processes

MDPI

Article

Hybrid ELM and MARS-Based Prediction Model for Bearing Capacity of Shallow Foundation

Manish Kumar [1], Vinay Kumar [2], Rahul Biswas [3], Pijush Samui [2], Mosbeh R. Kaloop [4,5,*], Majed Alzara [6] and Ahmed M. Yosri [6,7]

1 Department of Civil Engineering, SRM Institute of Science and Technology (SRMIST), Tiruchirappalli 621105, India; mkumarod@gmail.com
2 Department of Civil Engineering, NIT, Patna 800005, India; vinay1.ce16@nitp.ac.in (V.K.); pijush@nitp.ac.in (P.S.)
3 Department of Civil Engineering, NIT, Ravangla 737139, India; rahulbiswas@nitsikkim.ac.in
4 Department of Public Works Engineering, Mansoura University, Mansoura 35516, Egypt
5 Department of Civil and Environmental Engineering, Incheon National University, Incheon 22012, Korea
6 Department of Civil Engineering, College of Engineering, Jouf University, Sakakah 72388, Saudi Arabia; alzaramajed@gmail.com (M.A.); amyosri@ju.edu.sa (A.M.Y.)
7 Civil Engineering Department, Faculty of Engineering, Delta University for Science and Technology, Belkas 11152, Egypt
* Correspondence: mosbeh@inu.ac.kr

Abstract: The nature of soil varies horizontally as well as vertically, owing to the process of the formation of soil. Thus, ensuring the safe design of geotechnical structures has been a major challenge. In shallow foundations, conducting field tests is expensive and time-consuming and often conducted on significantly scaled-down models. Empirical models, too, have been found to be the least reliable in the literature. The study proposes AI-based techniques to predict the bearing capacity of a shallow foundation, simulated using the datasets obtained in experiments conducted in different laboratories in the literature. The results of the ELM-EO and ELM-PSO hybrid models are compared with that of the ELM and MARS models. The performance of the models is analyzed and compared with each other using various performance parameters. The models are graded to each other using rank analysis and the visual interpretations are provided using error matrices and REC curves. ELM-EO is concluded to be the best performing model (R^2 and RMSE equal to 0.995 and 0.01, respectively, in the testing phase), closely followed by ELM-PSO, MARS, and ELM. The performance of MARS is better than ELM (R^2 equals 0.97 and 0.5, respectively, in the testing phase); however, hybridization greatly enhances the performance of the ELM and the hybrid models perform better than MARS. The paper concludes that AI-based models are robust and hybridization of regression models with optimization techniques should be encouraged in further research. Sensitivity analysis suggests that all the input parameters have a significant influence on the output, with friction angle being the highest.

Keywords: shallow foundation; AI; ELM; MARS; PSO; EO

Citation: Kumar, M.; Kumar, V.; Biswas, R.; Samui, P.; Kaloop, M.R.; Alzara, M.; Yosri, A.M. Hybrid ELM and MARS-Based Prediction Model for Bearing Capacity of Shallow Foundation. *Processes* **2022**, *10*, 1013. https://doi.org/10.3390/pr10051013

Academic Editor: Anthony Rossiter

Received: 15 April 2022
Accepted: 16 May 2022
Published: 19 May 2022

1. Introduction

Foundations are the most important part of a structure, as they transfer the load of the superstructure to bearing strata. A shallow foundation is generally defined as a foundation having a depth less than or equal to breadth. Shallow foundations have a large base and small thickness, apart from the shallow embedment. It is a system in which the resistance is developed only from its base, and the failure is within the shallow depth extending to the surface. The bearing capacity of soils, also known as the load-carrying capacity of the soil, is one of the most significant topics in soil mechanics and foundation engineering. The greatest value of the load applied for which no point of the subsoil reaches the failure point is the bearing capacity of a shallow foundation [1]. The frictional resistance has

a negligible contribution to the bearing capacity in shallow foundations. Constructions of shallow foundations require less cost, less time, and the least geo-surface disturbance. The ultimate bearing capacity of a foundation is the load per unit area of the foundation at which shear failure occurs. The evaluation of the ultimate bearing capacity can be performed either through field tests such as pile load tests or empirical relations developed by various researchers in the literature. Experimental studies are typically conducted on smaller scale models that are highly scaled-down versions of real footings. For evaluating the final bearing capacity, the size of the foundation is a significant element for both square and rectangular footings. As a result, laboratory-built micro footing models differ from real-world footings in terms of behavior and stress distribution. This is known as the scale effect, and it has been investigated for many years by different researchers [2,3]. Although testing the actual size footing is crucial to understanding genuine soil–foundation behavior, it is an expensive, time-consuming, and analytically challenging process. The plate bearing test, standard penetration test, and pressuremeter test are all field tests that can be used to measure a soil's bearing capacity. These field tests, on the other hand, are time-consuming, costly, and difficult to manage and operate. The expense of testing the structures is so high that it exceeds the structure's cost. As a result, contractors frequently supply estimated footing sizes based on arbitrary assumptions of soil bearing capacity extrapolated from previous site experiences, which is cost-effective but erroneous. Various researchers in the literature proposed empirical relations to predict the bearing capacity of a shallow foundation [4–7]; however, all of these conventional formulations have some limitations and assumptions. As a result, they do not always produce realistic results when compared to experimental data [8–11]. As per the study of Rybak and Krol [12], the limit state is rarely achieved and it is often impossible to estimate the ultimate load.

Soil is naturally heterogeneous and thus uncertainties and variability are involved in its index and material properties. Uncertainty refers to the assessor's lack of awareness (degree of ignorance) of the elements that define the physical system being modeled. The numerous values a property has at different positions, times, or instances are referred to as variability. Due to the heterogeneous nature of the soil, various sources of uncertainties, and various degrees of variability involved, there is a tendency to search for reliable machine learning (ML)-based soft computing models to predict the bearing capacity of shallow foundations. The goal of artificial intelligence is to create machine parts that study human thought patterns and reflect them in reality. Artificial intelligence has found a wide range of applications in civil engineering in recent years [13–17]. Regression analysis is a statistical method of curve fitting that analyzes the relationship between the dependent variable(s) and the predictor variable. Artificial Neural Network (ANN) has been successfully applied in regression analysis of shallow foundation [18–21]; however, the success of ANN is subject to various shortcomings, including 'black-box approach', 'overfitting', low generalization capability, and local minimum. Adem et al. and Padmini et al. [8,22] applied ANN, FIS, and ANFIS to predict the bearing capacity of shallow foundations. ANN performs superior to FIS, but it is proved to be less robust than ANFIS. Baginska and Srokosz [21] applied a deep neural network (DNN) to improve the performance of the neural networks and the best results are achieved for the optimal number of layers 5 to 7. Gaussian process regression (GPR) was tested as a simulation model for the bearing capacity of a shallow foundation and the developed model was concluded to be robust [23]. To improve the sluggish learning speed of traditional feedforward neural networks, Huang et al. [24] proposed an extreme learning machine (ELM), which significantly reduces the training time and improves the generalization performance of the single layer feedforward neural network (SLFN). Though the application of ELM in shallow foundation analysis has been very limited, the model has been found to be robust in various disciplines of civil engineering [25–29]. Khaleel et al. [30] concluded that hybrid ELM models fare better than hybrid multiple linear regression (MLR) models. Traditional ML algorithms, despite yielding better results than statistical techniques, are more vulnerable to being trapped in local minima than catching the actual global minima. This has unfavorable

consequences. As a result, researchers are applying optimization methods with the objective of improving the classical ML parameters and presenting major outcomes to address this challenge [31–33]. Mesut Gor applied hybrid ANN models for the prediction of the bearing capacity of shallow foundations and observed significant improvement in the prediction accuracy post-optimization [34]. Moaeyedi et al. developed a multi-layer perceptron (MLP) combined with an imperialist competitive algorithm (ICA) and obtained an improved R^2 from 0.83 (for ANN) to 0.983 [35]. ELM-PSO has not been applied to the pile foundation problems, although its application in other domains has been very encouraging [36–38]. Several drawbacks of PSO include high computational complexity and premature convergence [39–41]. EO is a fast and powerful metaheuristic optimization technique with high population-based performance [42]. ELM-EO have not been applied in many domains so far; however, it has been proven to be robust in limited applications [43]. The present paper makes a comparative study of the performance of ELM-PSO and ELM-EO with traditional ELM and multivariate adaptive regression splines (MARS) models. MARS has been proven as a significant performing model for the analysis of civil engineering problems [44–49] and also in shallow foundations [47,50].

2. Details of AI-Based Models Used

2.1. MARS

Friedman developed the multivariate adaptive regression spline (MARS) [44,51,52]. MARS is based on the methodology of non-parametric and nonlinear regression techniques. Multivariate adaptive regression analysis helps a large number of independent parameters read a continuous output variable. It is the integration of recursive regression, additive regression, recursive partitioning regression, and spline regression. Multivariate adaptive regression splines select the factors by algorithms using "forward" and "backward" algorithms. The prediction accuracy of MARS with respect to other methods is relatively high and it is also highly adaptive. In general, the equation of non-parametric and nonlinear regression is as follows:

$$y_i = f(x_{i1}, x_{i2}, \ldots, x_{ik}) + \varepsilon_i \quad (1)$$

where

$$f(x_{i1}, x_{i2}, \ldots, x_{ik}) = \text{regression function and should be a smooth, continuous function.}$$

$$\varepsilon_i = \text{estimate of error involved.}$$

The developed model using MARS for predicting the output of given input in the form of y are as follows:

$$y = C_0 + \sum_{m=1}^{M} C_m B_m(x) \quad (2)$$

where C_0 is a constant, $B_m(x)$ is basis function, x is the input variable, and C_m is the coefficient of $B_m(x)$. The spline function consists of two parts, i.e., the left-sided truncated function in Equation (3a) and the right-sided truncated function in Equation (3b), which is as follows:

$$b_q^-(x-t) = [-(x-t)]_+^q = \begin{cases} (t-x)^q & \text{if } x < t \\ 0 & otherwise \end{cases} \quad (3a)$$

$$b_q^+(x-t) = [+(x-t)]_+^q = \begin{cases} (x-t)^q & \text{if } x > t \\ 0 & otherwise \end{cases} \quad (3b)$$

where $b_q^+(x-t)$ and $b_q^-(x-t)$ are the spline functions and t is the knot location. In general, any model based on multivariate adaptive regressions spline (MARS) follows three basic steps such as:

(a) Constructive phase.
(b) Pruning phase.
(c) Selection of optimum MARS.

At the beginning of the constructive phase, the basis function plays an important role in the formation of Equation (2) and the selection of the basis function depends on the generalized cross-validation (GCV). The GCV is adopted as the residual sum of squares of input parameters. The GCV parameter is used as a penalty for model complexity to prevent using a large number of spline functions. The GCV value is calculated from the following Equation (4), which is as follows:

$$GCV(M) = \left(\frac{1}{n}\right) \frac{\sum_{m=1}^{M} (y_i - \hat{y}_i)^2}{\left[1 - \frac{C(M)}{n}\right]^2} \tag{4}$$

The value of $C(M)$ is calculated from the following equation:

$$C(M) = M + dM \tag{5}$$

where

$$y_i = \text{response value for object} i.$$

$$\hat{y}_i = \text{predicted response value for object} i.$$

$$C(M) = \text{penalty factor.}$$

$$n = \text{total number of data objects.}$$

There is a cost penalty factor for maximizing each basis function in Equation (5). Overfitting of data is possible for many basis functions. To circumvent this problem, these basis functions are deleted during the pruning stage. After completing all needed processes, the best MARS model is selected.

2.2. ELM

Huang developed the extreme learning machine in 2004 and published it in 2006 [24,53]. The sigmoid activation principal is used in ELM and consists of three layers of the neural network. The advantages of the extreme learning machine compared to the conventional gradient-based learning methods are as follows:

- It avoids a number of issues that are difficult to deal with in traditional methods, such as halting criteria, learning rate, learning epochs, and local minimums.
- In most circumstances, it can provide better generalized performance than backpropagation (BP) since ELM is a one-pass learning technique that does not require re-iteration.
- It may be used to activate practically any nonlinear function.

The mathematical model of the ELM is described as follows:

Let us consider N samples of data (x_i, t_i) where $x_i = [x_{i1}, \ldots\ldots, x_{im}] \in R^m$ $t_i = [t_{i1}, \ldots\ldots.., t_{iq}] \in R^q$. The extreme learning machine (ELM) algorithm consists of a single hidden layer feedforward neural network with N hidden nodes and the activation function g(x) is shown as:

$$\sum_{i=1}^{\bar{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\bar{N}} \beta_i g_i(w_i.x_j + b_i) = o_j, j = 1, \ldots., N. \tag{6}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{im}]^T$ is the weight vector of the connectors from the input node to the i^{th} hidden node, and $\beta_i = \left[\beta_{i1}, \beta_{i2}, \ldots, \beta_{iq}\right]^T$ is the weight vector of the connectors between the i^{th} hidden node and the output nodes. The variable b_i is the threshold of the i^{th} hidden node.

2.3. PSO

Kennedy and Eberhart [36–38] were the first to suggest particle swarm optimization. A set of iterative techniques capable of leading the search operation to a quality result is referred to as a meta-heuristic. PSO models population social and cooperative behavior (e.g., fish schooling and flocking of birds) during food search. In PSO, each person in the population is referred to as a particle, and the entire population is referred to as a swarm. The swarm is defined as a set:

$$S = \{P_1,\ P_2,\ P_3, \ldots\ldots\ldots, P_N\} \tag{7}$$

of (N) particles (candidate solutions), defined as:

$$P_i = \ (p_{i1},\ p_{i2},\ p_{i3}, \ldots\ldots\ldots, p_{im})^T \ \in \ A, i = 1, 2, 3,\ 4, \ldots\ldots., N \tag{8}$$

where P_i represents an individual particle in defined swarm (S) inside the search space A. Each P_i contains the required number of dimension/control variables designated by $(p_{i1},\ p_{i2},\ \ldots\ldots., p_{im})^T$

$$P_i = \ (x_{il},\ x_{ir},\ \alpha_{il},\ \alpha_{ir})^T \ \in \ \mathrm{A}; i = 1, 2, 3,\ 4, \ldots\ldots.., N \tag{9}$$

The position (X_i^k) and velocity (V_i^k) are constantly updated as the particles move inside the search space A. Here,k represents the number of iteration steps in the PSO algorithm; therefore,

$$X_i^k = \left(x_{i1}^k, x_{i2}^k, x_{i3}^k, x_{i4}^k\right) \ \in \ \mathrm{A}, i = 1, 2, 3,\ 4, \ldots\ldots.., N \tag{10a}$$

$$V_i^k = \ (\ v_{i1}^k, v_{i2}^k,\ v_{i3}^k\ , v_{i4}^k\Big)^T i = 1, 2, 3,\ 4, \ldots\ldots.., N \tag{10b}$$

In the early form of PSO, each particle employs Equation (11) to update its velocity.

$$V_i^{k+1} = V_i^k + c_1 \times rand_1 \times \left(X_{pbest}^k - X_i^k\right) + c_2 \times rand_2 \times \left(X_{sbest}^k - X_i^k\right) \tag{11}$$

The problem of premature convergence is observed to plague the early PSO variant [54,55]. To alleviate this problem, another parameter (ω), the inertia weight coefficient is introduced to the original equation resulting in the newly formed velocity Equation (12) of PSO

$$V_i^{k+1} = \omega^{\mathrm{k}} \times V_i^k + c_1 \times rand_1 \times \left(X_{pbest}^k - X_i^k\right) + c_2 \times rand_2 \times \ \left(X_{sbest}^k - X_i^k\right) \tag{12}$$

The inertia weight (ω) can be assumed to follow a linearly decreasing pattern between maximum $(\omega_{\max})$ and minimum $(\omega_{\min})$ values.

$$\omega^k = \omega_{\max} - (\omega_{\max} - \omega_{\min})\ \frac{k}{k_{max}} \tag{13}$$

Later, a contemporary standard PSO (CS-PSO) version was developed by Clerc and Kennedy (2002), in which the velocity of the particles is updated as follows:

$$V_i^{k+1} = \eta \times V_i^k + c_1 \times rand_1 \times \left(X_{pbest}^k - X_i^k\right) + c_2 \times rand_2 \times \ \left(X_{sbest}^k - X_i^k\right) \tag{14}$$

For all the above cases mentioned, the particle's positions are updated as follows:

$$X_i^{k+1} = X_i^k + V_i^{k+1} \tag{15}$$

The constriction coefficient (η) is determined as suggested by Clerc and Kennedy (2002):

$$\eta = \frac{2}{\left|2 - \phi - \sqrt{\phi^2 - 4\phi}\right|} \tag{16}$$

where $\phi = c_1 + c_2$.

The cognitive (c_1) and social (c_2) parameters in Equation (10) in the CS-PSO variant is equal to 2.05. In this present work, contemporary standard PSO with velocity clamping is used. This is required so that the modified velocity does not make the particle being moved away from the domain of interest.

2.4. EO

The equilibrium optimizer (EO) is a newly developed meta-heuristics optimization technique by Faramarzi et al. [42]. The equilibrium optimizer algorithm was developed by modeling the theory of dynamic mass balance mathematically. The EO functions similarly to other meta-heuristics in that it searches through a list of potential solutions. Concentration vectors are the names given to the possible solutions in the EO. The initialization of these vectors is performed as follows:

$$C_i = C^{\min} + rand \times (C^{\max} - C^{\min})\ i = 1, 2, \ldots, N \tag{17}$$

where C_i signifies the i^{th} concentration vector, the value of the uniformly distributed random number, which is denoted by rand lies 0 and 1; $C^{\min}$ are lower bound for the concentration vectors C_i and $C^{\max}$ are upper bound vectors. After each concentration vector has been initialized, it is iterated and undergoes updating of parameters to offer an optimal solution. Equation (18) gives the formula for updating the concentration vectors:

$$C_i^{t+1} = C_{EQ}^t + (C_i^t - C_{EQ}^t) \times F_i^t + (1 - F_i^t) \times \frac{G_i^t}{\lambda_i^t V_i^t} \tag{18}$$

where t and t + 1 denote the concentration vector iterations (C_i^t and C_i^{t+1} respectively). The first and second term represents the equilibrium concentrations and the global search in the exploration phase, respectively. The third and final term aids the exploitation mechanism by extracting important information from the search space's examined search areas. C_{EQ}^t It is a concentration vector drawn at random from the equilibrium pool.

$$C_{EQ}^t = \left\{C_{EQ1}^t, C_{EQ2}^t, C_{EQ3}^t, C_{EQ4}^t\right\} \tag{19}$$

In terms of fitness, C_{EQ1}^t, C_{EQ2}^t, C_{EQ3}^t, and C_{EQ4}^t vectors represent the first four best concentration vectors. When the optimization search process begins, there is no pre-information of the algorithm's equilibrium state or state of final convergence. As a result, these vectors are thought of as approximations of equilibrium states. F_i^t is used to balance the E&E, and is given mathematically by:

$$F_i^t = \exp((T - T_0)\lambda_i^t) \tag{20}$$

$$T = \left(1 - \frac{t}{t_{\max}}\right)^{a_2 \times \frac{t}{t_{\max}}} \tag{21}$$

$$T_0 = T + \frac{1}{\lambda_i^t} \times \ln(-a_1 sign(r - 0.5)(1 - e^{\lambda T})) \tag{22}$$

where t and $t_{\max}$ indicates the current iteration and the maximum number of iterations, respectively. The strength of the exploration search is set by the value of a_1. A higher value of a_1 stimulates more exploration and vice-versa.$sign(r - 0.5)$ is the factor which controls the direction of E&E. Here r is a random number distributed uniformly between 0 to1. The

a_1 and a_2 are assigned fixed values 2 and 1, respectively, for the optimal model. The final expression of F_i^t can be obtained now:

$$F_i^t = a_1 \times sign(r - 0.5)(e^{-\lambda_i^t T} - 1) \quad (23)$$

G_i^t iteration rate aids in the exploration of the search space and is calculated using Equation (24).

$$G_i^t = G_0^t \times \exp((T - T_0) \times \lambda_i^t) \quad (24)$$

$$G_0^t = p_i^t \times (C_{EQ}^t - \lambda_i^t C_i^t) \quad (25)$$

$$P_i^t = \begin{cases} 0.5 r_1 r_2 \geq g_p^t \\ 0 \; otherwise \end{cases} \quad (26)$$

where r_1 and r_2 are the two random numbers in the range (0,1). P_i^t symbolizes the control parameters' iteration rate, which includes the possibility of iteration term contribution during the search procedure, as well as the probability of this contribution, which specifies how many particles use generation term to update their states, as determined by iteration probability (g_p^t).

2.5. Regression Optimization

Regression optimization is a method of integrating regression techniques and optimization approaches in a single framework for optimizing process control setpoints. To obtain accurate results, regression models have various tunable parameters that are unknown values, such as weight, bias, number of neurons in the intermediate hidden layer, and linear and nonlinear parameters, all of which must be optimized using a robust and reliable optimization procedure. The parameters to be optimized in ELM models are input weights, output weights, and hidden biases. The output weight of ELM is set by random initialization and the pseudo-inverse matrix; however, optimization techniques such as PSO and EO can improve its performance further. It is worth mentioning that initialization weights and biases may have non-optimal values, resulting in poor performance. ELM requires a large number of hidden layer nodes to achieve an expected result, which might lead to overfitting. The ELM parameters are optimized using PSO and EO in this study. In PSO-ELM, the fitness function is the root mean square error (RMSE), and the probable solution is weight and biases in the hidden layer. Before ML learning parameter optimization, the EO algorithm is set up. The population size, maximum iteration count, lower and higher boundaries, and ELM hidden layer neuron count are all variables to consider.

3. Details of Dataset

This study's dataset was compiled from experiments conducted in several laboratories in the literature [21]. These data have five input parameters and one output parameter. The input parameters are the width of the foundation (B), depth of the foundation (D), length to width ratio (L/B), density (γ), and angle of friction (φ). The output parameter is the bearing capacity of the foundation (Q_u). The descriptive statistics of the data are given in Table 1. The histogram with Pearson correlation matrix is shown in Figure 1. As can be seen, the sample variances are scattered in the range of 0, 0.25 to 8860.48, which indicates that the present dataset has a wide range of input and output parameters. In addition, the values of standard error (scattered in the range of 0.01 to 13.18) confirm that the present database consists of a wide range of variables and is hence useful for soft computing modeling.

Table 1. Descriptive statistics of the field data.

	B (m)	D (m)	L/B (-)	γ (KN/m^2)	φ (°)	Q (KPa)
Mean	0.11	0.08	3.92	16.45	38.95	192.84
Minimum	0.06	0.03	1.00	15.70	34.00	58.50
Maximum	0.15	0.15	6.00	17.10	42.50	423.60
Standard Error	0.01	0.01	0.35	0.07	0.44	13.18
Standard Deviation	0.04	0.04	2.47	0.50	3.11	94.13
Sample Variance	0.00	0.00	6.09	0.25	9.67	8860.48
Kurtosis	−1.55	−0.82	−1.94	−1.28	−1.21	−0.38
Skewness	−0.03	0.61	−0.37	−0.22	−0.45	0.65
Range	0.09	0.12	5.00	1.40	8.50	365.10

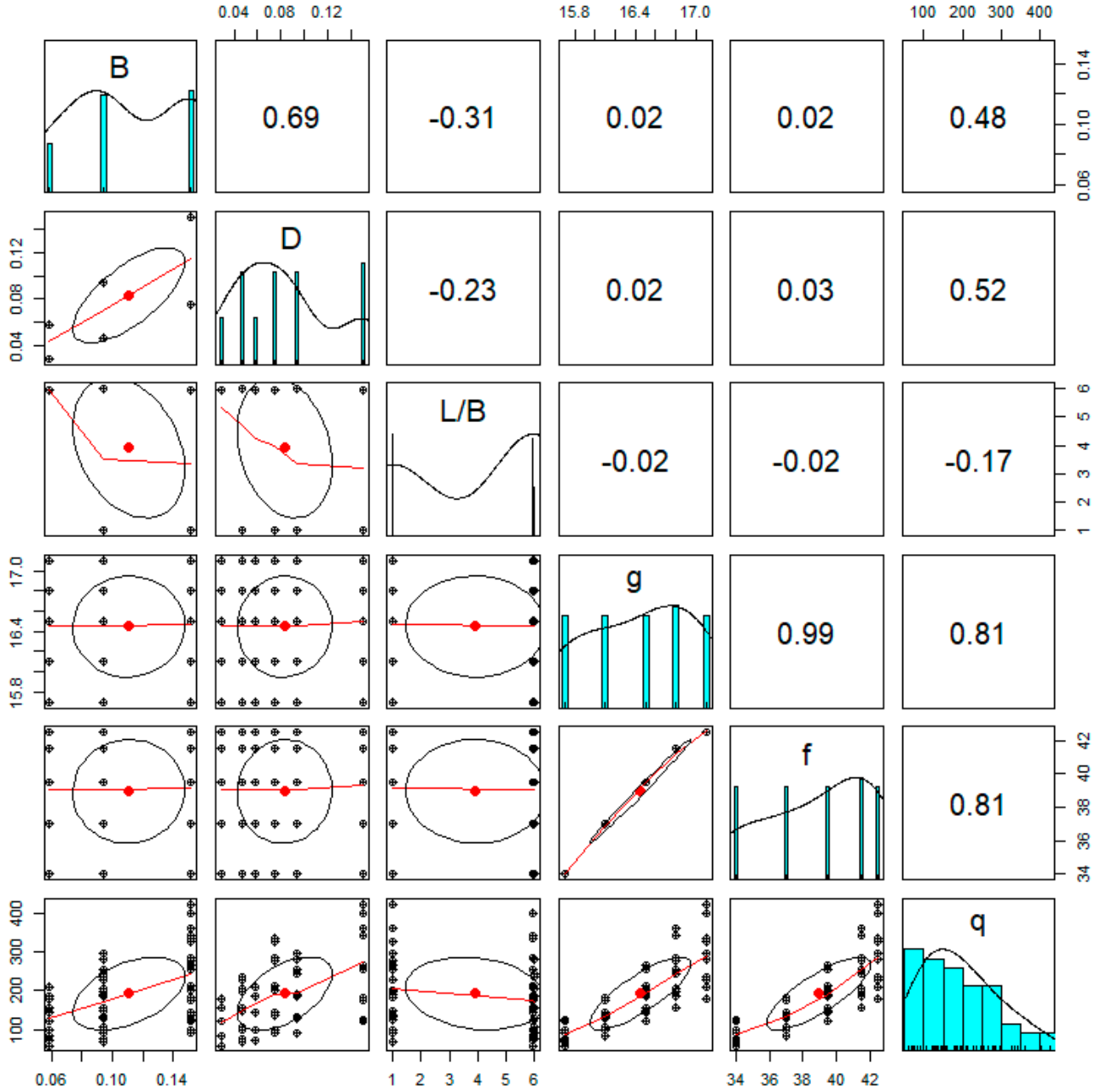

Figure 1. Histogram of the dataset.

4. Research Methodology

The first step in the research methodology is data normalization, i.e., normalizing the data in the range 0 to 1 using the 'min-max method' to bring all the predictor variables in the same range and to reduce the errors. The goal of data normalization is to achieve stable convergence of weight and biases in ML models. In the next step, the normalized data are divided into the training subset (70% of the data) to train the model and the testing subset (30% of the data) to validate the trained model [56]. The model learns from the correlation between input and output variables. The performance of the model was checked using statistical performance parameters. Based on the cost function, several iterations were carried out and the best performing model was selected using rank analysis. The following 14 performance parameters were used in the present paper to evaluate the performance of the simulation models study, namely, weighted mean absolute percentage error (WMAPE), root mean square error (RMSE), variance account factor (VAF), coefficient of determination (R^2), adjusted determination coefficient (Adj. R^2), Nash–Sutcliffe efficiency (NS), performance index (PI), root mean square error to observation's standard deviation ratio (RSR), normalized mean bias error (NMBE), bias, Willmott's index of agreement (WI), mean absolute error (MAE), mean bias error (MBE), and Legate and McCabe's Index (LMI) [57–62].

$$WMAPE = \frac{\sum_{i=1}^{N} \left| \frac{d_i - y_i}{d_i} \right| \times d_i}{\sum_{i=1}^{N} d_i} \tag{27}$$

$$WMAPE = \frac{\sum_{i=1}^{N} \left| \frac{d_i - y_i}{d_i} \right| \times d_i}{\sum_{i=1}^{N} d_i} \tag{28}$$

$$VAF = \left(1 - \frac{var(d_i - y_i)}{var(d_i)}\right) \times 100 \tag{29}$$

$$R^2 = \frac{\sum_{i=1}^{N} (d_i - d_{mean})^2 - \sum_{i=1}^{N} (d_i - y_i)^2}{\sum_{i=1}^{N} (d_i - d_{mean})^2} \tag{30}$$

$$AdjR^2 = 1 - \frac{(n-1)}{(n-p-1)}(1 - R^2) \tag{31}$$

$$PI = adj.R^2 + (0.01 \times VAF) - RMSE \tag{32}$$

$$NS = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - y_{mean})^2} \tag{33}$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (d_i - d_{mean})^2}} \tag{34}$$

$$\text{Bias Factor} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{d_i} \tag{35}$$

$$NMBE(\%) = \frac{\frac{1}{N} \sum_{i=1}^{n} (y_i - d_i)}{\frac{1}{N} \sum_{i=1}^{n} d_i} \tag{36}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{N} (d_i - y_i)^2}{\sum_{i=1}^{N} \{|y_i - d_{mean}| + |d_i - d_{mean}|\}^2} \right] \tag{37}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |(y_i - d_i)| \tag{38}$$

$$MBE = \frac{1}{N} \sum_{i=1}^{n} (y_i - d_i) \tag{39}$$

$$LMI = 1 - \left[\frac{\sum_{i=1}^{N} |d_i - y_i|}{\sum_{i=1}^{N} |d_i - d_{mean}|} \right], \; 0 < LMI \leq 1 \tag{40}$$

where d_i is the observed i^{th} value, y_i is the predicted i^{th} value, d_{mean} is the average of the observed value, and N is the number of data samples. Note that, for an ideal model, the values of these indices should be equal to their ideal values, the details of which are presented in Table 2.

Table 2. Optimal values of effective parameters of MARS model.

Parameters	MARS-L
GCV penalty per knot	0
Cubic modelling	0 (No)
Self-interactions	1 (No)
Maximum interactions	2
Prune	1 (true)
No. of F_b in the final model	15

5. Results and Discussion

5.1. Configuration of the Models

The development of the MARS model was made in MATLAB. MARS model started with 10 basis functions and the final best performing MARS model was taken for nine basis functions after the pruning phase. The equations of the basis functions are shown in Table 3. The model is piecewise-cubic, but the equations of basis functions are shown as piecewise-linear. The final output equation is provided in Equation (41).

$$Q = 0.169 + 0.628{*}\text{BF1} - 0.233{*}\text{BF2} + 0.344{*}\text{BF3} - 0.612{*}\text{BF4} + 0.073{*}\text{BF5} - 0.116{*}\text{BF6} + 0.535{*}\text{BF7} + 0.149{*}\text{BF8} - 0.108{*}\text{BF9} \rightarrow \tag{41}$$

Table 3. Equation of basic functions in MARS model.

SL.NO	Basic Function	Equation
1	BF1	max(0, φ − 0.352)
2	BF2	max(0, 0.352 − φ)
3	BF3	BF1 × max(0, D − 0.380)
4	BF4	Bf1 × max(0, 0.380 − D)
5	BF5	max(0, B − 0.379)
6	BF6	max(0, 0.379 − B)
7	BF7	BF5 × max(0, γ − 0.57)
8	BF8	max(0, D − 0.53)
9	BF9	max(0, 0.53 − D)

The ELM model is configured with 25 hidden neurons and a population size of 50. There are five input neurons and one output neuron. The maximum number of iterations was set as 200 and the cost function was RMSE. To optimize the learning parameters, PSO and EO optimization techniques are used in hybrid ELM models. The OAs are initialized to optimize the learning parameters of the ELM such as population size (n_s), the maximum number of iterations (k), lower bound (lb), upper bound (ub), and other parameters besides the number of hidden neurons (Nh) of ELMs. Then, using the training dataset, OAs optimize the weight and biases of the ELMs. RMSE was used as the cost function to estimate the optimal weight and bias values. It is worth noting that, while the values of n_s, k, lb, and ub were kept constant throughout the optimization process, the optimal value of learning parameters differed in each case.

The optimum number of hidden neurons was found by the trial-and-error method by varying the Nh of the same training dataset in the range of 10 to 30 and the configuration for the best performance was taken to simulate the model. The utmost apt value obtained was

25 for ELM, ELM-PSO, and ELM-EO. The values of other parameters were set as n_s = 50, k = 200, lb = −1, and ub = +1; therefore, the optimum number of optimized weight and biases are 176 (5 × 25 + 25 + 25 + 1). The detailed configuration of the developed ELM and optimized ELM models is presented in Table 4.

Table 4. Configuration of optimum hybrid ELM models.

Parameters	n_s	N_h	k	lb	ub
ELM		25			
ELM-PSO	50	25	100	−1	+1
ELM-EO	50	25	100	−1	+1

The actual vs. predicted graph is presented in Figure 2. The straight-line inclined at 45 degrees shows the perfect fit model with actual value = predicted value. As obvious from Figure 2, ELM-EO and ELM-PSO are the best fit model as all the data points are exactly close to the 'actual = predicted' line. The simulation of MARS and ELM is also satisfactory as all the data points are scattered around the straight line but less than in the hybrid ELM models.

5.2. Performance Parameters

R^2 is a statistical performance parameter for curve fitting problems. It evaluates the fit of the predicted data with that of a horizontal straight line (actual = predicted line). The ideal parameter is 1, which represents the perfect fit of the model. If the coefficient is 0.80, then the regression line should contain 80 percent of the points. Adj. R^2 is an improvement to R^2, which adjusts with a number of independent parameters. RMSE is the cost function used to analyze the accuracy of the model. It measures the magnitude of error of all the data points in the regression line—the lower the RMSE, the better fit the model. The ideal value of RMSE is 0. WMAPE is a statistical measure of the accuracy of the simulation model. It is an improvement over mean absolute percentage error where weighted errors are calculated. NS is a ratio of residual error variance (noise) to measured variance in observed data. NS values of less than one are unacceptable. Values ≥ 1 are desirable. VAF is the variance accounted for among original and predicted values of regression models. Perfect models have 100% VAF. NMBE computes the correlation between the predicted value and the mean value. It normalizes the MBE by dividing it with the mean of the observed values. A bias factor value of 1 means balanced prediction, a value greater than 1 means overprediction, and a value less than 1 means underprediction. The ideal values of the performance parameters are given in Table 5.

It can be inferred from Table 6 that the output performance parameter values for all the models are close to the ideal values for both the training and testing phases. The R^2 and RMSE for ELM-EO are 0.999 and 0 for training and 0.995 and 0.01 for testing, respectively. Further, the comparison between the performances of the models is made using rank analysis, which is elaborated on in the next section. The values of performance parameters for ELM are comparatively less satisfactory (R^2 = 0.94 and RMSE = 0.0558 in testing).

5.3. Rank Analysis

Rank analysis is the most straightforward and widely used method for determining the effectiveness of developed models and comparing their robustness. The statistical parameters are used to assign the score value in this study, with their ideal values serving as the benchmark. It depends on how many models are used. The greatest score is given to the best performing results model, and vice versa. The ranking ratings for two models with the same outcomes may be the same. The overall score of a model is calculated by adding

the scores value of the training phase and testing phase. The equation used to calculate the total score is given as

$$Total\ score = \left[\sum_{i=1}^{m} X_i + \sum_{j=1}^{n} X_j\right] \quad (42)$$

where X_i and Xj are the scores of the performance indicators for the training and testing phase, respectively. The number of performance indicators in the training and testing phase is represented by m and n, respectively.

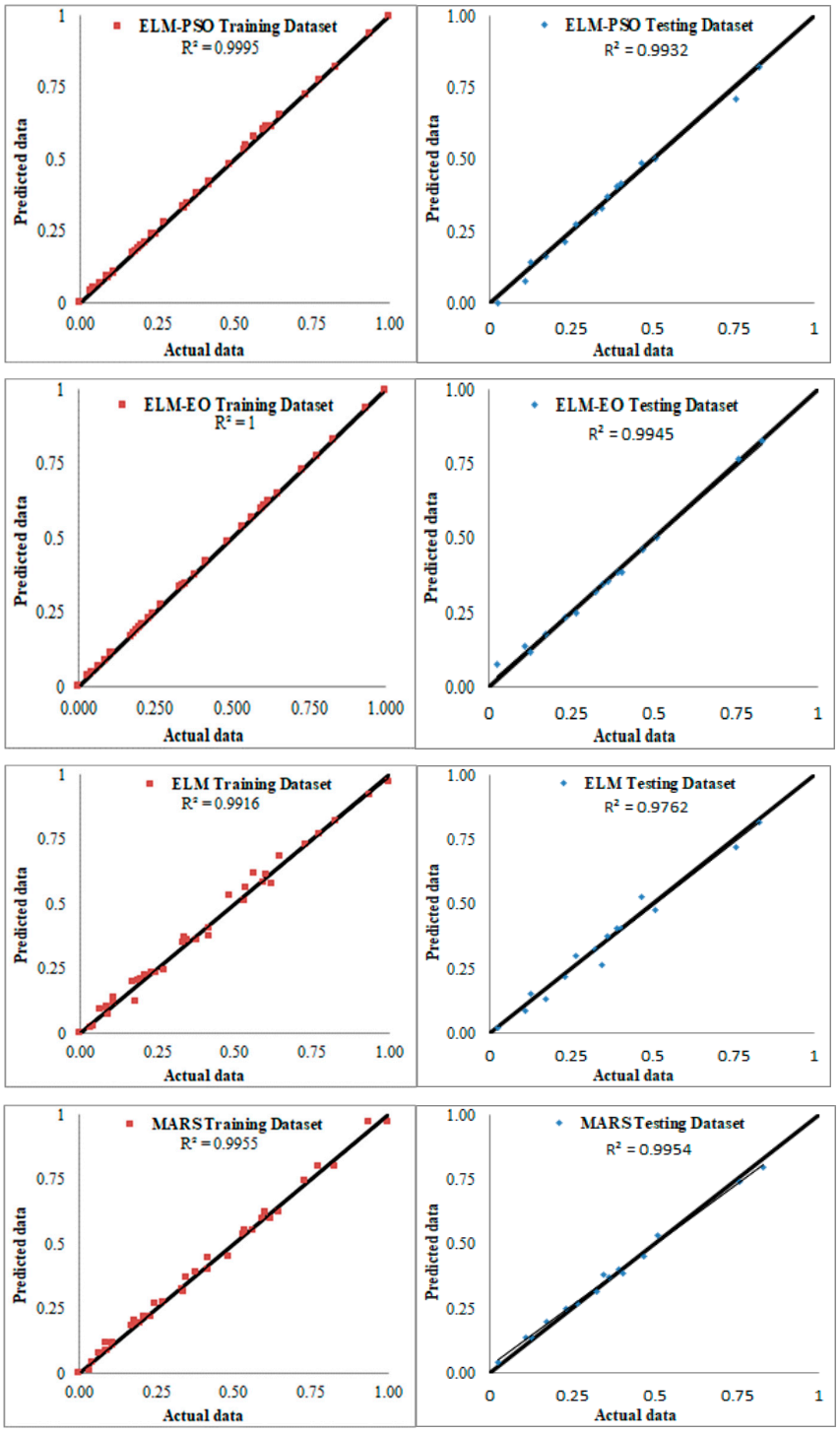

Figure 2. Scatter plot of the applied models.

Table 5. Ideal limit of statistical parameters.

Parameters	Ideal Value	Parameters	Ideal Value
VAF	100	RMSE	0
R^2	1	WMAPE	0
PI	2	MAE	0
WI	1	MBE	0
Adj. R^2	1	NMBE	0
NS	1	LMI	0
RSR	0	Bias	1

Table 6. Values of the performance parameters.

Model Statistical Parameters	ELM	ELM-EO	ELM-PSO	MARS	ELM	ELM-EO	ELM-PSO	MARS
	Testing Performance				Training Performance			
WMAPE	0.0797	0.0306	0.0441	0.0498	0.0543	0.0030	0.0127	0.0396
RMSE	0.0558	0.0170	0.0186	0.0199	0.0248	0.0014	0.0060	0.0180
VAF	93.921	99.3963	99.3155	99.3155	99.1566	99.9973	99.951	99.5517
R^2	0.9425	0.9945	0.9932	0.9954	0.9915	0.9999	0.9995	0.9955
Adj. R^2	0.9413	0.9872	0.9840	0.9946	0.9910	0.9999	0.9993	0.9952
NS	0.9386	0.9938	0.9926	0.9916	0.9915	0.9999	0.9995	0.9955
PI	1.8247	1.9641	1.9586	1.9673	1.9578	1.9985	1.9930	1.9727
RSR	0.2477	0.0785	0.0858	0.0916	0.0919	0.0052	0.02200	0.0669
Bias	1.0237	1.1431	0.9178	1.0876	0.9723	0.9731	0.9799	0.9383
NMBE	−1.0848	0.6977	1.205	1.8913	0.1616	0.0087	0.04750	0.1163
WI	0.9830	0.9984	0.9982	0.9978	0.9979	0.9999	0.9998	0.9988
MAE	0.0398	0.01088	0.0157	0.0177	0.0202	0.0012	0.0048	0.0150
MBE	−0.0054	0.00248	−0.0049	0.0067	0.0006	3.24×10^{-5}	0.00018	0.00044
LMI	0.7823	0.93410	0.9052	0.8929	0.9114	0.9950	0.9791	0.9343

The score attained by ELM-EO is the highest in the training phase (52), followed by ELM-PSO (42) and MARS (30) as presented in Table 7. ELM attains the least score value in both testing and training phases (18 each). In the testing phase, too, ELM-EO outperforms all the models (49), while closely followed by ELM-PSO and MARS (38 and 37, respectively). The total rank value attained by ELM-EO in both the phases combined is 101 (52 + 49), which is far ahead of ELM-PSO 80 (42 + 38) and MARS 67 (30 + 37). It can be inferred from the rank values that hybridization has a significant impact, and the efficiency of ELM is enhanced many times. While ELM lags behind MARS in both testing and training phases, hybrid ELM models have performances superior to MARS. EO is more robust in enhancing the performance of ELM than PSO and ELM-EO is the clear winner in terms of performance compared to the other applied models.

5.4. Error Matrix

The error matrix is a tool for displaying the correctness of a model. Figure 3 depicts the amount of error associated with hybrid models based on numerous performance parameters in this section [31]. In this study, the error values for indices R^2, Adj R^2, and RMSE in the range of 0% to 1%, 0% to 6%, and 0% to 6%, which is very satisfactory. Similarly, error value for indices WMAPE, NS, VAF, PI, Bias, WI, MAE, MBE, and LMI are obtained in the range of 0% to 8%, 0% to 6%, 0% to 6%, 0% to 9%, 3% to 14%, 3% to 14%, 0% to 2%, 0% to 4%, 0% to 1%, and 0% to 22%, respectively, in models for training and testing dataset. It is obvious from Figure 3 that the ELM-EO model achieves the least error compared to the other model for all the performance parameters in both training and testing phases. The error for all the parameters is close to 0 in both the phases except NMBE in testing. Thus, ELM-EO can be concluded as the most accurate and robust model from the error matrix. The error percentage for ELM is highest for all the parameters (6% for R^2 and RMSE compared to 0% and 2%, respectively, for MARS and 1% and 2%, respectively, for both ELM-EO and ELM-PSO) in both phases and thus concluded to possess the least accuracy among the applied models.

Table 7. Rank analysis of the simulated model outcomes for testing and training dataset.

Model Statistical Parameters		ELM	ELM-EO	ELM-PSO	MARS	ELM	ELM-EO	ELM-PSO	MARS
		Testing Performance				Training Performance			
WMAPE	Value	0.0797	0.0306	0.0441	0.0498	0.0543	0.0030	0.0127	0.0396
	Score	1	4	3	2	1	4	3	2
RMSE	Value	0.0558	0.0170	0.0186	0.0199	0.0248	0.0014	0.0060	0.0180
	Score	1	4	3	3	1	4	3	3
VAF	Value	93.921	99.3963	99.3155	99.3155	99.1566	99.9973	99.951	99.5517
	Score	1	4	3	3	1	4	3	3
R^2	Value	0.9425	0.9945	0.9932	0.9954	0.9915	0.9999	0.9995	0.9955
	Score	1	3	2	4	1	4	3	2
Adj. R^2	Value	0.9413	0.9872	0.9840	0.9946	0.9910	0.9999	0.9993	0.9952
	Score	1	3	2	4	1	4	3	2
NS	Value	0.9386	0.9938	0.9926	0.9916	0.9915	0.9999	0.9995	0.9955
	Score	1	4	3	2	1	4	3	2
PI	Value	1.8247	1.9641	1.9586	1.9673	1.9578	1.9985	1.9930	1.9727
	Score	1	3	2	4	1	4	3	2
RSR	Value	0.2477	0.0785	0.0858	0.0916	0.0919	0.0052	0.02200	0.0669
	Score	1	4	3	2	1	4	3	2
Bias	Value	1.0237	1.1431	0.9178	1.0876	0.9723	0.9731	0.9799	0.9383
	Score	2	3	1	3	2	3	4	1
NMBE	Value	−1.0848	0.6977	12.0509	1.8913	0.1616	0.0087	0.04750	0.1163
	Score	1	2	4	3	4	1	2	3
WI	Value	0.9830	0.9984	0.9982	0.9978	0.9979	0.9999	0.9998	0.9988
	Score	1	4	3	2	1	4	3	2
MAE	Value	0.0398	0.01088	0.0157	0.0177	0.0202	0.0012	0.0048	0.0150
	Score	1	4	3	2	1	4	3	2
MBE	Value	−0.0054	0.00248	−0.0049	0.0067	0.0006	3.24×10^{-5}	0.00018	0.00044
	Score	4	2	3	1	1	4	3	2
LMI	Value	0.7823	0.93410	0.9052	0.8929	0.9114	0.9950	0.9791	0.9343
	Score	1	4	3	2	1	4	3	2
	Total	18	49	38	37	18	52	42	30

Figure 3. Error matrices for the applied models for (**a**) training dataset and (**b**) testing dataset.

5.5. Sensitivity Analysis

In general, sensitivity analysis (SA) is a technique that is used to determine how changes in input parameters affect the response of the proposed models. This will assist us in identifying the input parameters based on their influence on the result. The cosine amplitude method [63] is used in this work to calculate the amount of influence of the inputs on the response, i.e., the bearing capacity of the pile foundation. The data pairings in this study are represented in a data array, X, as follows:

$$X = \{x_1, x_2, x_3, \ldots, x_i, \ldots, x_n\} \tag{43}$$

and variable *xi* in *X*, is a length vector of *m* as

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}\} \tag{44}$$

The correlation between the strength of the relation (R_{ij}) and datasets of x_i and x_j is provided by

$$R_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}} \tag{45}$$

The graphical representation using a pie chart of R_{ij} in Figure 4 shows the relation between the bearing capacity of soil and the input parameters, as shown in Figure 3. SA reveals that φ has the greatest influence on pile total capacity with a strength value of 0.93 followed by B with a strength value of 0.92. The parameters γ and D have a strength of 0.91; whereas L/B has the minimum effect on the capacity of the pile, i.e., 0.72. It can be concluded that all five parameters have stronger influences on the pile bearing capacity and hence are considered in predicting the output.

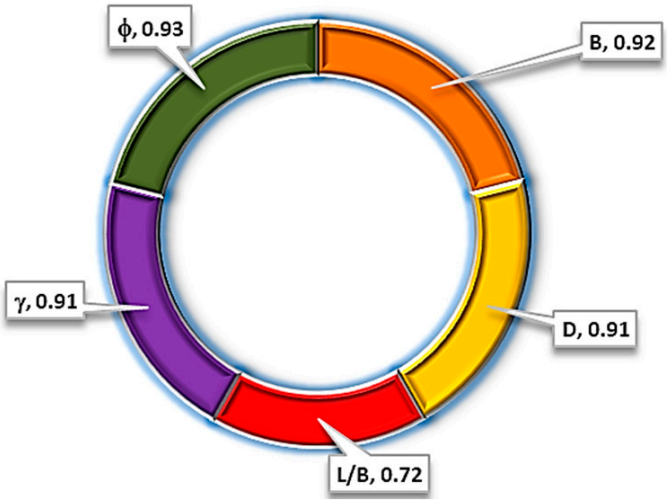

Figure 4. Sensitivity analysis using a pie chart.

5.6. REC Curves

The graph of error tolerance versus the percentage of points predicted inside the tolerance is plotted by the regression error characteristic (REC) curve. The error tolerance and accuracy of a regression function are represented by the x and y axes, respectively. The predicted error is approximated by the area over the REC curve (AOC). The lower the AOC, the better the models' performance. As a result, ROC curves provide for a quick and accurate visual assessment of model performance.

The REC curves for the models are plotted in Figure 5 for both phases. It can be concluded by the visual interpretation itself that ELM is the least accurate model in terms of prediction accuracy. Other models are very close to each other and we need to check the values of AOC for comparison of their performance. The values of the AOC are plotted

in Figure 5. In both training and testing phases the ELM-EO model is a better performing model than other models (AOC value 0.0057 and 0.0122, respectively). In the training phase, the lines for ELM-PSO and ELM-EO are almost overlapping (green line and yellow line) and the AOC values too are close to each other (0.0045 and 0.0057, respectively). Thus, the performances of ELM-PSO and ELM-EO are equally likely in the training phase. In the testing phase, the AOC value of ELM-EO is much better than ELM-PSO (0.0122 and 0.0142) and, of course, far better than MARS (0.0166) and ELM (0.0235).

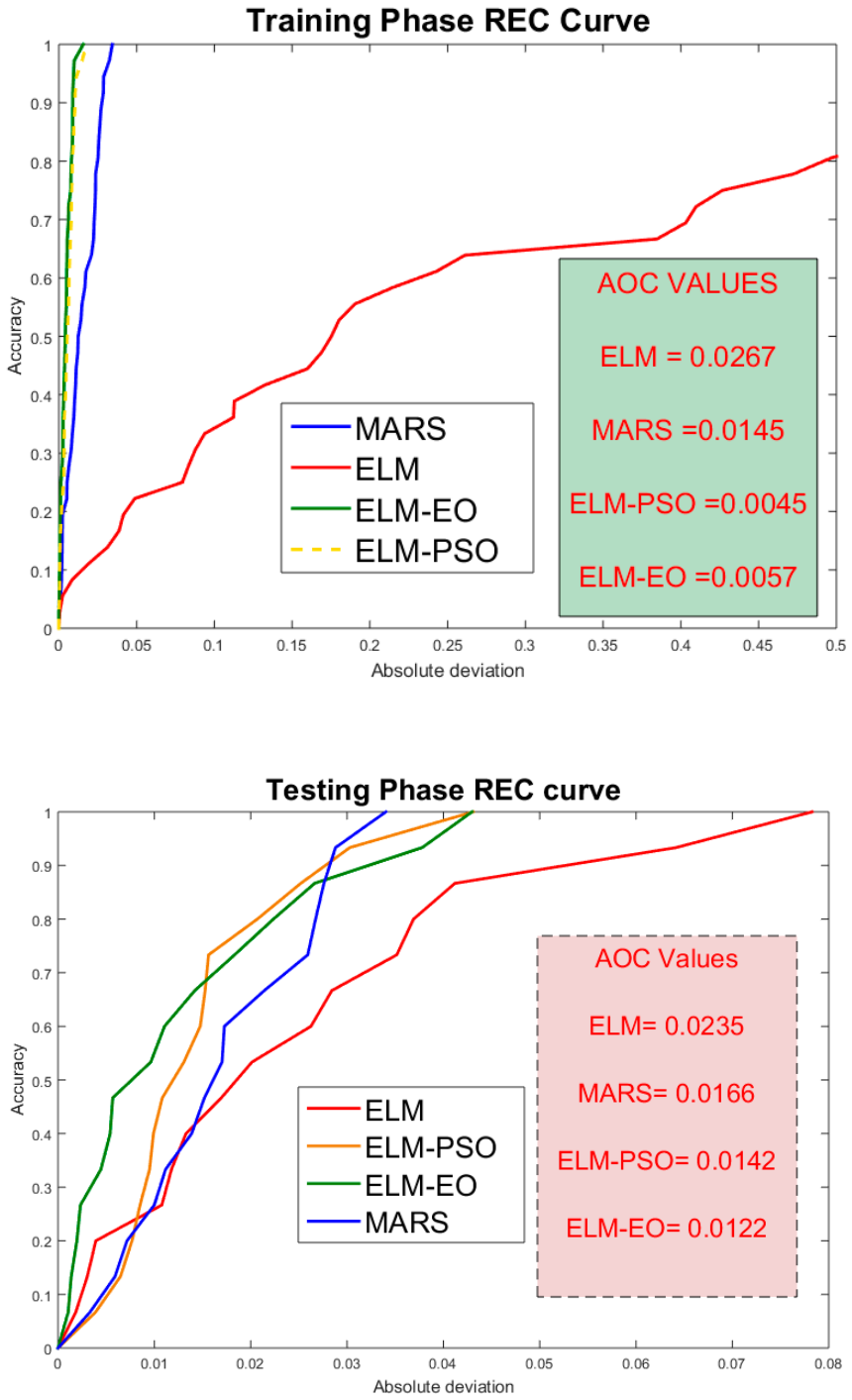

Figure 5. REC curve plot of the simulated models and respective AOC values.

6. Conclusions

The paper presents AI-based models for the prediction of the bearing capacity of shallow foundation as an alternative to the traditional methods, which suffers several practical and performance-based drawbacks. ELM, MARS, ELM-PSO, and ELM-EO models were trained and validated on field data and all the models were found to perform well on the yardsticks of various performance parameters used. In rank analysis, error matrix and REC curves, ELM-EO is concluded to outperform the other models (R^2 = 1, RMSE = 0.004, AOC = 0.0057 in the training phase and R^2 = 0.9945, RMSE = 0.017, AOC = 0.0122 in the testing phase). ELM-EO and ELM-PSO have equally likely performance in the training phase (R^2 close to 1); however, ELM-EO is the best model in the testing phase (R^2 = 0.995 for ELM-EO and 0.993 for ELM-PSO)—it is noteworthy that testing performance is the most important factor in the robustness of the model. The final rank values of ELM, ELM-PSO, and ELM-EO are 36, 80, and 101, respectively. R^2 for MARS is 0.995 in both training and testing phases; thus, the hybridization of the ELM model is concluded to enhance the performance of ELM many miles and further hybridization with various other optimization techniques should be encouraged in future research. The unique advantages of the proposed ELM-EO model are higher prediction accuracy, ease of implementation with the existing datasets, and high generalization capability. On the other hand, the predicting expression of MARS can be used as a user-friendly equation to determine the bearing capacity of the pile. The hybrid ELM models can be extended to other engineering applications once the corresponding database is created. Sensitivity analysis is conducted to assess the impact of input parameters on the output. All the input parameters were found to have a significant impact on the output, friction angle, and L/B ratio having the highest and lowest impact, respectively.

Author Contributions: Conceptualization, M.K., V.K., R.B. and A.M.Y.; data curation, M.K., V.K. and A.M.Y.; funding acquisition, M.R.K.; methodology, M.K., V.K., R.B., P.S. and A.M.Y.; supervision, P.S.; visualization, M.R.K., M.A. and P.S.; writing—original draft, M.K., V.K., R.B. and A.M.Y.; writing—review and editing, M.K., V.K., R.B., M.R.K., M.A. and A.M.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning, Republic of Korea (2019R1I1A1A01062202).

Data Availability Statement: All the data used in the city is properly reported within the text.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Bowles, J.E. *Foundation Analysis and Design*, 4th ed.; McGraw-Hill Education: New York, NY, USA, 1988; p. 1004.
2. Cerato, A.B.; Lutenegger, A.J. Scale Effects of Shallow Foundation Bearing Capacity on Granular Material. *J. Geotech. Geoenviron. Eng.* **2007**, *133*, 1192–1202. [CrossRef]
3. Fukushima, H.I.; Nishimoto, S.; Tomisawa, K. *Scale Effect of Spread Foundation Loading Tests Using Various Size Plates*; Independent Administrative Institution Civil Engineering Research Institute for Cold Region: Hokkaido, Japan, 2005.
4. Terzaghi, K. *Theoretical Soil Mechanics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1943.
5. Meyerhof, G.G. Some Recent Research on the Bearing Capacity of Foundations. *Can. Geotech. J.* **1963**, *1*, 16–26. [CrossRef]
6. Vesic, A.S. Analysis of Ultimate Loads of Shallow Foundations. *ASCE J. Soil Mech. Found. Div.* **1973**, *99*, 45–73. [CrossRef]
7. Aksoy, H.S.; Gör, M.; İnal, E. A New Design Chart for Estimating Friction Angle between Soil and Pile Materials. *Geomech. Eng.* **2016**, *10*, 315. [CrossRef]
8. Kalinli, A.; Acar, M.C.; Gündüz, Z. New Approaches to Determine the Ultimate Bearing Capacity of Shallow Foundations Based on Artificial Neural Networks and Ant Colony Optimization. *Eng. Geol.* **2011**, *117*, 29–38. [CrossRef]
9. Momeni, E.; Nazir, R.; Jahed Armaghani, D.; Maizir, H. Prediction of Pile Bearing Capacity Using a Hybrid Genetic Algorithm-Based ANN. *Meas. J. Int. Meas. Confed.* **2014**, *57*, 122–131. [CrossRef]
10. Kutter, B.L.; Abghari, A.; Cheney, J.A. Strength Parameters for Bearing Capacity of Sand. *J. Geotech. Eng.* **1988**, *114*, 491–498. [CrossRef]

11. Van Baars, S. Numerical Check of the Meyerhof Bearing Capacity Equation for Shallow Foundations. *Innov. Infrastruct. Solut.* **2017**, *31*, 9. [CrossRef]
12. Rybak, J.; Król, M. Limitations and Risk Related to Static Capacity Testing of Piles-"unfortunate Case" Studies. In Proceedings of the MATEC Web of Conferences, Online, 23 February 2018; Juhásová Šenitková, I., Ed.; EDP Sciences: Les Ulis, France, 2018; Volume 146, p. 02006.
13. Farooq, F.; Ahmed, W.; Akbar, A.; Aslam, F.; Alyousef, R. Predictive Modeling for Sustainable High-Performance Concrete from Industrial Wastes: A Comparison and Optimization of Models Using Ensemble Learners. *J. Clean. Prod.* **2021**, *292*, 126032. [CrossRef]
14. Farooq, F.; Czarnecki, S.; Niewiadomski, P.; Aslam, F.; Alabduljabbar, H.; Ostrowski, K.A.; Śliwa-Wieczorek, K.; Nowobilski, T.; Malazdrewicz, S. A Comparative Study for the Prediction of the Compressive Strength of Self-Compacting Concrete Modified with Fly Ash. *Materials* **2021**, *14*, 4934. [CrossRef]
15. Javed, M.F.; Amin, M.N.; Shah, M.I.; Khan, K.; Iftikhar, B.; Farooq, F.; Aslam, F.; Alyousef, R.; Alabduljabbar, H. Applications of Gene Expression Programming and Regression Techniques for Estimating Compressive Strength of Bagasse Ash Based Concrete. *Crystals* **2020**, *10*, 737. [CrossRef]
16. Khan, M.A.; Farooq, F.; Javed, M.F.; Zafar, A.; Ostrowski, K.A.; Aslam, F.; Malazdrewicz, S.; Maślak, M. Simulation of Depth of Wear of Eco-Friendly Concrete Using Machine Learning Based Computational Approaches. *Materials* **2021**, *15*, 58. [CrossRef] [PubMed]
17. Song, H.; Ahmad, A.; Farooq, F.; Ostrowski, K.A.; Maślak, M.; Czarnecki, S.; Aslam, F. Predicting the Compressive Strength of Concrete with Fly Ash Admixture Using Machine Learning Algorithms. *Constr. Build. Mater.* **2021**, *308*, 125021. [CrossRef]
18. Ray, R.; Kumar, D.; Samui, P.; Roy, L.B.; Goh, A.T.C.; Zhang, W. Application of Soft Computing Techniques for Shallow Foundation Reliability in Geotechnical Engineering. *Geosci. Front.* **2021**, *12*, 375–383. [CrossRef]
19. Debnath, S.; Sultana, P. Prediction of Settlement of Shallow Foundation on Cohesionless Soil Using Artificial Neural Network. In *Proceedings of the 7th Indian Young Geotechnical Engineers Conference*; Springer: Singapore, 2022; pp. 477–486. [CrossRef]
20. Samui, P. Application of Statistical Learning Algorithms to Ultimate Bearing Capacity of Shallow Foundation on Cohesionless Soil. *Int. J. Numer. Anal. Methods Geomech.* **2012**, *36*, 100–110. [CrossRef]
21. Bagińska, M.; Srokosz, P.E. The Optimal ANN Model for Predicting Bearing Capacity of Shallow Foundations Trained on Scarce Data. *KSCE J. Civ. Eng.* **2018**, *23*, 130–137. [CrossRef]
22. Padmini, D.; Ilamparuthi, K.; Sudheer, K.P. Ultimate Bearing Capacity Prediction of Shallow Foundations on Cohesionless Soils Using Neurofuzzy Models. *Comput. Geotech.* **2008**, *35*, 33–46. [CrossRef]
23. Ahmad, M.; Ahmad, F.; Wróblewski, P.; Al-Mansob, R.A.; Olczak, P.; Kamiński, P.; Safdar, M.; Rai, P. Prediction of Ultimate Bearing Capacity of Shallow Foundations on Cohesionless Soils: A Gaussian Process Regression Approach. *Appl. Sci.* **2021**, *11*, 10317. [CrossRef]
24. Huang, G.-B.; Kheong Siew, C.; Zhu, Q.-Y.; Siew, C.-K. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks Sentence Level Sentiment Analysis View Project Neural Networks View Project Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004. [CrossRef]
25. Liu, Z.; Shao, J.; Xu, W.; Chen, H.; Zhang, Y. An Extreme Learning Machine Approach for Slope Stability Evaluation and Prediction. *Nat. Hazards* **2014**, *73*, 787–804. [CrossRef]
26. Samui, P.; Kim, D.; Jagan, J.; Roy, S.S. Determination of Uplift Capacity of Suction Caisson Using Gaussian Process Regression, Minimax Probability Machine Regression and Extreme Learning Machine. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2019**, *43*, 651–657. [CrossRef]
27. Samui, P. Application of Artificial Intelligence in Geo-Engineering. In *Springer Series in Geomechanics and Geoengineering*; Springer: Amsterdam, The Netherlands, 2019; pp. 30–44. [CrossRef]
28. Ghani, S.; Kumari, S.; Choudhary, A.K.; Jha, J.N. Experimental and Computational Response of Strip Footing Resting on Prestressed Geotextile-Reinforced Industrial Waste. *Innov. Infrastruct. Solut.* **2021**, *62*, 98. [CrossRef]
29. Kang, F.; Li, J.-S.; Wang, Y.; Li, J. Extreme Learning Machine-Based Surrogate Model for Analyzing System Reliability of Soil Slopes. *Eur. J. Environ. Civ. Eng.* **2017**, *21*, 1341–1362. [CrossRef]
30. Khaleel, F.; Hameed, M.M.; Khaleel, D.; AlOmar, M.K. *Applying an Efficient AI Approach for the Prediction of Bearing Capacity of Shallow Foundations*; Springer: Cham, Switzerland, 2022; pp. 310–323. [CrossRef]
31. Kardani, N.; Bardhan, A.; Samui, P.; Nazem, M.; Zhou, A.; Armaghani, D.J. A Novel Technique Based on the Improved Firefly Algorithm Coupled with Extreme Learning Machine (ELM-IFF) for Predicting the Thermal Conductivity of Soil. *Eng. Comput.* **2021**, 1–20. [CrossRef]
32. Bardhan, A.; GuhaRay, A.; Gupta, S.; Pradhan, B.; Gokceoglu, C. A Novel Integrated Approach of ELM and Modified Equilibrium Optimizer for Predicting Soil Compression Index of Subgrade Layer of Dedicated Freight Corridor. *Transp. Geotech.* **2022**, *32*, 100678. [CrossRef]
33. Kardani, N.; Bardhan, A.; Roy, B.; Samui, P.; Nazem, M.; Armaghani, D.J.; Zhou, A. A Novel Improved Harris Hawks Optimization Algorithm Coupled with ELM for Predicting Permeability of Tight Carbonates. *Eng. Comput.* **2021**, 1–24. [CrossRef]
34. Gör, M. Analyzing the Bearing Capacity of Shallow Foundations on Two-Layered Soil Using Two Novel Cosmology-Based Optimization Techniques. *Smart Struct. Syst.* **2022**, *29*, 513. [CrossRef]

35. Moayedi, H.; Gör, M.; Kok Foong, L.; Bahiraei, M. Imperialist Competitive Algorithm Hybridized with Multilayer Perceptron to Predict the Load-Settlement of Square Footing on Layered Soils. *Measurement* **2021**, *172*, 108837. [CrossRef]
36. Jing, Z. Study on Deformation Law of Foundation Pit by Multifractal Detrended Fluctuation Analysis and Extreme Learning Machine Improved by Particle Swarm Optimization. *J. Yangtze River Sci. Res. Inst.* **2019**, *36*, 53. [CrossRef]
37. Li, W.; Li, B.; Guo, H.; Fang, Y.; Qiao, F.; Zhou, S. The Ecg Signal Classification Based on Ensemble Learning of Pso-Elm Algorithm. *Neural Netw. World* **2020**, *30*, 265–279. [CrossRef]
38. Zeng, J.; Roy, B.; Kumar, D.; Mohammed, A.S.; Armaghani, D.J.; Zhou, J.; Mohamad, E.T. Proposing Several Hybrid PSO-Extreme Learning Machine Techniques to Predict TBM Performance. *Eng. Comput.* **2021**, *1*, 1–17. [CrossRef]
39. Chen, F.; Sun, X.; Wei, D.; Tang, Y. Tradeoff Strategy between Exploration and Exploitation for PSO. *Proc. 2011 7th Int. Conf. Nat. Comput. ICNC* **2011**, *3*, 1216–1222. [CrossRef]
40. Grimaldi, E.A.; Grimaccia, F.; Mussetta, M.; Zich, R.E. PSO as an Effective Learning Algorithm for Neural Network Applications. In Proceedings of the ICCEA 2004. 2004 3rd International Conference on Computational Electromagnetics and its Applications, Beijing, China, 1–4 November 2004; pp. 557–560. [CrossRef]
41. Askarzadeh, A.; Rezazadeh, A. Artificial Bee Swarm Optimization Algorithm for Parameters Identification of Solar Cell Models. *Appl. Energy* **2013**, *102*, 943–949. [CrossRef]
42. Faramarzi, A.; Heidarinejad, M.; Stephens, B.; Mirjalili, S. Equilibrium Optimizer: A Novel Optimization Algorithm. *Knowl.-Based Syst.* **2020**, *191*, 105190. [CrossRef]
43. Kardani, N.; Bardhan, A.; Gupta, S.; Samui, P.; Nazem, M.; Zhang, Y.; Zhou, A. Predicting Permeability of Tight Carbonates Using a Hybrid Machine Learning Approach of Modified Equilibrium Optimizer and Extreme Learning Machine. *Acta Geotech.* **2021**, *17*, 1239–1255. [CrossRef]
44. Samui, P. Determination of Ultimate Capacity of Driven Piles in Cohesionless Soil: A Multivariate Adaptive Regression Spline Approach. *Int. J. Numer. Anal. Methods Geomech.* **2012**, *36*, 1434–1439. [CrossRef]
45. Zhang, W.; Wu, C. Machine Learning Predictive Models for Pile Drivability: An Evaluation of Random Forest Regression and Multivariate Adaptive Regression Splines. In *Springer Series in Geomechanics and Geoengineering*; Springer: Cham, Switzerland, 2020.
46. Samui, P.; Kim, D. Least Square Support Vector Machine and Multivariate Adaptive Regression Spline for Modeling Lateral Load Capacity of Piles. *Neural Comput. Appl.* **2013**, *23*, 1123–1127. [CrossRef]
47. Luat, N.V.; Nguyen, V.Q.; Lee, S.; Woo, S.; Lee, K. An Evolutionary Hybrid Optimization of MARS Model in Predicting Settlement of Shallow Foundations on Sandy Soils. *Geomech. Eng.* **2020**, *21*, 583–598. [CrossRef]
48. Dong, J.; Zhu, Y.; Jia, X.; Shao, M.; Han, X.; Qiao, J.; Bai, C.; Tang, X. Nation-Scale Reference Evapotranspiration Estimation by Using Deep Learning and Classical Machine Learning Models in China. *J. Hydrol.* **2022**, *604*, 127207. [CrossRef]
49. Rahgoshay, M.; Feiznia, S.; Arian, M.; Hashemi, S.A.A. Simulation of Daily Suspended Sediment Load Using an Improved Model of Support Vector Machine and Genetic Algorithms and Particle Swarm. *Arab. J. Geosci.* **2019**, *12*, 227. [CrossRef]
50. Zheng, G.; Zhang, W.; Zhou, H.; Yang, P. Multivariate Adaptive Regression Splines Model for Prediction of the Liquefaction-Induced Settlement of Shallow Foundations. *Soil Dyn. Earthq. Eng.* **2020**, *132*, 106097. [CrossRef]
51. Friedman, J.H. Multivariate Adaptive Regression Splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
52. Kumar, V.; Himanshu, N.; Burman, A. Rock Slope Analysis with Nonlinear Hoek–Brown Criterion Incorporating Equivalent Mohr–Coulomb Parameters. *Geotech. Geol. Eng.* **2019**, *37*, 4741–4757. [CrossRef]
53. Seifi, A.; Ehteram, M.; Singh, V.P.; Mosavi, A. Modeling and Uncertainty Analysis of Groundwater Level Using Six Evolutionary Optimization Algorithms Hybridized with ANFIS, SVM, and ANN. *Sustainability* **2020**, *12*, 4023. [CrossRef]
54. Eberhart, R.C.; Shi, Y. Comparison between Genetic Algorithms and Particle Swarm Optimization. In *Evolutionary Programming VII*; Poroto Saravanam, W.N., Waagen, D., Eiben, A.E., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 611–616.
55. Shi, Y.; Eberhart, R.C. Empirical Study of Particle Swarm Optimization. In Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), Washington, DC, USA, 6–9 July 1999; pp. 1945–1950.
56. Samui, P.; Sitharam, T.G. Site Characterization Model Using Artificial Neural Network and Kriging. *Int. J. Geomech.* **2010**, *10*, 171–180. [CrossRef]
57. Legates, D.R.; Mccabe, G.J. A Refined Index of Model Performance: A Rejoinder. *Int. J. Climatol.* **2013**, *33*, 1053–1056. [CrossRef]
58. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE* **2007**, *50*, 885–900. [CrossRef]
59. Willmott, C.J. On the Evaluation of Model Performance in Physical Geography. In *Spatial Statistics and Models*; Springer: Cham, Switzerland, 1984.
60. Kumar, S.; Rai, B.; Biswas, R.; Samui, P.; Kim, D. Prediction of Rapid Chloride Permeability of Self-Compacting Concrete Using Multivariate Adaptive Regression Spline and Minimax Probability Machine Regression. *J. Build. Eng.* **2020**, *32*, 101490. [CrossRef]
61. Biswas, R.; Samui, P.; Rai, B. Determination of Compressive Strength Using Relevance Vector Machine and Emotional Neural Network. *Asian J. Civ. Eng.* **2019**, *20*, 1109–1118. [CrossRef]
62. Biswas, R.; Rai, B.; Samui, P.; Roy, S.S. Estimating Concrete Compressive Strength Using MARS, LSSVM and GP. *Eng. J.* **2020**, *24*, 41–52. [CrossRef]
63. Biswas, R.; Bardhan, A.; Samui, P.; Rai, B.; Nayak, S.; Armaghani, D.J. Efficient Soft Computing Techniques for the Prediction of Compressive Strength of Geopolymer Concrete. *Comput. Concr.* **2021**, *28*, 221–232. [CrossRef]