning trees), RSS-REPT, and RSS-ET (RSS-extra trees). The accuracy of the models was assessed using numerous performance indexes, such as the root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R^2).

2. Short Literature Review on Soft Computing Techniques for Estimation of the California Bearing Ratio

The story of predictive models based on available test data to forecast the CBR value can be traced back to Black (1962) [26], who developed a correlation scheme for CBR estimation of cohesive soils based on index properties, such as the plasticity index (PI) and liquidity index (LI). Johnson and Bhatia in (1969) [27] suggested a correlation for CBR based on the suitability index, which relies on plasticity and particle size distribution data from lateritic gravel soils in Ghana. Agarwal and Ghanekar in (1970) [28], based on experiments on fine-graded soils, proposed a correlation between CBR and Atterberg limits. Yet, the authors suggested that poor correlations were revealed among the input parameters, while only OMC and LL were found to be significant; therefore, the applicability of the method was limited to preliminary soil identifications.

Later, Al-Refeai and Al-Suhaibani (1997) [29], relying on data obtained from different soils ranging from clay to gravely sand, from dynamic cone penetrometer tests, suggested

a relationship between the moisture content, dry density, and penetration depth with CBR, and performed tests for a range of moisture content and density conditions.

Stephens in (1992) [30], based on series of data from the Natal Roads Department in Pietermaritzburg, South Africa, reviewed the performances of earlier published models and commented that most of them were unsatisfactory for universal use, particularly for problematic soils like expansive and shrinking soils, widely exposed in South Africa. One of the key issues identified was the impact of the clay fraction determination on CBR value estimation. The author proposed the use of shrinkage properties and gradation to allow for CBR estimation for shrinking and non-shrinking soils.

In the USA, under the National Cooperative Highway Research Program (NCHRP) (2004) [31], two empirical predictive models were suggested, using the index properties of soil for clean (i.e., fine content $\geq$ 12%) and coarse-grained soils. Thereafter, Kin in (2006) [32] reviewed the correlation equations given by NCHRP and found limitations in CBR estimation for coarse-grained soils, whereas for fine-grained soils, the NCHRP models were found to be of moderate accuracy.

During the last decade, many authors used statistics, proposing simple and multiple regression predictive models for CBR value estimation, based on gradation analysis, Atterberg limits, optimum moisture content, and maximum dry density among others [1,3,33–38].

A close review of the literature presented from Black (1962) [26] of the latest studies suggests that the prediction of CBR values from presented and reviewed statistical models is generally problematic, has poor applicability and potential for universal models, and is likely to be limited to the local datasets. This is partly attributed to the small number of observations available used in the predictive models, the complexity and non-linearity of the problem itself, the uncertainty involved in the determination of soil properties, the scatter of data, and the particulate nature of soil, which is seldom elastic, isotropic, or homogeneous.

Machine learning (ML) is an area of computational mechanics that sufficiently handles complex problems, exposing non-linear characteristics, including a high level of uncertainty as evidenced by recently published studies in the broader spectrum of geotechnical engineering. Advanced ML techniques, which are more competent in non-linear modelling, provide a feasible tool for simulating several multifaceted processes [39].

Several AI-based models have been used in the last decade to predict the CBR value of soils. Various biologically inspired algorithms, such as artificial neural network (ANN), support vector machines (SVMs), gene expression programming (GEP), generalized regression neural networks (GRNNs), multi-layer perceptron neural networks (MLPNs), and group method of data handling (GMDH), have been exploited and produced promising results, as shown in Table 1.

Taskiran (2010) [40] proposed an ANN and a GEP model to forecast the CBR value of fine-grained soils. The compiled data set included 151 CBR test data for soils classified as A-4 to A-7. The performance of the models was found to range ($R^2 > 0.90$) for both the models.

Yildirim and Gunaydin (2011) [41], Kumar et al. (2013) [42], Varghese et al. (2013) [43], Bhatt and Jain (2014) [34], Sabat (2015) [44], Ghorbani and Hasanzadehshooiili (2018) [45], Suthar and Aggarwal (2018) [33], Alam et al. (2020) [46], and Islam and Roy (2020) [47] developed further models relying on limited data sets (in the range of 20 to 158 observations). Their findings suggested accuracies of 0.81 $R^2 < 1.00$. Later SVM models developed by Sabat (2015) [44], using 49 CBR test data of stabilized soils, were found to show a performance of $R^2 = 0.96$. Only recently, Taha et el. (2019) [48] used larger data sets of 218 laboratory tests and developed an ANN-based model with a prediction accuracy of $R^2 = 0.88$. In addition, Tenpe and Patel (2020) [49] used 389 soil test data and produced two models using SVM and GEP algorithms, with a performance accuracy ranging between 0.83 < R^2 < 0.90. Al-Busultan et al. (2020) [50] used a dataset of 358 tests and developed an ANN model with an $R^2 = 0.78$.

The prediction capability of the presented models, as was quantified through R^2 metrics, suggests the paradox of higher predictive accuracies based on predictive models developed using smaller datasets, compared to predictive models that used larger data sets and yielded comparatively lower R^2 values of moderate accuracy [49,50]. This is most probably a result of overfitting and network memorizing of the particular local dataset, which results in the models being weak in generalization. It is very common for back-propagation-based models to become trapped in local minima, leading to erroneous results [51]. It is important to highlight that studies based on larger training data sets include a more representative description of the specific geotechnical problem, and therefore are expected to be more reliable for future predictions. The reliability of a model depends on the comprehensiveness of the input data set. The incorporation of a wide variety of soils, as per the Unified Soil Classification System, which covers a range of engineering properties that affect the stiffness of a soil, such as soil index properties and particle size distribution, satisfies the criteria for a promising prediction model.

To address these deficiencies, scholars have proposed hybrid models by integrating optimization algorithms (OAs) and common soft computing models to search for the exact global minimum instead of finding the local minima [47,52]. Hybridization of OAs and CSC techniques balances the exploration and exploitation processes and generates optimized values of learning parameters (weights and biases), which in turn are used to enhance the performance of CSC techniques.

Bradhan et al., 2021a [53] and Bradhan et al., 2021b [54] proposed a novel integration of extreme learning machine (ELM) and adaptive neuro swarm intelligence (ANSI) techniques for the determination of the California Bearing Ratio (CBR) of soils and the results were prominent.

Onyelowe et al., 2021 [16] applied evolutionary hybrid algorithms of ANN, Levenberg–Marquardt back-propagation (LMBP), Bayesian programming (BP), and conjugate gradient (CG) algorithms to predict the CBR value of ash-treated expansive soil, and the correlation was found to be $R^2 = 0.9$.

Raza et al., 2021 [55], identified a gap in the literature in the prediction of geosynthetic-reinforced subgrade soil, and used data-driven-based machine learning models to estimate the CBR value. Several intelligent models, such as artificial neural network (ANN), least median of squares regression, Gaussian processes regression, elastic net regularization regression, lazy K-star, M-5 model trees, alternating model trees, and random forest, were proposed and the prediction accuracy was found to be $0.80 < R^2 < 0.92$.

Table 1. Prediction accuracy of the soft computing models predicting the CBR of soil reported in the reviewed literature.

Reference	Model(s) Employed	Prediction Accuracy	Total Nr of Data/Types of Soils
Bardhan et al., 2021a [53]	MARS-L	$R^2 = 0.96$ RMSE = 0.0359	362
Bardhan et al., 2021b [54]	extreme learning machine (ELM)-based models	$0.81 < R^2 < 0.91$	312
Onyelowe et al., 2021 [16]	Levenberg–Marquardt backpropagation (LMBP), Bayesian programming (BP), and conjugate gradient (CG)	$R^2 = 0.90$	129
Raza et al. 2021 [55]	least median of squares regression, Gaussian processes regression, elastic net regularization regression, lazy K-star, M-5 model trees, alternating model trees, and random forest	$0.80 < R^2 < 0.92$	97 tests

In the current study, the proposed hybrid model aimed to develop subspace optimization-based hybrid computing models for the prediction of CBR using 10 input variables: gravel % (G), coarse sand % (CS), fine sand % (FS), fine material (silt and clay %—passing sieve No. 200) (SC), organic matter content (O), liquid limit (LL), plastic limit (PL), plasticity index (PI), optimum moisture content (OMC), and maximum dry density (MDD), using a comprehensive data set of 214 CBR tests of various soil types.

3. Materials and Methods

3.1. Database for the Training of Soft Computing Models

The reliability of the database used to train soft computing models is undoubtedly a critical parameter affecting the reliability of the actual model prediction. Interestingly though, the reliability of the actual database used to train soft computing models has generally received less attention than, for example, the model architecture and the various transfer functions used. A reliable database should not only comprise a statistically significant amount of representative data, but the data distribution should also comply with fundamental statistical analysis principles and the experimental/field data should be reported in compliance with international standards.

In light of the above, an experimental database was compiled from samples collected at the Van Don - Mong Cai expressway project, in the Quang Ninh province of Vietnam. The starting point of the route is located at 70 + 108 km and intersects the endpoint of the main road connecting the Van Don zone in the Doan Ket commune, the Van Don District, and the Quang Ninh province. The endpoint of the route is located at 150 + 338 km and intersects the 335 provincial road, at the starting point of the Bac Luan 2 Bridge Path project in the Hai Hoa ward, Mong Cai city of the Quang Ninh province [56]. A total of 214 samples were collected during the period spanning from November 8, 2019 to July 1, 2021. The soil samples were then transferred to the laboratory, whereupon the particle size distribution (AASHTO T 88 [57] and ASTM D 422 [58]), liquid limit (AASHTO T 89 [59] and ASTM D 4318 [60]), organic content (AASHTO T 267 [61]), compaction curves (ASTM D 4253 [62] and ASTM D 4254 [63] [64]), and the CBR of the soil were determined (Figures 1 and 2). The statistical parameters, such as the minimum, average, maximum, and standard deviations, are presented in detail in Table 2.

Table 2. Statistical analysis of the input and output parameters.

Variable	Symbol	Units	Category	Statistics			
				Min	Average	Max	STD
Gravel content	G	%	Input	0	51.4	24.75	13.295
Coarse Sand content	CS	%	Input	3	46.3	23.7	7.017
Fine Sand content	FS	%	Input	2.5	41.5	7.25	6.468
Silt-Clay content	SC	%	Input	17.87	88.7	44.55	10.447
Organic content	OC	%	Input	0.12	2.94	1.51	0.373
Liquid Limit	LL	%	Input	2.08	48.45	39.99	6.173
Plastic Limit	PL	%	Input	1.17	28.49	20.835	3.068
Plasticity Index	PI	%	Input	0.91	27.48	18.435	4.078
Optimum Moisture Content	OMC	%	Input	9.3	21.5	14.275	2.619
Maximum Dry Density	MDD	g/cm^3	Input	1.672	2.14	1.871	0.118
California Bearing Ratio	CBR	-	Output	3.09	41.26	7.95	8.175

The California Bearing Ratio (CBR) relates the penetration resistance of laboratory-compacted soil material to that of well-graded (poorly sorted), durable, and crushed rock material [65]. The CBR was developed by the American Society for Testing and Materials [66] in North America [65] and the American Association of State Highway and Transportation Officials [67] for assessing the penetration resistance of subbase and subgrade pavement materials. The test involves compaction of the soil in a standard mold size (177.8 mm height and 152.3 mm diameter). The moisture content and compaction energy may vary with individual project's requirements. The load is applied through a 49.6 mm diameter steel piston at a 1.3 mm penetration rate per minute. The load required to incur a 2.54 mm and 5.08 mm penetration is continuously measured and converted to stress by dividing it with the area of the steel piston. The CBR is then calculated as the ratio of the required laboratory stress over the corresponding crushed aggregate standard penetration load [65]. A minimum CBR of 10 is generally required for subgrade design [64].

The CBR of soil is influenced by a number of parameters including, for example, the particle size, soil fabric, plasticity index, moisture content, suction, and dry density [11,12]. In the laboratory, the CBR is determined as the in situ moisture content and corresponding dry density. Whilst the in situ dry density of the soil can be determined with reasonable accuracy, determining the in situ moisture content may be challenging. In general, as the moisture content is reduced and the suction increases, the soil shifts from a bulk-water-regulated to a menisci-water-regulated response and the CBR is reduced significantly at the wet side of the optimum [68,69]. In this research, the CBR at the equilibrium moisture content (4 days soaked CBR) was measured [70].
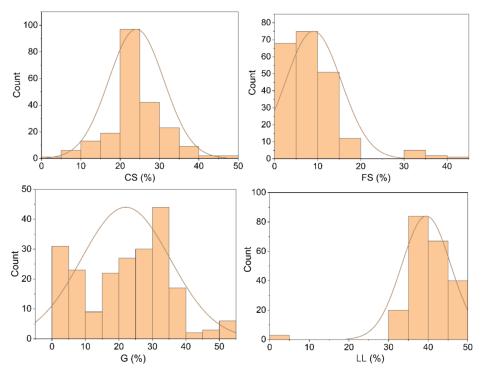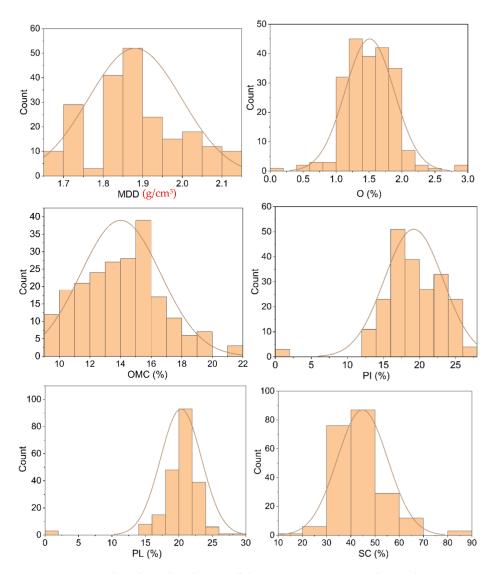

Figure 1. *Cont.*

Figure 1. Statistical analysis distribution of the input parameters in this study.
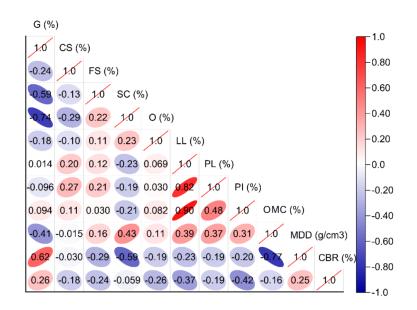

Figure 2. Correlation matrix analysis input variables in this study.

3.2. Sensitivity Analysis of the Input Parameters Predicting the CBR of Soil

A sensitivity analysis was performed to identify which of the 10 input parameters (gravel percentage (G), coarse sand percentage (CS), fine sand percentage (FS), fine soil material (silt clay percentage) (SC), organic matter content (O), liquid limit (LL), plastic limit (PL), plasticity index (PI), optimum moisture content (OMC), and maximum dry density (MDD)) significantly affected the predicted CBR of soil. The aim of the sensitivity analysis was to remove the input parameters that have the smallest influence on the predicted output parameter, thereby significantly reducing the required complexity and training time of the model. In this research, the cosine amplitude method (CAM) was used to perform the sensitivity analysis [71,72]. In CAM, data pairs are used to construct a data array, X, as follows:

$$X = \{x_1, x_2, x_3, \ldots, x_i, \ldots, x_n\} \tag{1}$$

where the x_i variable, in the X array, is a m length vector, which may be expressed as:

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}\} \tag{2}$$

The relationship between R_{ij} (strength of the relation) and the x_i and x_j datasets may be expressed as:

$$R_{ij} = \frac{\sum_{k=1}^{m} x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^{m} x^2_{ik} \sum_{k=1}^{m} x^2_{ik}}} \tag{3}$$

The results of the sensitivity analysis presented in Figure 3 show that the highest and smallest relative strength effect (RSE) on the CBR of soil was obtained for the maximum dry density (MDD)) (RSE = 0.8301) and the fine sand percentage (FS) (RSE = 0.5915), respectively. The other eight input parameters registered moderate RSE values ranging between 0.7548 and 0.7974.

Figure 3. Sensitivity analysis of the input parameters predicting the CBR of soil.

3.3. Methods Used

This section presents the methodology used to train and develop the soft computing models. A database comprising 10 input parameters: gravel percentage (G), coarse sand percentage (CS), fine sand percentage (FS), fine soil material (silt clay percentage) (SC), organic matter content (O), liquid limit (LL), plastic limit (PL), plasticity index (PI), optimum moisture content (OMC), and maximum dry density (MDD), for the prediction of the soil's CBR was compiled (Figure 4). The database was split into training and testing datasets at

a ratio of 70% to 30%. For the modelling of the California Bearing Ratio (CBR), three soft computing models were trained and developed, such as REPT, RSS-REPT, and RSS-ET, and their accuracy was evaluated using a variety of performance indexes, such as RMSE, MAE, and R^2. The RSS-REPT and RSS-ET are hybrid models, which were developed using a combination of RSS ensemble techniques and two predictors, namely REPT and ET. In the hybrid models, RSS ensemble was firstly used to optimize the training dataset, and then the optimal training dataset was used to train the predictors (REPT and ET). The Weka software was used as a platform for training and validating the models. Detailed and in-depth background theory of the methods used herein is presented in the following sections.
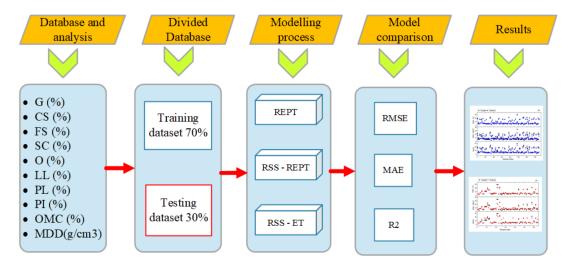

Figure 4. Methodological framework for the predicted values in this study.

3.3.1. Random Subspace (RSS)

The random subspace method (RSS) is a random sampling ensemble method used to produce different representations that could be employed in generating a variety of decision agents [73,74]. A typical RS model comprises an integrated algorithm that establishes a DT based on a classifier supporting the maximum accuracy in the case of training data. This method is used to improve the performance of weak classifiers [75]. Thereafter, the RSS incorporates randomness inside the problem representation by randomly choosing specific variables that are to be replaced [74]. According to Plumpton et al., [76], the RS approach is an efficacious ensemble and it exhibits many diverse classifiers since it combines the accuracy of the weak classifiers [77]. Moreover, it resembles the bagging algorithm in terms of stochastic discrimination theory as a random selection is made by the original training set [78]; however, the RSS is chosen using the original training set of characteristics [79].

This technique has been applied to a significant number of nonlinear problems [78], in various disciplines (medical science, computer science, and banking). The application of the RSS technique in transportation engineering is still limited [73]. A detailed presentation of the RSS technique is as follows:

It is assumed that $X = [x_1, x_2, \ldots, x_n]$ refers to a vector of n number of affecting parameters. Constructing an RSS ensemble to consolidate various classifiers for cataloging purposes, N samples having a size of Z are arbitrarily selected using a uniform distribution over X so that no replacement is required. Every specimen depicts the associated individual subset that expresses a subspace of X. After that, the training of a classifier takes place considering either a single subset or a whole training set [73]. However, the aforementioned amendment is done in the feature space (instead of the instance space). The pseudo-code in the case of the RSS algorithm (Algorithm 1) is reported by [80].

Algorithm 1. RSS algorithm

Input: *Data set D = [(x₁,y₁), (x₂,y₂), . . . , (xₙ,yₙ),)]*
  Base classifier algorithm L;
  Number of subspace rate k;
  Number of learning rounds T.
Process:
For, t = 1,2, . . . , T
Dₜ = RS (D,k);   *% Random generate a subspace sample from D*
Hₜ = L(Dₜ);   *% Train a base classifier hₜ from the subspace sample*
End.
*Output:H(x) =*arg *$max_{y \in x} \sum_{t=1}^{T} 1(y = h_t(x))$; % the value of 1(α) is 1 if α is true*
                          % and 0 otherwise

If the dataset exhibits a variety of redundant or irrelevant parameters, then the reliable base classifiers could be attained in random subspaces in contrast to the original feature space [80].

3.3.2. Reduced Error Pruning Trees (REPT)

The reduced error pruning tree ("REPT") is a mixture of the reduced error pruning (REP) and the decision tree (DT) algorithm technique, which comprises various splits and pruning steps. In this research, the DT was implemented to simplify the modeling process and the REP was incorporated to reduce the complexity of the tree structure. In addition, the REPT uses the validation dataset to accurately predict the generalization error [81,82]. It is important to mention that the pruning phenomenon involved in the REPT algorithm is attributed to the backward over-fitting issue. The REPT algorithm intends to search for the minimal version of the excellent sub-tree on the basis of the post-pruning technique [83].

The aim of the REPT is to reduce the level of modeling complexity when dealing with numerous input data. According to Pham, Jaafari, Nguyen-Thoi, Van Phong, Nguyen, Satyam, Masroor, Rehman, Sajjad and Sahana [82], the REPT technique has been used by numerous researchers to determine an optimal subtree by using the post-pruning technique. The REPT refers to a robust DT learning, such that it establishes a DT on the basis of information gain or variance reduction [84]. The performance of the REPT model is either associated with the information gain obtained from entropy or reducing the variance (as shown in Equation (4) below) and reduced error pruning methods [83]:

$$\text{Gain Ratio}(x, \ Z) = \frac{Entropy(Z) - \sum_{i=1}^{n} \frac{|Z_i|}{|Z|} Entropy(Z_i)}{-\sum_{i=1}^{n} \frac{|Z_i|}{|Z|} \log_2 \frac{|Z_i|}{|Z|}} \tag{4}$$

where attribute x is attributed to a training dataset Z with subsets Z_i, i = 1, 2, . . . , n.

Moreover, REPT may be used to reduce the size of DTs by reducing the complexity of the final classifier. The REPT also increases the degree of estimation accuracy of the classifier since it controls the over-fitting problem alongside the removal of the tree sections that tend to create noisy or erroneous data [85].

Usually, two different techniques are used for pruning the DTs by applying the information gain ratio, i.e., (i) pre-pruning and (ii) post-pruning [83]. Pre-pruning is applied when the number of instances falls below the training set percentage, signifying that this node is aggregated. Post-pruning is used when the DT has been developed to a point such that no problem is encountered in the training set [82]. While comparing the two aforementioned approaches, it is obvious that pre-pruning has the advantage of producing trees faster, while post-pruning has the capability to generate more effective trees [86]. Pre-pruning occurs when the tree expansion is stopped during the data building process. The main advantage of the REPT technique is the reduction of the complexity of

the DT structure, thus avoiding the over-fitting issue in the process of learning such that the accuracy degree is not affected [87].

3.3.3. Extra Tree (ET)

Introduced by Geurts, et al. [88], extra trees (ET) are also known as extremely randomized trees [89], and are essentially an extension of random forest (RF) regression and they incorporate stochasticity in the induction generation of classical DTs, thus forming a more computationally robust AI algorithm. Furthermore, ETs are considered to be the evolutionary version of the RF, yielding good results while simulating complex problems. Note that both these models comprise a series of regression tree models that are formed independently [90,91]. The ETs are different from the RF regression in terms of selecting data to train the model (ET utilizes the whole data while RF uses only a bootstrap replica) and picking the optimal feature for splitting the note (ET picks a much better feature than that of RF). Moreover, ET comprises three main governing parameters: (i) K is the number of randomly chosen variables in order to disintegrate a node, n_{min} depicts the minimum number of specimens needed for splitting an internal node, and M is the number of trees developed inside the model [91].

Multiple DTs are utilized, which accomplish classification as well as regression processes. The feature bagging-based split occurs in two major stages. First, the random subset of features is chosen out of the previously chosen training data subset. After that, in the second stage, the excellent subset feature alongside its corresponding value is selected for performing the decision split. Generally, the most appropriate and robust feature is chosen on the basis of Gini criteria or information gain [92].

3.3.4. Prediction Accuracy Indicators

In this research, the following three performance indicators were used to assess the prediction accuracy of the developed models: the root mean square error (RMSE) [93–96], mean absolute error (MAE) [19,97,98], and correlation coefficient (R^2) [16,97,98]:

$$R = \frac{\sum_{i=1}^{m} (y_{cai} - \overline{y_{ca}})(y_{mei} - \overline{y_{me}})}{\sqrt{\sum_{i=1}^{m} (y_{cai} - \overline{y_{ca}})^2 (y_{mei} - \overline{y_{me}})^2}} \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{m} (y_{cai} - y_{mei})^2} \tag{6}$$

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |y_{cai} - y_{mei}| \tag{7}$$

4. Results and Discussion

4.1. Prediction Accuracy

Figure 5 shows the comparison of the predicted and measured California Bearing Ratio for the three different models. The correlation coefficient R^2 of the various models during the training stage was 0.937, 0.939, and 0.995 for the REPT, RSS-REPT, and RSS-ET models, respectively (Table 3). Correlation coefficient values in excess of 0.8 are generally considered to establish a close agreement between the measured and predicted values [16–18,93,99–103]. However, the prediction accuracy of the testing dataset dropped significantly to $R^2 = 0.709$ and $R^2 = 0.783$ for the REPT and RSS-REPT models, respectively. Whilst these correlation coefficient values are greater than 0.8, they may indicate overfitting issues. The correlation coefficient of the REPT-ET model during the testing stage was $R^2 = 0.968$, establishing the robustness of this model as compared to the REPT and RSS-REPT models. No indication of overfitting issues occurred for the RSS-ET model, which registered similar correlation coefficient values during the training and testing stage.
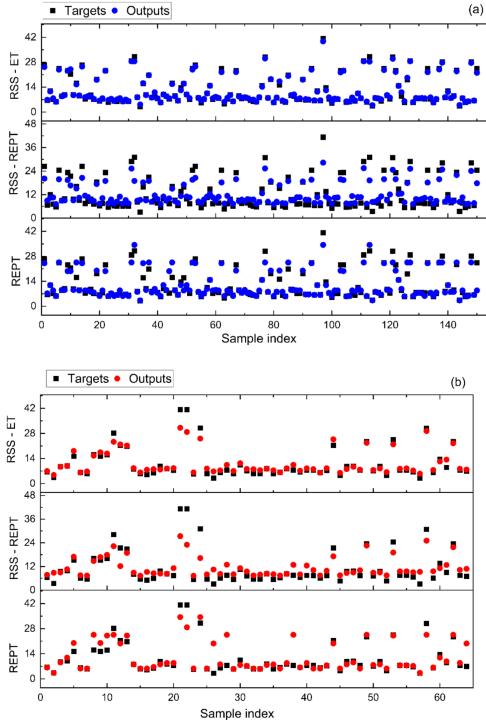
Figure 5. Comparison of the predicted and actual results: (**a**) Training and (**b**) test data.

Table 3. Statistical analysis results stemming from the current study.

No	Parameters	Training	Testing
	R^2		
1	REPT	0.937	0.709
2	Randomsubspace—REPT	0.939	0.783
3	Randomsubspace—Extra Tree	0.995	0.968
	MAE		
1	REPT	1.515	2.956
2	Randomsubspace—REPT	2.586	3.534
3	Randomsubspace—Extra Tree	0.530	1.728
	RMSE		
1	REPT	2.765	6.424
2	Randomsubspace—REPT	3.630	5.885
3	Randomsubspace—Extra Tree	0.960	2.725

The relative error of the developed models generally ranges between ±12.5 for the training and ±17.5 for the testing dataset (Figure 6). The RSS-ET outperformed both the REPT and RSS-REPT models in terms of prediction accuracy. The error for the RSS-ET model denoted by the green line for both the training and testing datasets is the smallest. Table 3 summarizes the prediction accuracy of the models using a variety of performance indexes. The correlation and error analysis data establish the suitability of the RSS-ET model in predicting CBR values within the input data range to which the model was trained and developed.
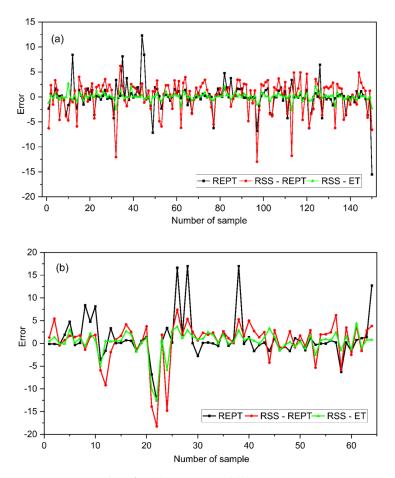

Figure 6. Error values for (**a**) training and (**b**) testing.

4.2. Comparison of Developed Models

Figure 7 shows the accuracy of the developed models during the training and testing stage using the root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R^2). The results show that the random subspace-based extra tree (RSS-ET) model outperforms the random subspace-based (RSS-REPT) and reduced error pruning tree (REPT) models regardless of the performance index is used. Moreover, the prediction accuracy of the random subspace-based extra tree (RSS-ET) model ($R^2 = 0.968$) developed in this research is higher than the prediction accuracy of the soft computing models currently reported in the literature [49,104–107].
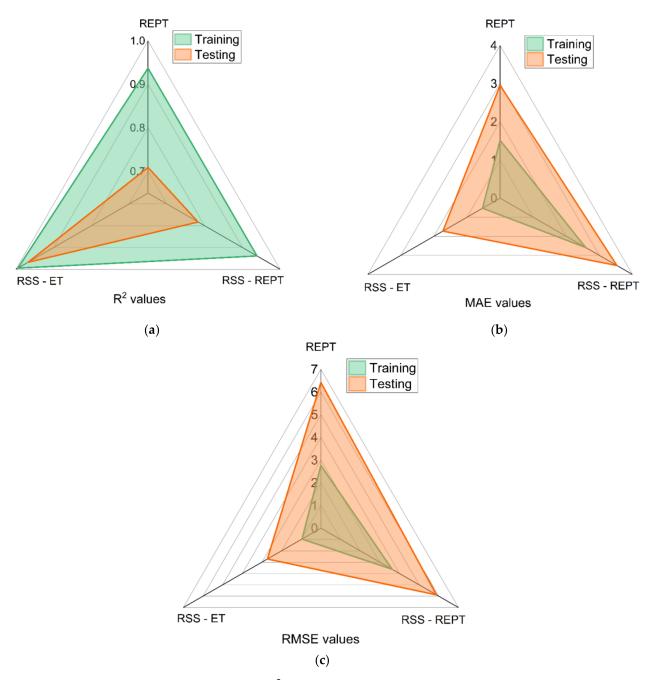

Figure 7. Comparative analysis of R^2 (**a**), RMSE (**b**), and MAE (**c**) of the developed models.

5. Limitations

The proposed random subspace-based extra tree (RSS-ET) neural network can predict the CBR of soil, strictly within the range of parameter values used to train and develop it, which are presented in Table 2. The predictive accuracy of the optimum RSS-ET model may also be affected by the distribution of the parameter values used for training and development. For example, the available fine sand content data are particularly limited within the 20–30% range. As part of the ongoing research, the authors aim to enrich the parameter value range, where a limited amount of data is available and to calibrate the developed random subspace-based extra tree (RSS-ET) over the enriched experimental database.

6. Conclusions

In this research, reduced error pruning trees (REPTs), random subsurface-based REPT (RSS-REPT), and RSS-based extra tree (RSS-ET) models were trained and developed for the prediction of the CBR of soil. The input parameters of the models include the gravel content (G), coarse sand content (CS), fine sand content (FS), silt clay content (SC), organic content (O), liquid limit (LL), plastic limit (PL), plasticity index (PI), optimum moisture content (OMC), and maximum dry density (MDD). The following main conclusions may be drawn:

1. The results show that the random subspace-based extra tree (RSS-ET) model outperformed the random subspace-based REPT (RSS-REPT) and reduced error pruning tree (REPT) models independent of which following performance indices were used: root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R^2).

2. The accuracy of the developed random subspace-based extra tree (RSS-ET) model to predict the CBR of soil was R^2 –0.968 and is higher than the prediction accuracy of the soft computing models currently reported in the literature (Table 1) [16,53,54]. Whilst this is a significantly high prediction accuracy, it is strictly associated with the database used in this research. As part of ongoing research, the authors aim to enrich the parameter value range where a limited amount of data is available and to calibrate the developed random subspace-based extra tree (RSS-ET) over the enriched experimental database.

3. During the testing stage, the correlation coefficient (R^2) values of the REPT and RSS-REPT models were significantly smaller than those obtained during the training stage, indicating overfitting issues. No indication of overfitting issues was observed for the RSS-ET model, which registered similar correlation coefficient values during the training and testing stage.

sios T. Chountalas, Evgenios A. Kotsonis and Athanasia D. Skentou. We thank you the University of Transport Technology for supporting this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANN	Artificial neural network
ANFIS	Adaptive Neuro-Fuzzy Inference System
CBR	California Bearing Ratio
CS	Coarse Sand
DT	Decision Tree
FS	Fine Sand
G	Gravel content
GEP	Gene Expression Programming
GRNN	Generalized Regression Neural Networks
GMDH	Group Method of Data Handling
LL	Liquid Limit
MAE	Mean absolute error
MDD	Maximum Dry Density
MLPN	Multi-Layer Perceptron Neural Networks
ML	Machine learning
O	Organic content
OMC	Optimum Moisture Content
PL	Plastic Limit
PI	Plasticity Index
REPT	Reduced Error Pruning Trees
RMSE	Root Mean Square Error
RSS	Random Subspace
RSS-REPT	Random Subspace based Reduced Error Pruning Trees
RSS-ET	Random Subspace based Extra Tree
R	Correlation coefficient
SC	Soft computing
SC	Silt day Content
SVM	Support Vector Machine
NCHRP	National Cooperative Highway Research Program
LMBP	Levenberg–Marquardt Backpropagation
BP	Bayesian Programming
$\overline{y}_{im}$	mean of measured values
n	Total number of data
y_{im}	Measured values
y_{ip}	Predicted values

References

1. Rehman, Z.; Khalid, U.; Farooq, K.; Mujtaba, H. Prediction of CBR value from index properties of different soils. *Technol. J. Univ. Eng. Technol. (UET)* **2017**, *22*, 17–26.
2. Haupt, F.; Netterberg, F. Prediction of California Bearing Ratio and compaction characteristics of Transvaal soils from indicator properties. *J. S. Afr. Inst. Civ. Eng.* **2021**, *63*, 47–56.
3. Katte, V.Y.; Mfoyet, S.M.; Manefouet, B.; Wouatong, A.S.L.; Bezeng, L.A. Correlation of California bearing ratio (CBR) value with soil properties of road subgrade soil. *Geotech. Geol. Eng.* **2019**, *37*, 217–234. [CrossRef]
4. Nagaraju, T.V.; Prasad, C.D.; Raju, M.J. Prediction of California bearing ratio using particle swarm optimization. In *Soft Computing for Problem Solving*; Springer: New York, NY, USA, 2020; pp. 795–803.
5. Mendoza, C.; Caicedo, B. Elastoplastic framework of relationships between CBR and Young's modulus for granular material. *Road Mater. Pavement Des.* **2018**, *19*, 1796–1815. [CrossRef]
6. Mendoza, C.; Caicedo, B. Elastoplastic framework of relationships between CBR and Young's modulus for fine grained materials. *Transp. Geotech.* **2019**, *21*, 100280. [CrossRef]
7. Brown, S.J.G. Thirty-sixth Rankine Lecture: Soil mechanics in pavement engineering. *Geotechnique* **1996**, *46*, 381–426. [CrossRef]
8. Farias, I.G.; Araujo, W.; Ruiz, G. Prediction of California bearing ratio from index properties of soils using parametric and non-parametric models. *Geotech. Geol. Eng.* **2018**, *36*, 3485–3498. [CrossRef]

9. Abdulnabi, T.Y.; Abdulrazzaq, Z.G. An estimated correlation between California bearing ratio (CBR) with some soil parameters of gypseous silty sandy soils. *Tikrit J. Eng. Sci.* **2020**, *27*, 58–64. [CrossRef]

10. Mousavi, F.; Abdi, E.; Rahimi, H. Effect of polymer stabilizer on swelling potential and CBR of forest road material. *KSCE J. Civ. Eng.* **2014**, *18*, 2064–2071. [CrossRef]

11. Hight, D.W.; Stevens, M.G.H. An analysis of the California Bearing Ratio test in saturated clays. *Geotechnique* **1982**, *32*, 315–322. [CrossRef]

12. Mishra, D.; Tutumluer, E.; Butt, A.A. Quantifying effects of particle shape and type and amount of fines on unbound aggregate performance through controlled gradation. *Transp. Res. Rec. J. Transp. Res. Board* **2010**, *2167*, 61–71. [CrossRef]

13. Wang, G.; Ma, J. A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine. *Expert Syst. Appl.* **2012**, *39*, 5325–5331. [CrossRef]

14. Zeng, J.; Roussis, P.C.; Mohammed, A.S.; Maraveas, C.; Fatemi, S.A.; Armaghani, D.J.; Asteris, P.G. Prediction of Peak Particle Velocity Caused by Blasting through the Combinations of Boosted-CHAID and SVM Models with Various Kernels. *Appl. Sci.* **2021**, *11*, 3705. [CrossRef]

15. Asteris, P.G.; Skentou, A.D.; Bardhan, A.; Samui, P.; Pilakoutas, K. Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. *Cem. Concr. Res.* **2021**, *145*, 106449. [CrossRef]

16. Onyelowe, K.C.; Iqbal, M.; Jalal, F.E.; Onyia, M.E.; Onuoha, I.C. Application of 3-algorithm ANN programming to predict the strength performance of hydrated-lime activated rice husk ash treated soil. *Multiscale Multidiscip. Modeling Exp. Des.* **2021**, 1–16. [CrossRef]

17. Iqbal, M.; Onyelowe, K.C.; Jalal, F.E. Smart computing models of California bearing ratio, unconfined compressive strength, and resistance value of activated ash-modified soft clay soil with adaptive neuro-fuzzy inference system and ensemble random forest regression techniques. *Multiscale Multidiscip. Modeling Exp. Des.* **2021**, *4*, 1–19.

18. Jalal, F.E.; Xu, Y.; Iqbal, M.; Javed, M.F.; Jamhiri, B. Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP. *J. Environ. Manag.* **2021**, *289*, 112420. [CrossRef]

19. Jalal, F.E.; Xu, Y.; Iqbal, M.; Jamhiri, B.; Javed, M.F. Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms. *Transp. Geotech.* **2021**, *30*, 100608. [CrossRef]

20. Iqbal, M.; Zhang, D.; Jalal, F.E.; Javed, M.F. Computational AI prediction models for residual tensile strength of GFRP bars aged in the alkaline concrete environment. *Ocean. Eng.* **2021**, *232*, 109134. [CrossRef]

21. Luo, X.; Lin, F.; Chen, Y.; Zhu, S.; Xu, Z.; Huo, Z.; Yu, M.; Peng, J. Coupling logistic model tree and random subspace to predict the landslide susceptibility areas with considering the uncertainty of environmental features. *Sci. Rep.* **2019**, *9*, 15369. [CrossRef] [PubMed]

22. Pham, B.T.; Phong, T.V.; Nguyen-Thoi, T.; Parial, K.; Singh, K.S.; Ly, H.-B.; Nguyen, K.T.; Ho, L.S.; Le, H.V.; Prakash, I. Ensemble modeling of landslide susceptibility using random subspace learner and different decision tree classifiers. *Geocarto Int.* **2020**, 1–23. [CrossRef]

23. Albashish, D.; Hammouri, A.I.; Braik, M.; Atwan, J.; Sahran, S. Binary biogeography-based optimization based SVM-RFE for feature selection. *Appl. Soft Comput.* **2021**, *101*, 107026. [CrossRef]

24. Wang, G.; Ma, J.; Yang, S. An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Syst. Appl.* **2014**, *41*, 2353–2361. [CrossRef]

25. Tan, M.; Yuan, S.; Li, S.; Su, Y.; Li, H.; He, F. Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning. *IEEE Trans. Power Syst.* **2019**, *35*, 2937–2948. [CrossRef]

26. Black, W. A method of estimating the California bearing ratio of cohesive soils from plasticity data. *Geotechnique* **1962**, *12*, 271–282. [CrossRef]

27. de Graft-Johnson, J.; Bhatia, H.; Gidigasu, D. The engineering characteristics of the laterite gravels of Ghana. In Proceedings of the Speciality Session on Engineering Properties of Lateritic Soils. Seventh International Conference on Soil Mechanics and Foundation Engineering, Mexico City, Mexico, 26 November–5 December 1969.

28. Agarwal, K.; Ghanekar, K. Prediction of CBR from plasticity characteristics of soil. In Proceedings of the 2nd South-East Asian Conference on Soil Engineering, Singapore, 11–15 June 1970; pp. 571–576.

29. Al-Suhaibani, A.; Al-Refeai, T.; Noureldin, A. Characterization of Subgrade Soil in Saudi Arabia; A study of Resilient Behavior. KACST Project No. AR-12-51, Final Report; 1997. *Emir. J. Eng. Res.* **2003**, *8*, 41–49.

30. Stephens, D. Variation of the California bearing ratio in some synthetic clayey soils. *Civ. Eng. Siviele Ing.* **1992**, *1992*, 379–380.

31. Peshkin, D.; Hoerner, T.; Zimmerman, K. *National Cooperative Highway Research Program, NCHRP, Report 523 Optimal Timing of Pavement Preventive Maintenance Treatment Applications*; Transportation Research Board, TRB, National Research Council: Washington, DC, USA, 2004.

32. Kin, M. Judul: California Bearing Ratio Correlation with Soil Index Properties. Master's Thesis, University Technology Malaysia, Skudai, Malaysia, 2006.

33. Suthar, M.; Aggarwal, P. Predicting CBR value of stabilized pond ash with lime and lime sludge using ANN and MR models. *Int. J. Geosynth. Ground Eng.* **2018**, *4*, 1–7. [CrossRef]

34. Bhatt, S.; Jain, P.K.; Pradesh, M. Prediction of California bearing ratio of soils using artificial neural network. *Am. Int. J. Res. Sci. Technol. Eng. Math.* **2014**, *8*, 156–161.

35. Alawi, M.; Rajab, M. Prediction of California bearing ratio of subbase layer using multiple linear regression models. *Road Mater. Pavement Des.* **2013**, *14*, 211–219. [CrossRef]
36. Kurnaz, T.F.; Kaya, Y. Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network. *Eur. Phys. J. Plus* **2019**, *134*, 326. [CrossRef]
37. Talukdar, D.K. A study of correlation between California Bearing Ratio (CBR) value with other properties of soil. *Int. J. Emerg. Technol. Adv. Eng.* **2014**, *4*, 559–562.
38. Erzin, Y.; Turkoz, D. Use of neural networks for the prediction of the CBR value of some Aegean sands. *Neural Comput. Appl.* **2016**, *27*, 1415–1426. [CrossRef]
39. Merghadi, A.; Abderrahmane, B.; Tien Bui, D. Landslide susceptibility assessment at Mila Basin (Algeria): A comparative assessment of prediction capability of advanced machine learning methods. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 268. [CrossRef]
40. Taskiran, T. Prediction of California bearing ratio (CBR) of fine grained soils by AI methods. *Adv. Eng. Softw.* **2010**, *41*, 886–892. [CrossRef]
41. Yildirim, B.; Gunaydin, O. Estimation of California bearing ratio by using soft computing systems. *Expert Syst. Appl.* **2011**, *38*, 6381–6391. [CrossRef]
42. Kumar, S.A.; Kumar, J.P.; Rajeev, J. Application Of Machine Learning Techniques To Predict Soaked CBR Of Remolded Soils. *IJERT* **2013**, *2*, 3019–3024.
43. Varghese, V.K.; Babu, S.S.; Bijukumar, R.; Cyrus, S.; Abraham, B.M. Artificial neural networks: A solution to the ambiguity in prediction of engineering properties of fine-grained soils. *Geotech. Geol. Eng.* **2013**, *31*, 1187–1205. [CrossRef]
44. Sabat, A.K. Prediction of California bearing ratio of a stabilized expansive soil using artificial neural network and support vector machine. *Electron. J. Geotech. Eng.* **2015**, *20*, 981–991.
45. Ghorbani, A.; Hasanzadehshooiili, H. Prediction of UCS and CBR of microsilica-lime stabilized sulfate silty sand using ANN and EPR models; application to the deep soil mixing. *Soils Found.* **2018**, *58*, 34–49. [CrossRef]
46. Alam, S.K.; Mondal, A.; Shiuly, A. Prediction of CBR value of fine grained soils of Bengal Basin by genetic expression programming, artificial neural network and krigging method. *J. Geol. Soc. India* **2020**, *95*, 190–196. [CrossRef]
47. Roy, B.; Singh, M.P. An empirical-based rainfall-runoff modelling using optimization technique. *Int. J. River Basin Manag.* **2020**, *18*, 49–67. [CrossRef]
48. Taha, S.; Gabr, A.; El-Badawy, S. Regression and Neural Network Models for California Bearing Ratio Prediction of Typical Granular Materials in Egypt. *Arab. J. Sci. Eng.* **2019**, *44*, 8691–8705. [CrossRef]
49. Tenpe, A.R.; Patel, A. Application of genetic expression programming and artificial neural network for prediction of CBR. *Road Mater. Pavement Des.* **2020**, *21*, 1183–1200. [CrossRef]
50. Al-Busultan, S.; Aswed, G.K.; Almuhanna, R.R.; Rasheed, S.E. Application of artificial neural networks in predicting subbase CBR values using soil indices data. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *671*, 012106. [CrossRef]
51. Ferentinou, M.; Sakellariou, M. Computational intelligence tools for the prediction of slope performance. *Comput. Geotech.* **2007**, *34*, 362–384. [CrossRef]
52. Murlidhar, B.R.; Kumar, D.; Armaghani, D.J.; Mohamad, E.T.; Roy, B.; Pham, B.T. A novel intelligent ELM-BBO technique for predicting distance of mine blasting-induced flyrock. *Nat. Resour. Res.* **2020**, *29*, 4103–4120. [CrossRef]
53. Bardhan, A.; Gokceoglu, C.; Burman, A.; Samui, P.; Asteris, P.G. Efficient computational techniques for predicting the California bearing ratio of soil in soaked conditions. *Eng. Geol.* **2021**, *291*, 106239. [CrossRef]
54. Bardhan, A.; Samui, P.; Ghosh, K.; Gandomi, A.H.; Bhattacharyya, S. ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions. *Appl. Soft Comput.* **2021**, *110*, 107595. [CrossRef]
55. Raja, M.N.A.; Shukla, S.K.; Khan, M.U.A. An intelligent approach for predicting the strength of geosynthetic-reinforced subgrade soil. *Int. J. Pavement Eng.* **2021**, 1–17. [CrossRef]
56. Ministry of Transport and Communications. *TCVN 5729:2012: Expressway—Specifications for Design*; Ministry of Transport and Communications: Hanoi, Vietnam, 2012.
57. AASHTO. *Code T 88-20, Standard Method of Test for Particle Size Analysis of Soils*; American Association of State and Highway Transportation Officials: Washington, DC, USA, 2020.
58. ASTM. *D422-63(2007)e2, Standard Test Method for Particle-Size Analysis of Soils (Withdrawn 2016)*; ASTM International: West Conshohocken, PA, USA, 2007. [CrossRef]
59. AASHTO. *Code T 89-13 (2017), Standard Method of Test for Determining the Liquid Limit of Soils*; American Association of State Highway and Transportation Officials: Washington, DC, USA, 2017.
60. ASTM. *D4318-17e1, Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils*; ASTM International: West Conshohocken, PA, USA, 2017. [CrossRef]
61. AASHTO. *Code T 267-86 (2018), Standard Method of Test for Determination of Organic Content in Soils by Loss of Ignition*; American Association of State and Highway Transportation Officials: Washington, DC, USA, 2018.
62. ASTM. *D4253-16e1, Standard Test Methods for Maximum Index Density and Unit Weight of Soils Using a Vibratory Table*; ASTM International: West Conshohocken, PA, USA, 2016. [CrossRef]
63. ASTM. *D4254-16, Standard Test Methods for Minimum Index Density and Unit Weight of Soils and Calculation of Relative Density*; ASTM International: West Conshohocken, PA, USA, 2016. [CrossRef]

64.	Schaefer, V.R.; Stevens, L.; White, D.; Ceylan, H. *Design Guide for Subgrades and Subbases*; Tech Transfer Summaries: Ames, IA, USA, 2008; Volume 60.

65.	Zhou, W. Cone Penetrometer. In *Encyclopedia of Engineering Geology*; Bobrowsky, P.T., Marker, B., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 182–183.

66.	ASTM. *D1883-16, Standard Test Method for California Bearing Ratio (CBR) of Laboratory-Compacted Soils*; ASTM International: West Conshohocken, PA, USA, 2016. [CrossRef]

67.	AASHTO. *Code T 193-13 (2017), Standard Method of Test for the California Bearing Ratio*; American Association of State and Highway Transportation Officials: Washington, DC, USA, 2017.

68.	Ariema, F.; Butler, B.E. Chapter 6. Embankment Foundations. In *Guide to Earthwork Construction; State of the Art Report 8*; Transportation Research Board National Research Council: Washington, DC, USA, 1990.

69.	Atkins, H. *Highway Materials, Soils and Concretes*; Prentice Hall: Hoboken, NJ, USA, 1997.

70.	Ampadu, S.I.K. A laboratory investigation into the effect of water content on the CBR of a subgrade soil. In *Experimental Unsaturated Soil Mechanics*; Springer: New York, NY, USA, 2007; pp. 137–144.

71.	Khandelwal, M.; Armaghani, D.J.; Faradonbeh, R.S.; Ranjith, P.; Ghoraba, S. A new model based on gene expression programming to estimate air flow in a single rock joint. *Environ. Earth Sci.* **2016**, *75*, 739. [CrossRef]

72.	Momeni, E.; Armaghani, D.J.; Hajihassani, M.; Amin, M.F.M. Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. *Measurement* **2015**, *60*, 50–63. [CrossRef]

73.	Pham, B.T.; Prakash, I.; Bui, D.T. Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology* **2018**, *303*, 256–270. [CrossRef]

74.	Li, H.; Lee, Y.-C.; Zhou, Y.-C.; Sun, J. The random subspace binary logit (RSBL) model for bankruptcy prediction. *Knowl. Based Syst.* **2011**, *24*, 1380–1388. [CrossRef]

75.	Nhu, V.-H.; Khosravi, K.; Cooper, J.R.; Karimi, M.; Kisi, O.; Pham, B.T.; Lyu, Z. Monthly suspended sediment load prediction using artificial intelligence: Testing of a new random subspace method. *Hydrol. Sci. J.* **2020**, *65*, 2116–2127. [CrossRef]

76.	Plumpton, C.O.; Kuncheva, L.I.; Oosterhof, N.N.; Johnston, S.J. Naive random subspace ensemble with linear classifiers for real-time classification of fMRI data. *Pattern Recognit.* **2012**, *45*, 2101–2108. [CrossRef]

77.	Al-rimy, B.A.S.; Maarof, M.A.; Shaid, S.Z.M. Crypto-ransomware early detection model using novel incremental bagging with enhanced semi-random subspace selection. *Future Gener. Comput. Syst.* **2019**, *101*, 476–491. [CrossRef]

78.	García-Pedrajas, N.; Ortiz-Boyer, D. Boosting random subspace method. *Neural Netw.* **2008**, *21*, 1344–1362. [CrossRef] [PubMed]

79.	Hong, H.; Liu, J.; Zhu, A.-X.; Shahabi, H.; Pham, B.T.; Chen, W.; Pradhan, B.; Bui, D.T. A novel hybrid integration model using support vector machines and random subspace for weather-triggered landslide susceptibility assessment in the Wuning area (China). *Environ. Earth Sci.* **2017**, *76*, 1–19. [CrossRef]

80.	Wang, G.; Ma, J. Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Syst. Appl.* **2011**, *38*, 13871–13878. [CrossRef]

81.	Nhu, V.-H.; Shahabi, H.; Nohani, E.; Shirzadi, A.; Al-Ansari, N.; Bahrami, S.; Miraki, S.; Geertsema, M.; Nguyen, H. Daily Water Level Prediction of Zrebar Lake (Iran): A Comparison between M5P, Random Forest, Random Tree and Reduced Error Pruning Trees Algorithms. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 479. [CrossRef]

82.	Pham, B.T.; Jaafari, A.; Nguyen-Thoi, T.; Van Phong, T.; Nguyen, H.D.; Satyam, N.; Masroor, M.; Rehman, S.; Sajjad, H.; Sahana, M. Ensemble machine learning models based on Reduced Error Pruning Tree for prediction of rainfall-induced landslides. *Int. J. Digit. Earth* **2021**, *14*, 575–596. [CrossRef]

83.	Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [CrossRef]

84.	Mohamed, W.N.H.W.; Salleh, M.N.M.; Omar, A.H. A comparative study of reduced error pruning method in decision tree algorithms. In Proceedings of the 2012 IEEE International conference on control system, computing and engineering, Penang, Malaysia, 23–25 November 2012; pp. 392–397.

85.	Pham, B.T.; Prakash, I. A novel hybrid intelligent approach of random subspace ensemble and reduced error pruning trees for landslide susceptibility modeling: A case study at mu cang chai district, yen bai province, viet nam. In Proceedings of the International Conference on Geo-Spatial Technologies and Earth Resources, Hanoi, Vietnam, 5–6 October 2017; pp. 255–269.

86.	Chen, W.; Hong, H.; Li, S.; Shahabi, H.; Wang, Y.; Wang, X.; Ahmad, B.B. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* **2019**, *575*, 864–873. [CrossRef]

87.	Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total. Environ.* **2018**, *627*, 744–755. [CrossRef] [PubMed]

88.	Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]

89.	Marani, A.; Nehdi, M.L. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites. *Constr. Build. Mater.* **2020**, *265*, 120286. [CrossRef]

90.	Alizamir, M.; Kim, S.; Kisi, O.; Zounemat-Kermani, M. Deep echo state network: A novel machine learning approach to model dew point temperature using meteorological variables. *Hydrol. Sci. J.* **2020**, *65*, 1173–1190. [CrossRef]

91.	Ahmad, M.W.; Mouraud, A.; Rezgui, Y.; Mourshed, M. Deep highway networks and tree-based ensemble for predicting short-term building energy consumption. *Energies* **2018**, *11*, 3408. [CrossRef]

92. John, V.; Liu, Z.; Guo, C.; Mita, S.; Kidono, K. Real-time lane estimation using deep features and extra trees regression. In Proceedings of the Image and Video Technology, Sydney, NSW, Australia, 18–22 November 2015; pp. 721–733.

93. Asteris, P.G.; Koopialipoor, M.; Armaghani, D.J.; Kotsonis, E.A.; Lourenço, P.B. Prediction of Cement-based Mortars Compressive Strength using Machine Learning Techniques. *Neural Comput. Appl.* **2020**, *33*, 13089–13121. [CrossRef]

94. Ly, H.-B.; Pham, B.T.; Le, L.M.; Le, T.-T.; Le, V.M.; Asteris, P.G. Estimation of axial load-carrying capacity of concrete-filled steel tubes using surrogate models. *Neural Comput. Appl.* **2021**, *33*, 3437–3458. [CrossRef]

95. Asteris, P.G.; Skentou, A.D.; Bardhan, A.; Samui, P.; Lourenço, P.B. Soft computing techniques for the prediction of concrete compressive strength using Non-Destructive tests. *Constr. Build. Mater.* **2021**, *303*, 124450. [CrossRef]

96. Asteris, P.G.; Lemonis, M.E.; Le, T.T.; Tsavdaridis, K.D. Valuation of the ultimate eccentric load of rectangular CFSTs using advanced neural network modeling. *Eng. Struct.* **2021**, *248*, 113297. [CrossRef]

97. Asteris, P.G.; Cavaleri, L.; Ly, H.-B.; Pham, B.T. Surrogate models for the compressive strength mapping of cement mortar materials. *Soft Comput.* **2021**, *25*, 6347–6372. [CrossRef]

98. Armaghani, D.J.; Asteris, P.G. A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. *Neural Comput. Appl.* **2021**, *33*, 4501–4532. [CrossRef]

99. Asteris, P.G.; Lemonis, M.E.; Nguyen, T.-A.; Van Le, H.; Pham, B.T.J.S.; Structures, C. Soft computing-based estimation of ultimate axial load of rectangular concrete-filled steel tubes. *Steel Compos. Struct.* **2021**, *39*, 471.

100. Zhou, J.; Asteris, P.G.; Armaghani, D.J.; Pham, B.T. Prediction of ground vibration induced by blasting operations through the use of the Bayesian Network and random forest models. *Soil Dyn. Earthq. Eng.* **2020**, *139*, 106390. [CrossRef]

101. Asteris, P.G.; Lourenço, P.B.; Hajihassani, M.; Adami, C.-E.N.; Lemonis, M.E.; Skentou, A.D.; Marques, R.; Nguyen, H.; Rodrigues, H.; Varum, H. Soft computing based models for the prediction of masonry compressive strength. *Eng. Struct.* **2021**, *248*, 113276. [CrossRef]

102. Jalal, F.E.; Xu, Y.; Li, X.; Jamhiri, B.; Iqbal, M. Fractal approach in expansive clay-based materials with special focus on compacted GMZ bentonite in nuclear waste disposal: A systematic review. *Environ. Sci. Pollut. Res.* **2021**, *28*, 43287–43314. [CrossRef] [PubMed]

103. Iqbal, M.F.; Liu, Q.-f.; Azim, I.; Zhu, X.; Yang, J.; Javed, M.F.; Rauf, M. Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *J. Hazard. Mater.* **2020**, *384*, 121322. [CrossRef] [PubMed]

104. Azim, I.; Yang, J.; Javed, M.F.; Iqbal, M.F.; Mahmood, Z.; Wang, F.; Liu, Q.-f. Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. *Structures* **2020**, *25*, 212–228. [CrossRef]

105. Azim, I.; Yang, J.; Iqbal, M.F.; Javed, M.F.; Nazar, S.; Wang, F.; Liu, Q.-f. Semi-analytical model for compressive arch action capacity of RC frame structures. *Structures* **2020**, *27*, 1231–1245. [CrossRef]

106. Tran, T.-H.; Dam, N.D.; Jalal, F.E.; Al-Ansari, N.; Ho, L.S.; Phong, T.V.; Iqbal, M.; Le, H.V.; Nguyen, H.B.T.; Prakash, I.; et al. GIS-Based Soft Computing Models for Landslide Susceptibility Mapping: A Case Study of Pithoragarh District, Uttarakhand State, India. *Math. Probl. Eng.* **2021**, *2021*, 9914650. [CrossRef]

107. Rahimi, I.; Gandomi, A.H.; Asteris, P.G.; Chen, F. Analysis and Prediction of COVID-19 Using SIR, SEIQR, and Machine Learning Models: Australia, Italy, and UK Cases. *Information* **2021**, *12*, 109. [CrossRef]