

## Article

# Predicting the Ultimate Axial Capacity of Uniaxially Loaded CFST Columns Using Multiphysics Artificial Intelligence

Sangeen Khan <sup>1</sup>, Mohsin Ali Khan <sup>1,2,\*</sup>, Adeel Zafar <sup>1</sup>, Muhammad Faisal Javed <sup>3,\*</sup>, Fahid Aslam <sup>4</sup>,  
Muhammad Ali Musarat <sup>5</sup> and Nikolai Ivanovich Vatin <sup>6</sup>

- <sup>1</sup> Department of Structural Engineering, Military College of Engineering (MCE), National University of Science and Technology (NUST), Islamabad 44000, Pakistan; sakhan.pg18mce@student.nust.edu.pk (S.K.); adeel.zafar@mce.nust.edu.pk (A.Z.)
- <sup>2</sup> Civil Engineering Department, CECOS University of IT and Emerging Science, Peshawar 25000, Pakistan
- <sup>3</sup> Department of Civil Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan
- <sup>4</sup> Department of Civil Engineering, College of Engineering in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia; f.aslam@psau.edu.sa
- <sup>5</sup> Department of Civil and Environmental Engineering, University Teknologi PETRONAS, Bandar Seri Iskandar 32610, Malaysia; muhammad\_19000316@utp.edu.my
- <sup>6</sup> Peter the Great St. Petersburg Polytechnic University, 195291 St. Petersburg, Russia; vatin@mail.ru
- \* Correspondence: arbab\_faisal@yahoo.com (M.F.J.); moak.pg18mce@student.nust.edu.pk (M.A.K.)

**Abstract:** The object of this research is concrete-filled steel tubes (CFST). The article aimed to develop a prediction Multiphysics model for the circular CFST column by using the Artificial Neural Network (ANN), the Adaptive Neuro-Fuzzy Inference System (ANFIS) and the Gene Expression Program (GEP). The database for this study contains 1667 datapoints in which 702 are short CFST columns and 965 are long CFST columns. The input parameters are the geometric dimensions of the structural elements of the column and the mechanical properties of materials. The target parameters are the bearing capacity of columns, which determines their life cycle. A Multiphysics model was developed, and various statistical checks were applied using the three artificial intelligence techniques mentioned above. Parametric and sensitivity analyses were also performed on both short and long GEP models. The overall performance of the GEP model was better than the ANN and ANFIS models, and the prediction values of the GEP model were near actual values. The *PI* of the predicted  $N_{st}$  by GEP, ANN and ANFIS for training are 0.0416, 0.1423, and 0.1016, respectively, and for  $N_{lg}$  these values are 0.1169, 0.2990 and 0.1542, respectively. Corresponding *OF* values are 0.2300, 0.1200, and 0.090 for  $N_{st}$ , and 0.1000, 0.2700, and 0.1500 for  $N_{lg}$ . The superiority of the GEP method to the other techniques can be seen from the fact that the GEP technique provides suitable connections based on practical experimental work and does not rely on prior solutions. It is concluded that the GEP model can be used to predict the bearing capacity of circular CFST columns to avoid any laborious and time-consuming experimental work. It is also recommended that further research should be performed on the data to develop a prediction equation using other techniques such as Random Forest Regression and Multi Expression Program.

**Keywords:** concrete filled steel tube; artificial neural network; multi-physics model; Random Forest Regression; Adaptive Neuro-Fuzzy Inference System; gene expression programming; bearing capacity of columns



**Citation:** Khan, S.; Ali Khan, M.; Zafar, A.; Javed, M.F.; Aslam, F.; Musarat, M.A.; Vatin, N.I. Predicting the Ultimate Axial Capacity of Uniaxially Loaded CFST Columns Using Multiphysics Artificial Intelligence. *Materials* **2022**, *15*, 39. <https://doi.org/10.3390/ma15010039>

Academic Editors: Mingli Cao, Majid Ali, Mehran Khan and Jian-Guo Dai

Received: 24 September 2021  
Accepted: 12 December 2021  
Published: 22 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Concrete Filled Steel Tube Artificial Modelling

Concrete filled steel tube (CFST) is a composite construction element. CFST columns are advantageous due to greater seismic resistance and load bearing capacity, lesser size utilization, good aesthetics, and high fire resistance [1]. The composite action of steel tube

and infilled concrete improves the overall strength and ductility of CFST columns. CFST construction does not require concrete casting formwork, which helps in faster construction with lesser cost [2]. Different types of CFST cross-sections are used in the building industry, such as elliptical, square, rectangular, polygonal, and circular sections. However, this research is concerned with the circular CFST stub and long columns. Generally, the structural performance of circular CFST columns is better than that of polygonal CFST columns [3]. Circular CFST columns have higher ultimate capacities as compared to other cross-sectional shapes. The strength index value for the circular column is higher than 1, which shows the positive confinement effect of the circular CFST column. The rectangular CFST column has less confinement effect, and thus it can be seen from the strength index that the theoretical capacity is greater than the actual capacity [4].

Numerous researchers have conducted experimental studies to check the effect of different parameters on the axial capacity of circular short and long CFST columns [5]. These parameters include concrete compressive strength, diameter, height of the column, yield strength of steel tube, thickness of steel tube, and eccentricity. Furthermore, the experimental research is costly and needs expensive precise equipment. The accuracy of the experimental study relies on skilled labor, type of equipment, condition for casting and testing of specimens and appropriate instrumentation, while in numerical or analytical modelling high computational skills are necessary and need experimental tests merely to validate the model. Thus, developing an accurate, precise, and reliable empirical expression is essential to encompass all the important parameters. The short columns are described as those with  $L/D \leq 4$  (for circular columns) or  $L/B \leq 4$  (for rectangular columns), where  $L$  is the length of the specimens, and  $D$  and  $B$  represent the outer diameter of the circular section the width of the rectangular sections respectively, slender columns having  $L/B > 4$  or  $L/D > 4$  [6].

CFST columns have been used for operation in extreme conditions under high axial loads. High loads require greater dimensions of the CFST cross-section, high strength steel tube and concrete. The dimensions and strength of the material are limited in practice, and these empirically determined limitations are included in building codes. Each code differently interprets the effect of confinement on the overall bearing capacity of CFST columns. Chinese code [7] and Japanese code [8] use allowable concrete strength ( $f_c'$ ) of 67 MPa and 90 MPa and steel tube yield stress ( $f_y$ ) of 420 MPa and 440 MPa, respectively. The recently released AISC 360-16 [9] and AS/NZS 2327 [10] is applicable to  $f_c'$  of 69 MPa and 100 MPa and  $f_y$  is limited to 525 MPa and 690 MPa, respectively. The Eurocode [11] permits  $f_c'$  and  $f_y$  up to 50 MPa and 460 MPa, respectively. In comparison, Liew and Xiong [12] extended these limits to 90 MPa and 460 MPa, respectively. The use of high strength material is valuable for reduction of the size of CFST columns which eventually leads to the savage of floor space and lesser construction cost. The high strength steel tube enhances the elastic behavior and thus improves the confinement effect towards the concrete core. Use of concrete in CFST helps in the functional optimization of both materials. Advancements in the construction industry permit high strength materials to be practically utilized. In addition, the equations available in the mentioned standard codes do not agree with each other. Moreover, these codes are based on the pre-assumed stress–strain curve of CFST, which makes the validity of the presented equations suspicious. To tackle this issue, many researchers have conducted experimental studies on the utilization of high strength materials in CFST columns. Khan et al. [13] use  $f_c'$  and  $f_y$  up to 113 MPa and 762 MPa, respectively, in CFST columns. Mursi and Uy [14] and Sakino et al. use normal strength concrete in CFST columns with  $f_y$  up to 761 MPa and 853 MPa, respectively.

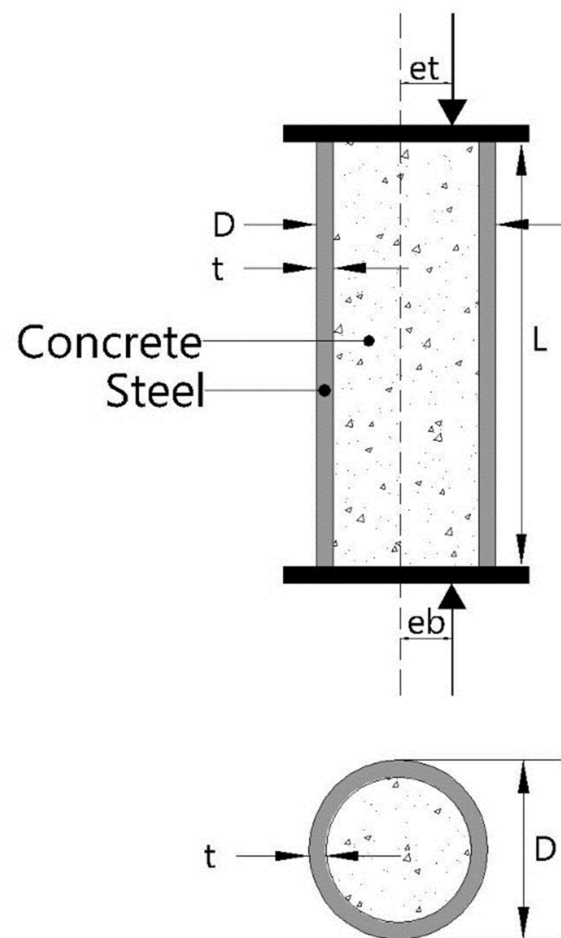
Different studies recommended various methods for the estimation of the bearing capacity of CFST columns [15]. Researchers have found nonlinear and linear regression methods to be highly effective in the civil engineering field [16]. However, the development of these models is based on pre-assumed equations, which makes them impracticable and unrealistic in terms of estimation perspective [17,18]. To tackle this problem, recently various artificial intelligence (AI) techniques, specifically machine learning methods, have

been extensively used in the field of civil engineering) [19–21]. Researchers have used Artificial neural network (ANN) [22], Support vector machine (SVM), [23] random forest regression (RFR) [24,25], adaptive neuro-fuzzy interface system (ANFIS) [26], feed-forward neural network (FNN) [27], particle swarm optimization (PSO) [28], genetic programming (GP) [29], gene expression programming (GEP) [30], etc. For the estimation of mechanical properties of different types of civil engineering materials and structures. Nguyen et al. [31] estimated the axial capacity of rectangular CFST via the FNN algorithm. Researchers also projected the relationship between load deformation and fire resistance of CFST stub columns using the ANN technique [32].

Other studies [33] proposed an ANN model to estimate the ultimate capacity of rectangular CFST beam-columns and circular CFST beams. It was concluded that the predictive model performed better than EC4 and AISC projected equations in both cases. Likewise, the authors focused on an alternative technique to estimate the ultimate axial capacity of stub CFST columns and confinement performance of infilled concrete ANFIS [34,35]. They stated that the ANFIS model performs significantly better and is more accurate than multiple linear regression (MLR) and multiple non-linear regression (NLMR). The researchers also formulated the punching shear strength of concrete slabs using ANN and GP algorithms [36]. The GEP empirical model for predicting the axial capacity of short circular CFST columns performed better than other formulae available in different design codes [37]. Similarly, researchers use gene expression programming (GEP) to deliver an empirical equation for the estimation of the axial capacity of concrete filled double skin tube (CFDST) columns [38] and short CFST columns [39]. They testified that GEP predictive models yield better performance than available equations, giving lesser error values with a higher correlation coefficient.

Furthermore, the soft computing algorithms solve problems by training the available data set to obtain results, which are then validated via validation set data [40]. However, ANN and ANFIS algorithm-based prediction needs many improvements to provide a practical equation for future use. Therefore they are black boxes [41]. Numerous hidden neurons collapse ANN and ANFIS algorithms to provide a practicable empirical equation between the explanatory variables and the response, and can be adopted to predict correlation purposes [41]. The complex structure of these models obstructs wide-scale adoption [41]. However, they can be effectively utilized as a predictor and to judge the accuracy of the GEP based model [41].

In addition, experimental research is costly and requires abundant resources and time to carry out an accurate strength analysis. A slight mistake in computing the quantities and casting process may cause a malicious impact on the strength. Besides, the machine learning algorithm only requires an initial data set to efficiently predict the desired property. In this research, an effort has been made to address the limitation of the provision of the standard code by developing a GEP based empirical equation for short and slender CFST beam columns considering several input variables, i.e.,  $D$ : diameter of the tube,  $t$ : thickness of the tube,  $L$ : length of the tube,  $L/D$ : length to diameter ratio,  $e_t$ : eccentricity at the top face or loading face,  $e_b$ : eccentricity at the bottom face,  $f_y$ : yield strength of the tube and  $f_c$ : compressive strength of the infilled concrete. The schematic layout showing the input variables used to predict the capacity of CFST columns has been provided in Figure 1. The GEP algorithm delivers a simplistic empirical hand-based expression that can be used for future unseen data. The ANN and ANFIS algorithms are also used as a predictor to confirm the validity of the equation. A detailed and comprehensive database has been developed from peer-reviewed internationally published articles. This widespread database ensures the applicability of the model for new data. Statistical error checks are employed to verify the performance of the established models. In the end, a permutation feature analysis and a parametric study were also conducted to arrive at an accurate, reliable model.

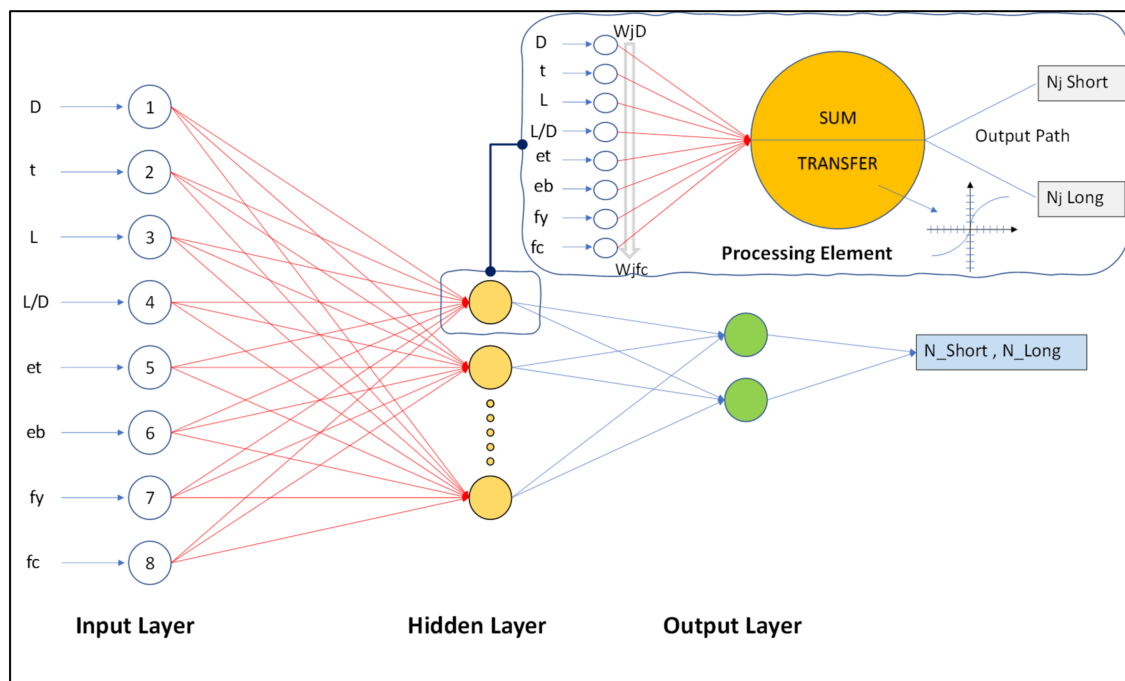


**Figure 1.** Schematic layout of circular concrete filled steel tube (CFST) column.

### 1.2. Detailed Description of Machine Learning Algorithms (ANN, ANFIS, GEP)

Artificial Neural Networks (ANNs) are computer algorithms that anticipate and classify the issues concerned with the effective processing of data [42]. As its name indicates, ANNs are based on mathematical models based on the human brain's neuron system [43]. ANNs have various layers of processing elements or nodes. Figure 2 shows the three layers with arranged nodes, i.e., input layer/s, output layer/s and hidden layer/s. The input layers have independent variables, output layers' target results, and hidden layers have concealed neurons/variables [44]. For each output ( $N_{st}/N_{lg}$ ), eight inputs were selected, while in the hidden layer the input parameters ( $D$ ,  $t$ ,  $L$ ,  $L/D$ ,  $e_t$ ,  $e_b$ ,  $f_y$  and  $f_c$ ) were multiplied by a suitable weight factor for connection. A threshold value at every node ( $\theta_j$ ) is added to the weighted input values after their summation. The resultant input ( $I_j$ ) is passed through the linear transfer function, which is called the transfer phase. The various activation transfer function (AFs) usually used in ANNs are the linear sigmoid, the stepped hyperbolic tangent and the logistic, among others [45]. An activation function is the key feature of a neural network which plays an important role in the artificial neural network model. It can be observed that these activated functions assist in appointing nonlinearity to the neural networks, due to which the selection of the appropriate activated function becomes very important [46]. Activated functions which have been used in the past include tangent hyperbolic and logistic sigmoid activated functions [47], the transcendental type parametric algebraic activated function [48], swish activated functions [49], and Multistate AF's to improve the DNN models [50] etc.





**Figure 2.** Schematic layout of feed-forward neural networks with eight explanatory variables.

In this research, the transfer functions used for the modelling of ANN models are TRANSIG and PURELIN. On the one hand, these transfer functions are capable of effectively increasing neurons in each layer and in each transfer function to improve the statistical indices of the training dataset. However, on the other hand, they also decrease the accuracy of the testing and validation datasets [51,52]. Dorofki et al. [53] observed that among various statistical functions, the performance of the Log-sigmoid transfer function was best, because these are differentiable, bounded and continuous. At the same time, Purelin TF gives much improved results. As a result, PE ( $N_{stj}$  or  $N_{lgj}$ ) is obtained as the resulting output, and the input of a PE is basically the output of the previous PE. For the hidden and output layer, every neuron utilizes the Logistic function (Equation (1)) as an activated function [54]. Moreover, the complete process can be observed from Equations (1)–(3).

$$f_h(z) = \frac{1}{1 + e^{-z}} \quad (1)$$

$$I_j = \left\{ \left( w_{jD} * D + w_{jt} * t + w_{jL} * L \dots w_{jfy} * f_y \right) \right\} + \theta_j; \text{ Summation} \quad (2)$$

$$N_{stj} \text{ or } N_{lgj} = f(I_j); \text{ Transfer} \quad (3)$$

To achieve an output with the least minimum error, the best combination of weights is achieved by adjusting the weight to the set rules at the time when the information from the input layer is passed by the ANN in the training stage. Another training set is used to validate the trained model. The method and implementation of ANN modeling is discussed in greater depth elsewhere and is beyond the scope of this review [55–57].

An attractive computation intelligence modelling tool, adaptive neuro-fuzzy inference scheme (ANFIS), blends the learning capabilities of ANNs with the reasoning capability of fuzzy logic. ANFIS has a better prediction potential and is a better alternative for computing nonlinear complex problems with greater precision [58]. With similar learning capability as ANN, ANFIS learns from training data containing a multiplex model and then gives the solutions in a fuzzy interface system (FIS) [43]. In MATLAB R2020b there is a tool called ANFIS that can train the input and output entities for the best connection between both the parameters. A basic FIS consists of several stages. First are feeding inputs to aid in the fuzzification of fuzzy sets according to the activation of linguistic rules. Then

basic laws are formulated by experts. These laws can also be derived from numerical results. Inference is the next step, which involves fuzzy mapping sets according to fixed laws. Finally, the fuzzy sets are defuzzified, which results in the final performance values.

The ANFIS technique is divided into five steps:

- (a) data collection,
- (b) ANFIS growth,
- (c) variables selection,
- (d) training and testing,
- (e) results

to express values in another way. In addition, the schematic layout of the ANFIS model for eight input variables ( $D, t, L, L/D, e_t, e_b, f_y$  and  $f_c$ ) is shown in Figure 3. The circle denotes the set nodes, while the square denotes the adaptive nodes. The two statements used for the presentation of the architecture of the ANFIS are IF-THEN statements are as follows.

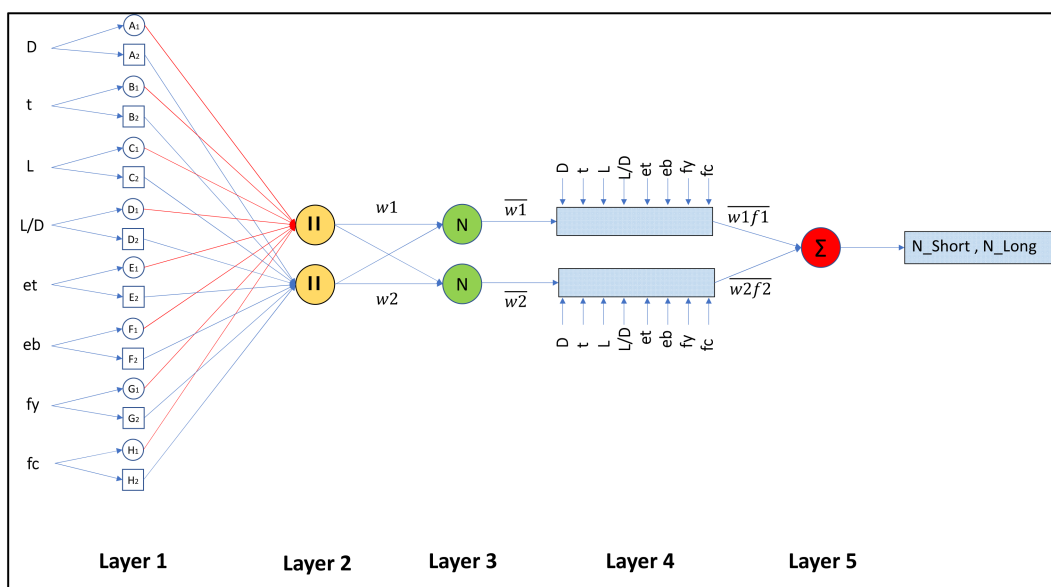


Figure 3. Sugeno ANFIS model schematic layout with eight explanatory variables.

Statement 1: IF ( $D$  is  $A_1$ ) and ( $t$  is  $B_1$ ) THEN,

$$\{f_1 = p_1(D) + q_1(t) + r_1\} \tag{4}$$

Statement 2: IF ( $D$  is  $A_2$ ) and ( $t$  is  $B_2$ ) THEN,

$$\{f_2 = p_2(D) + q_2(t) + r_2\} \tag{5}$$

where  $f_n$  denotes the fuzzy outputs ( $N_{st}, N_{lg}$ ) for the fuzzy inputs ( $D, t, L, L/D, e_t, e_b, f_y$  and  $f_c$ ), according to the fuzzy statement,  $A_i$  and  $B_i$  denote the fuzzy sets, and  $p_i, q_i$ , and  $r_i$  denote the arrangement elements determined in the training cycle.

An ANFIS model is made up of five layers [58], explained in detail below.

### 1.2.1. Layer 1

This layer is also called the fuzzification layer. The adaptive Pes provide outputs in Equations (6) and (7), which describe the fuzzy membership functions of the input model parameters and the original fuzzy rule foundation.

$$O_k^1 = \mu_{Ak}(D), \quad k = 1, 2 \tag{6}$$

$$O_k^1 = \mu_{Bk-2}(t), \quad k = 3, 4 \quad (7)$$

where  $\mu$  indicates the weight obtained by connecting the fuzzy membership function, and  $\mu_{Ak}(D)$  and  $\mu_{Bk-2}(t)$  differentiate the method of applying any fuzzy membership function. Equation (8) gives the  $\mu_{Ak}(D)$  for a bell-shaped membership function

$$\mu_{Ak}(D) = \frac{1}{1 + \left\{ \left( \frac{D-c_k}{a_k} \right) \right\}^{b_k}} \quad (8)$$

where  $a_k$ ,  $b_k$  and  $c_k$  are the factors affecting this membership function.

### 1.2.2. Layer 2

This layer's output is the preset rules' firing power for a given input pattern. The nodes in the second layer are constant and perform simple multiplication, with the output's parameters mentioned below (Equation (9)),

$$O_k^2 = w_k = \mu_{Ak}(D) \cdot \mu_{Bk}(t), \quad k = 1, 2 \quad (9)$$

### 1.2.3. Layer 3

Following the pattern of the second layer, the third layer also has fixed nodes for normalizing the firing strength of the previous layer. Equation (10) represents the output:

$$O_k^3 = \bar{w}_k = \frac{w_k}{w_1 + w_2}, \quad k = 1, 2 \quad (10)$$

### 1.2.4. Layer 4

In this layer, considering the first order Sugeno model, nodes are adaptive. Their outputs are represented as products of normalized firing intensity and first-order polynomial, with the first order Sugeno model taken into consideration. As a result, the output is given by (Equation (11)):

$$O_k^3 = \bar{w}_k f_k = \bar{w}_k \{p_k(D) + q_k(t) + r_i\} \quad (11)$$

### 1.2.5. Layer 5

In this layer, a fixed node summits the weighted magnitude of rules achieved from the prior layer, yielding Equation (12) as the model's output.

$$O^5 = \sum_{k=1}^2 \bar{w}_k f_k = \frac{\sum_{k=1}^2 w_k f_k}{w_1 + w_2} \quad (12)$$

It is worth noting that only the first and fourth layers of the ANFIS architecture are adaptive. In the first layer, the three adaptable parameters  $a_k$ ,  $b_k$ , and  $c_k$ , also known as premise parameters, are linked to input membership functions. Similarly, the three adaptable parameters  $p_k$ ,  $q_k$ , and  $r_k$ , also known as consequent parameters, are analogous to first-order polynomials and are found in the fourth sheet [59].

The gene expression programming approach, which is founded on Darwin's evolution theory and Mendel's genetic theory, is the most intellectually appealing computational knowledge algorithm [60,61]. There are two languages in GEP: (a) the gene's language, and (b) the expression tree's (Ets) language, and understanding one requires knowledge of the other's sequence or structure [62]. The following are the fundamental steps involved in traditional GEP modelling. A typical gene or chromosome contains two parts i-e:

- (a) Head consisting of function or terminal symbols
- (b) Tail containing only the terminal symbols.

The complexity of each parameter is represented by head size and the number of genes that control the number of sub-Ets.

Figure 4 shows how the chromosomes have set lengths that can be easily converted into an algebraic expression [63]. Every GEP gene has a series of words that are adapted from the function set; for example, arithmetic operations (+, −, ×, ÷), Boolean logic functions (AND, OR, NOT, etc.), trigonometric functions (cos, sin, ln), conditional functions (IF, THEN, ELSE), etc. [64].

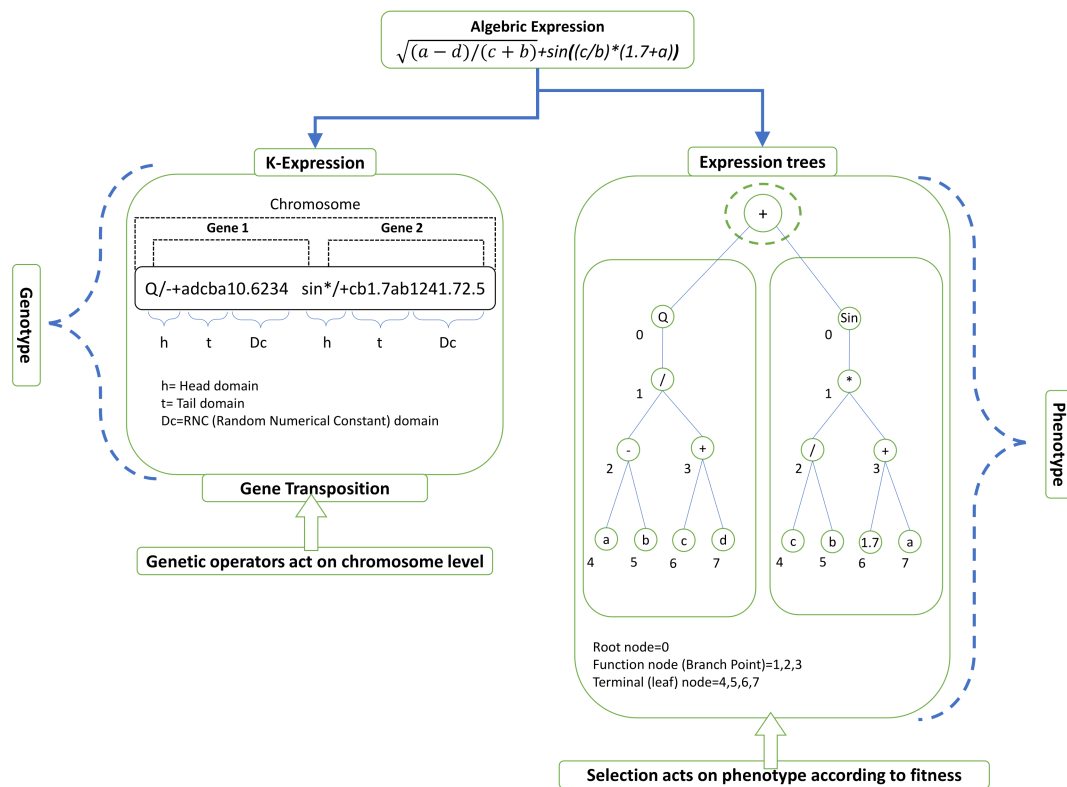


Figure 4. Illustration of a mathematical equation and its equivalent expression tree (ET).

The chromosomes are then expressed by ETs that come in a range of shapes and sizes. Then, in line with their percentages, the principal genetic operators of crossover, mutation, transposition, and recombination (1-point, 2-point, and gene recombination) are performed on the chromosomes [65]. Figure 4 illustrates a common expression tree (ET) and describes the crossover and mutation processes. Equation (13) also shows how the ET is expressed using Karva notation or a K-expression [36].

$$ET_{GEP} = \log\left(i - \frac{3}{j}\right) \quad (13)$$

When the stopping condition (the maximum number of generations or a satisfactory solution) is reached, the whole process is finished [66]. If the termination conditions for achieving the optimum iteration or the favorite fitness value are not satisfied, then the Roulette wheel procedure is used, which chooses the viable chromosomes of the first generation and moves them on to the next generation [67]. This method will be repeated for a certain number of generations or before the right solution is found [68].

### 1.3. The Aim of the Research

It follows from the above review that the variety of existing models does not allow rational predicting of the ultimate axial capacity of uniaxially loaded CFST columns. The article aimed to develop a prediction Multiphysics model for the circular CFST column by using Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Gene Expression Program (GEP).

## 2. Methods

### 2.1. Description and Division of Collected Data

A robust database is needed to successfully apply the machine learning algorithms (ANN, ANFIS, GEP). Thai et al. [6] collected the most recent experimental data of CFST columns and combined it with the existing database [69]. The Thai et al. [6] database is comprised of more than 3100 tests performed on CFST columns of different classes. A total of 1667 experimental results were extracted from the existing database, which comprised two different classes: circular short (702) and long (965) CFST columns loaded concentrically and eccentrically. The geometric features include physical dimensions such as length ( $L$ ), tube thickness ( $t$ ) and tube diameter ( $D$ ), and eccentricities at end supports ( $e_t$ ,  $e_b$ ). The material properties of steel and concrete include the yield stress ( $f_y$ ) and compressive strength ( $f_c$ ) of concrete. The concrete compressive strength obtained from the experimental tests collected from the literature was based on both available cylinder and cube specimens. Cube strength was converted to cylinder strength through related conversion factors. Furthermore, cylinder strength was used in the design equations [70] to avoid errors. Other material properties, such as steel and concrete moduli and steel ultimate stress, were considered of minor significance. In the case of concrete, the compressive strength for all the specimens is given, and the modulus is directly affected by this compressive strength, so there is no need to establish any relation between compressive strength and modulus of concrete, and only strength was incorporated in the model as a significant factor. A similar strategy was used to eliminate the need for steel's ultimate stress, while in the case of steel modulus, for example, 200 Gpa are probable, and this value is normal for all steel grades used in all columns.

For each class, descriptive statistics like distribution shape (kurtosis and skewness), central tendency (mean and median), dispersion of data (Standard deviations) and data extremities (maximum and minimum) for different geometric and material characteristics are provided in Table 1. For instance, the diameter ranges from 44.5 mm to 1020 mm for both classes, and the thickness ranges from 0.5 mm to 16.5 mm and 0.5 mm to 13.3 mm for long and short CFST columns, respectively. Similarly, the compressive strength range in filled concrete for both classes is from 7.7 MPa to 193.3 MPa, while that for yield strength of steel tube for long and short CFST columns is from 178.3 MPa to 853 MPa and from 185.7 MPa to 853 MPa, respectively. The range of  $L/D$  ratio is witness to the difference between short and long CFST columns. The reader is advised to note that these ranges surpass the design codes currently in use. As a result, this database can be used to build an ANN, ANFIS and GEP model with enhanced prediction capabilities that are more inclusive than codal provisions. It should be remembered that the magnitude eccentricity is a part of the affecting parameters in both classes. Therefore, the developed models can be confidently used for both axial and moment capacity of circular CFST columns. Skewness and kurtosis are related to the distribution of data. If the larger portion of the data for a particular variable is to the left of the mean, then this shows positive skewness (right tailed). Furthermore, if most of the data is to the right of mean value, then this shows negative skew (left tailed), and skewness is zero for perfectly symmetrical distribution (normal distribution). At the same time, the kurtosis indicates the heaviness of the tail related to the normal distribution. The leptokurtic or positive kurtosis dictates that the data is higher than the normal distribution, whereas the platykurtic or negative kurtosis reveals that the data is flatter than the normal distribution. The values of kurtosis and skewness for each input and output is also provided in Table 1. After constructing a reliable database, the available datapoints are divided into two sets, i.e., training and testing set [71].



**Table 1.** Descriptive statistical analysis of explanatory variables and response.

Category	Parameters	Mean	Median	Max	Min	S.D.	Kurtosis	Skewness
Long	<b>Inputs</b>							
	$D$ (mm)	147.2	121.0	1020.0	44.5	89.9	31.78	4.38
	$t$ (mm)	4.4	4.0	16.5	0.5	2.4	5.12	1.79
	$L$ (mm)	1438.3	1040.0	5560.0	152.3	1094.5	1.12	1.23
	$L/D$	11.2	8.6	51.5	0.8	8.9	1.88	1.35
	$e_t$ (mm)	13.0	0.0	300.0	0.0	28.1	25.87	4.08
	$e_b$ (mm)	11.2	0.0	300.0	0.0	27.5	29.50	4.38
	$f_y$ (MPa)	332.1	322.0	853.0	178.3	81.7	7.83	1.98
	$f_c$ (MPa)	46.6	40.1	193.3	7.7	26.8	7.75	2.36
	<b>Output</b>							
$N_{exp}$ (kN)	1616	848.5	46,000	45.2	3181.1	73.86	7.53	
Short	<b>Inputs</b>							
	$D$ (mm)	169.2	133.1	1020.0	48.0	112.5	23.19	4.17
	$t$ (mm)	4.2	4.0	13.3	0.5	2.3	1.56	1.15
	$L$ (mm)	498.7	399.5	3060.0	152.3	334.0	24.32	4.16
	$L/D$	3.0	3.0	4.0	0.8	0.6	0.19	−0.55
	$e_t$ (mm)	2.8	0.0	105.0	0.0	10.9	30.38	5.03
	$e_b$ (mm)	2.8	0.0	105.0	0.0	10.9	30.38	5.03
	$f_y$ (MPa)	336.8	322.7	853.0	185.7	97.5	10.53	2.52
	$f_c$ (MPa)	58.8	46.6	193.3	7.7	35.9	2.60	1.54
	<b>Output</b>							
$N_{exp}$ (kN)	2782.5	1678.1	46,000	199.9	4304.5	39.20	5.39	

Min. Minimum; Max. Maximum; S.D. standard deviation.

## 2.2. Structure of ANN, ANFIS and GEP Models

The specification of the significantly affecting input parameters is the first step in designing the appropriate model. The  $N_{st}$  and  $N_{lg}$  were found to be dependent on the following factors (Equation (14)):

$$N_{st} \text{ or } N_{lg} \text{ (kN)} = f(D, t, L, \frac{L}{D}, e_t, e_b, f_y, \text{ and } f_c) \quad (14)$$

Here,  $D$  is the diameter of the tube,  $t$  is the thickness of the tube,  $L$  is the length of the tube,  $L/D$  is the ratio between length and diameter of the tube,  $e_t$  and  $e_b$  is eccentricity at top and bottom face,  $f_y$  is the yield strength of tube, and  $f_c$  is the compressive strength of the tube.

Both ANN and ANFIS simulations were performed in the MATLAB R2020b environment using the neural network and fuzzy logic toolbox, respectively. The 702 experimental records of short CFST columns were randomly divided into 70% training (495 datapoints) and 30% testing (207 datapoints), and similarly for second class, i.e., long CFST columns, out of a total of 965 experimental records, 70% (676 datapoints) and 30% (289 datapoints) were accumulated in the training and testing set, respectively [36]. The training accuracy and time taken to train the model are essential [72] for comparison of the performance of each model. The input layer in this analysis had eight input nodes, one for each of the model inputs ( $D, t, L, L/D, e_t, e_b, f_y, f_c$ ), and the output layer had  $N_{st}$  and  $N_{lg}$  for ANN. After using the Levenberg-Marquardt algorithm and choosing random data division, the number of hidden neurons was set to ten. In addition, the network form was chosen as feed-forward back-propagation. To achieve a better output at the required number of hidden layers, trial and error methods should be used [54]. Table 2 lists the statistical parameters of modelling for ANNs in this research.

**Table 2.** The setting of different parameters of ANN, ANFIS, and GEP model used in current research.

Parameters	Class and Value	
	$N_{lg}$	$N_{st}$
Training dataset (70%)	676	495
Testing dataset (30%)	289	207
<b>ANN</b>		
Network type	Feed-forward back-propagation	
Data division	Random (un-biased)	
No. of hidden layer	8	
No. of hidden neurons	10	
Training algorithm	Levenberg-Marquardt	
Hidden layer's Transfer function	TANSIG	
Output layer's Transfer function	PURELIN	
No. of non-linear parameters	16	
No. of epochs	40	
Learning rate	0.01	
<b>ANFIS</b>		
No. of linear parameters	72	65
No. of nonlinear parameters	140	120
Total No. of parameters	176	154
No. of fuzzy rules	5	8
No. of MFs	5	8
No. of nodes	20	45
No. of Training epoch	30	30
Training error goal	0	0
Membership Function type	Trimf	
Fuzzy structure	Sugeno	
Type of FIS	Sub clustering	
Method of Optimization	Back propagation and least square	
Output function	Linear	
<b>GEP</b>		
Parameters		
General		
Number of chromosomes	100	
Number of Genes	3	
Head size	8	
Linking function	Addition	
Function set	+, −, ×, ÷	
Numerical constants		
Constant per gene	10	
Type of data	Floating number	
Maximum complexity	8	
Ephemeral random constant	[−10,10]	
Genetic operators		
Rate of mutation	0.00138	
Inversion rate		
IS transposition rate	0.00546	
RIS transposition rate		
One-point recombination rate		
Two-point recombination rate		
Gene recombination rate	0.00277	
Gene transposition rate		

ANFIS provides only a single output; unlike ANNs, the outputs were handled separately using the same set of input parameters in both ANN and ANFIS models. The sub-cluster FIS was first generated because the database contained a huge number of data points, in which subtractive clustering with a hybrid optimization technique (least

square and back-propagation technique) was used to train the FIS by building a triangular membership function (trimf) [73]. Furthermore, Venkatesh and Bind [42] also recommend using the grid portioning approach in which the maximum number of input parameter is taken as six. The various setting parameters for the training phase are presented in Table 2.

GeneXproTools version 5.0 was used to implement the GEP algorithm [74,75]. From data loading to code generation, GeneXproTools is a versatile data processor that supports categorical variables and missing values, greatly enhancing the performance and accuracy of the modelling process [76]. It can generate multiple models from large heterogeneous data and locate code in programming languages such as MATLAB, C++, and Visual Basic [54]. The parameters in the GEP algorithm were finalized based on previous literature guidelines and several initial runs [41,77]. The initial optimum combination of GEP parameters was calculated using the trial-and-error method. The effect of a single GEP parameter on prediction accuracy was then investigated using the optimum combination of parameters. Finally, the proposed model was formulated using the finalized optimum combination of GEP parameters for obtaining basic mathematical expressions to predict the bearing capacity of CFST columns. The model's complexity increases as the number of chromosomes, head size, and genes increase, and hence the length of the running program is determined.

Furthermore, multiple evolved models are strongly influenced by head size and genes. The error in the formulated model is reduced by the higher values of these setting parameters, which results in a higher value of the coefficient of regression. Table 2 shows the unique values for the parameters in the GEP algorithm for both classes of CFST columns considered in this study.

### 2.3. Evaluation of Models through Statistical Measures

Statistical performance for the  $N_{st}$  and  $N_{lg}$  prediction (ANN, ANFIS and GEP models) were measured using five standard statistical metrics, including correlation coefficient ( $R$ ), determination coefficient ( $R^2$ ), root mean square error (RMSE), mean absolute error (MAE), relative squared error (RSE), mean absolute percent error (MAPE) and relative root mean square error (RRMSE) in the training testing sets [77–79]. In addition, for all the proposed models, a performance index ( $PI$ ) has been calculated as another metric, ruled primarily by RRMSE and  $R$  [36]. Equation (15) to Equation (21) define these performance measures:

$$MAE = \frac{\sum_{i=1}^n |p_i - q_i|}{n} \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - q_i)^2}{n}} \quad (16)$$

$$RSE = \frac{\sum_{i=1}^n (p_i - q_i)^2}{\sum_{i=1}^n (\bar{p}_i - p_i)^2} \quad (17)$$

$$RRMSE = \frac{1}{|\bar{e}|} \sqrt{\frac{\sum_{i=1}^n (p_i - q_i)^2}{n}} \quad (18)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{q_i - p_i}{q_i} \right| \times 100 \quad (19)$$

$$R = \frac{\sum_{i=1}^n (p_i - \bar{p}_i)(q_i - \bar{q}_i)}{\sqrt{\sum_{i=1}^n (p_i - \bar{p}_i)^2 \sum_{i=1}^n (q_i - \bar{q}_i)^2}} \quad (20)$$

$$PI = \frac{RRMSE}{1 + R} \quad (21)$$

Here,  $p_i$  and  $q_i$  are the  $i$ th predicted and expected outcomes, respectively,  $\bar{p}_i$  and  $\bar{q}_i$  are the average of predicted and expected outcomes, and  $n$  is the total number of experiments.

To determine the relative correlation between models and experimental outputs ( $p_i$  and  $q_i$ ), the performance of  $R$  is used. When  $R > 0.8$ , the predicted and expected values are highly correlated [80]. However,  $R$  is insensitive to the division and multiplication of outcomes [77]. Thus,  $R^2$  was used because of its impartial assessment and comparatively better performance.  $R^2$  values equal to 1 and closer to each other demonstrate that much of the model's variation between input parameters was used [81]. In addition, RMSE is a common metric since significant errors in comparison to smaller errors are resolved very effectively. RMSE closer to 0 indicates that the prediction error is negligible [58]. It does not, however, ensure optimum efficiency in any conditions. MAE was also calculated and is enormously advantageous in the presence of smooth and continuous data [82]. To sum up, a greater value of  $R$  and smaller values for RMSE, MAE, RSE, and RRMSE provides a better standard calibration for model performance. In addition, Gandomi et al. [83] proposed that  $PI$  ranges from 0 to infinity and closer to zero indicate a good model performance.

In a range of machine learning techniques, due to unnecessary data training, the models appear to overfit [84] and lead to lower training error values and greater testing error values. In order to choose the best predictive model that can solve overfitting, a minimized objective function ( $OF$ ) is used, as shown in Equation (22) [36,85]:

$$OF = \left( \frac{n_T - n_v}{n} \right) PI_T + 2 \left( \frac{n_v}{n} \right) PI_v \quad (22)$$

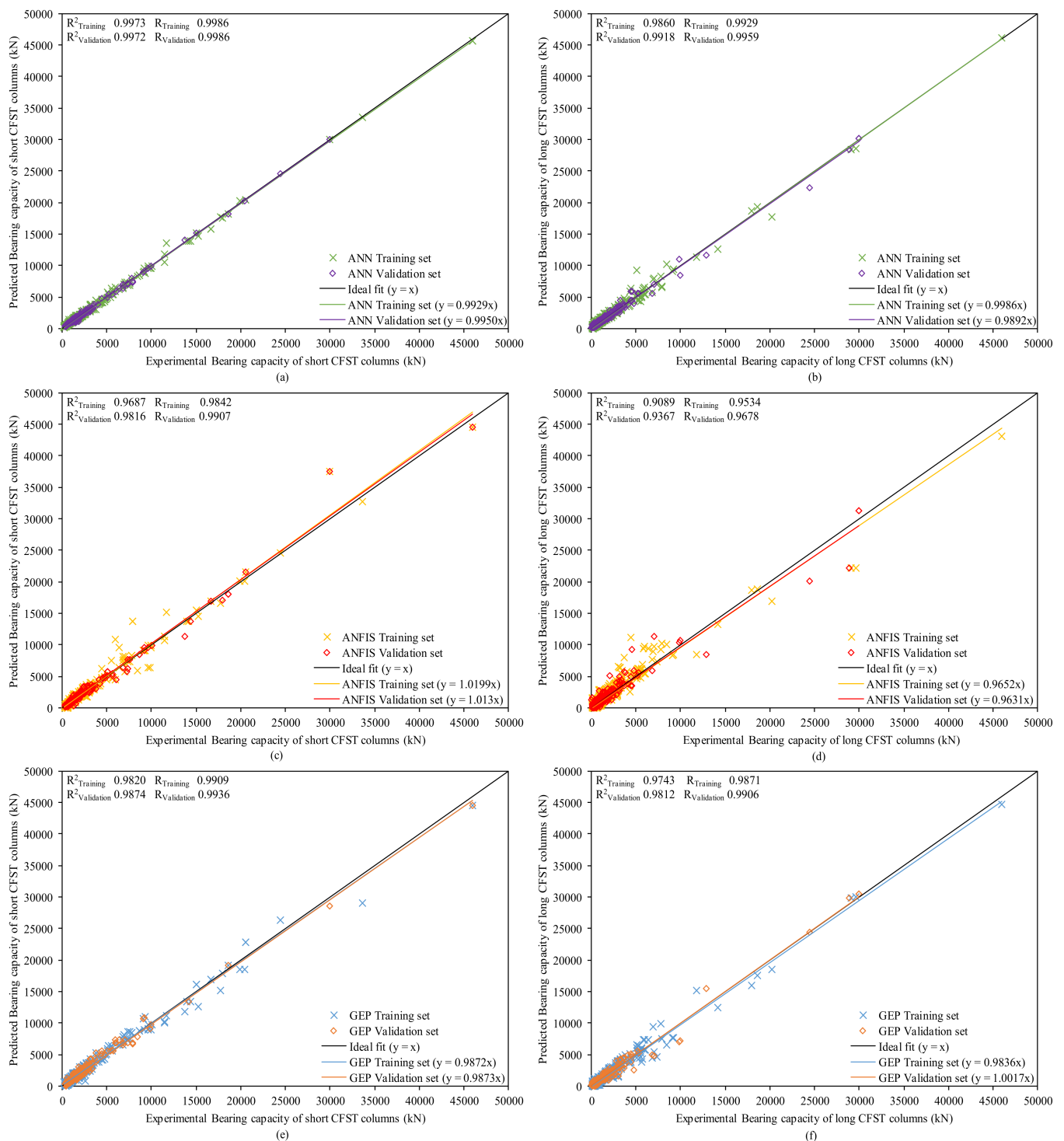
Here  $T$  and  $V$  (subscripts) correspond to training and testing set points, and  $n$  represents the total number of records. The better-predicted formula must have a lower OBF value (nearly equals to 0), as the consequence of  $R$ , RRMSE and proportional percentage of dataset records are considered. Eight different suitable parameter combinations were executed, and the least  $OF$  was chosen in this study.

### 3. Results and Discussion

#### 3.1. Regression Analysis of ANN, ANFIS and GEP Model

The regression plots between actual and predicted bearing capacities for short and long CFST columns of ANN, ANFIS and GEP models are clearly presented in Figure 5. The equation for the slope of the regression lines between actual and predicted bearing capacity for both training data and testing data has been shown in all plots. The closeness of the datapoints near the regression line drawn at  $45^\circ$  indicates the better performance of the established models [58]. The slope should be closer to 1 for an ideal fit [77,86] and for a strong correlation. For all three proposed models, a strong correlation can be seen as depicted from the slope of the regression lines. The values of slope for both stages, i.e., training and testing set, are quite similar in all the proposed models showing that the models are efficiently trained and hold a high generalization capacity. The spread of the datapoints in the training and testing set also shows that the issue of overfitting has been diminished.

Moreover, the coefficient of determination ( $R^2$ ) for all the proposed models is greater than 90% in both training and testing stages following the trend:  $R^2(\text{ANN}) > R^2(\text{GEP}) > R^2(\text{ANFIS})$ , reflecting the shortcoming of neural and fuzzy arrangement in the projected ANFIS model. The mean correlation coefficient ( $R$ ) in the projected models for  $N_{st}$  is also maximum for ANN (0.9986) tracked by GEP (0.9922) and ANFIS (0.9874). For a stronger correlation, the  $R$ -value will be higher (i.e.,  $R > 0.8$ ) for an acceptable model [77,87,88]. In the case of  $N_{st}$ , the  $R^2$  for ANN model is highest, i.e., 99.73% and 99.72% for training and testing data, respectively, and 98.20% and 98.74% for the GEP model, respectively.



**Figure 5.** Regression plot between actual and predicted bearing capacity of (a) ANN model for short CFST columns (b) ANN model for long CFST columns (c) ANFIS model for short CFST columns (d) ANFIS model for long CFST columns (e) GEP model for short CFST columns and (f) GEP model for long CFST columns.

### 3.2. GEP Based Formulation of Bearing Capacity of CFST Columns

The two GEP based empirical formulae for future prediction of bearing capacity of short and long circular CFST columns are derived using the GEP algorithm. The sub-ETs' links to these formulae for  $N_{st}$  and  $N_{lg}$  use four basic mathematical operations, i.e., +, −, × and ÷, as presented in Figure 6. The ETs in Figure 6a,b corresponding to three



different numbers of genes with addition used as a linking function, are decoded to derive a respective mathematical equation for  $N_{st}$  and  $N_{lg}$  as shown in Equations (23) and (24). Based on the total number of records, the projected formulae are in close agreement with standard limits for an ideal model and can be confidently and reliably used for the prediction of bearing capacity of short and long circular CFST columns [64,89,90].

$$N_{st}(kN) = \left( \left( \frac{-18.9L+11525.98}{f_c} \times \frac{L}{D} \right) - 1711.5 + f_y - D + (18.94 \times D) \right) + \left( f_c \times \left( \frac{D+L}{-0.38 \times f_c} + \frac{-45.67+D}{5.42} \right) - e_t \right) + \left( t \times \left( L + \frac{f_y}{5.48} - (3.07 \times e_b) - \left( \frac{L}{D} \times 88.45 \right) \right) \right) \tag{23}$$

$$N_{lg} = ((f_c + D + 9.55 + (7.1 \times f_c)) + (t \times 225.8)) + \left( \left( \frac{f_y-L}{f_c} - e_b - \frac{L}{D} - 11.82 + D \right) \times t \right) + \left( \left( t + \frac{t}{4.03 + 0.54 + \frac{L/D}{f_c}} \right) \times D \right) \tag{24}$$

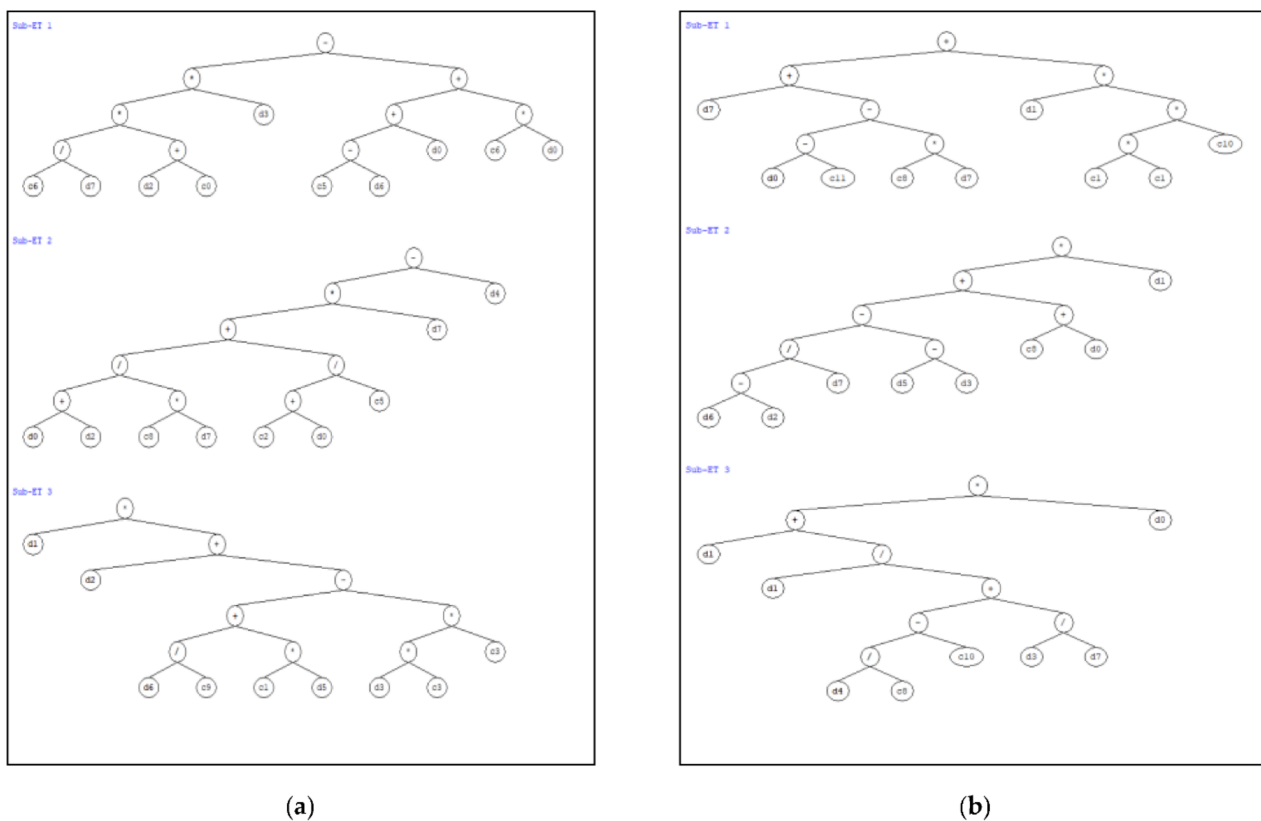


Figure 6. Representation of expression trees (a) Short circular CFST (b) Long circular CFST.

### 3.3. Performance Evaluation of Proposed Models Using Statistical Indicators

The reliability and accurateness of the models greatly depend on the number of data points [77]. In this research, a maximum number of records, i.e., for short CFST columns 495 training and 207 testing records and for long CFST columns 676 Training and 289 testing records, are used for the development of models and, hence, better accuracy has been achieved. It is suggested in the literature that, for a reliable model, the ratio between several records and input variables in both training and testing should be at least 5 [91]. In the training and testing stage of this study, the specified ratio is far beyond the limit, i.e., equal to 61.88 and 25.88, respectively, for  $N_{st}$  and 84.5 and 36.13, respectively, for  $N_{lg}$ .

The studies suggested that  $R$  or  $R^2$ , enumerates the linear dependency of response and explanatory variables. An acceptable value of  $R$  greater than 0.8 shows a strong

correlation between actual and predicted values [92]. Thus, the evaluation of the proposed models based on the slope of the regression line and regression or correlation coefficient is insufficient [92]. Therefore, the developed models are also assessed using different statistical metrics for evaluating their robustness.

### 3.3.1. ANN Model

It can be seen from Figure 7, that for short CFST columns, the R exceeds 0.90 and is nearly equal to 1, which shows that the ANN model has a better prediction capability for training and testing data sets that is perfectly equal to 0.9986 for training and testing both sets, which witnessed the outburst performance of the ANN model. Similarly, in the case of long CFST columns, these are 0.9929 and 0.9959 for the respective datasets. To interpret the statistics of MAPE, the absolute percent error plot and the error histogram of percent error are shown in Figure 7c,d. The percent error histograms indicate 58% and 36% ANN predicted values for  $N_{st}$  and  $N_{lg}$ , respectively, have an error less than 10%. As reflected in Figure 7a,b, the error values are scattered near zero, showing the outburst performance of the developed multi-physics-based ANN model. Furthermore, both the MAE and RMSE in Table 3, enumerate the magnitude of average error values and have their own importance. The RMSE squared the error before average and gave more weightage to larger error values [93]. At the same time, MAE gives low weightage to larger error values and is always lower than RMSE [80,93]. For the ANN model for  $N_{st}$ ,  $MAE_{testing}$  (145.96) is lower than  $RMSE_{testing}$  (193.20), satisfying the stated condition. Similarly, these values for  $N_{lg}$  are (196.22) and (325.12), respectively. The  $RSE_{testing}$  of ANN models for  $N_{st}$  and  $N_{lg}$  is also minimal and nearly equals zero, i.e., (0.00274) and (0.0086), respectively. The details of the statistical analysis of all the proposed models for  $N_{st}$  and  $N_{lg}$  in both stages (training and testing), are provided in Table 3.

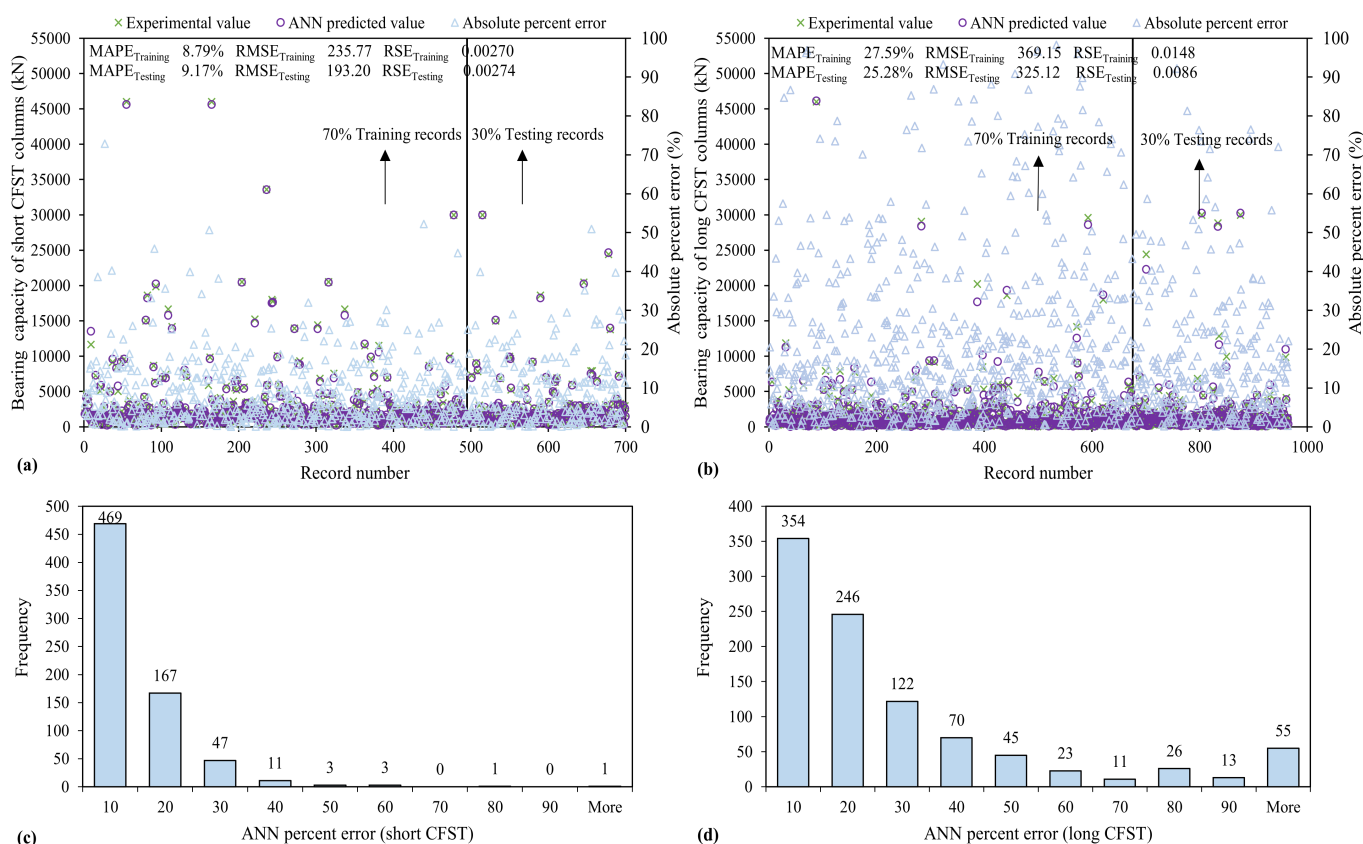


Figure 7. Variation of mean absolute error and error histograms of bearing capacity established using ANN algorithm (a,c) short CFST (b,d) long CFST columns.

**Table 3.** Statistical indicators for ANN, ANFIS and GEP models developed for bearing capacity of short and long circular CFST columns.

Model	Statistical Metrics	ANN		ANFIS		GEP	
		Training	Testing	Training	Testing	Training	Testing
Long	MAE	214.98	196.22	556.86	500.30	306.34	290.36
	MAPE	27.59	25.28	36.29	42.91	40.50	39.74
	RSE	0.0148	0.0086	0.0931	0.0635	0.0256	0.0195
	RMSE	369.15	325.12	925.33	882.60	485.46	489.06
	R	0.9929	0.9959	0.9534	0.9678	0.9871	0.9906
	RRMSE	0.2330	0.1922	0.5841	0.5219	0.3064	0.2892
	PI	0.1169	0.0963	0.2990	0.2652	0.1542	0.1452
	OF	0.1000		0.2700		0.1500	
Short	MAE	155.29	145.96	360.05	328.73	350.56	387.93
	MAPE	8.79	9.17	24.55	27.20	22.39	20.15
	RSE	0.00270	0.00274	0.0345	0.0197	0.0179	0.0126
	RMSE	235.77	193.20	750.68	681.67	552.17	527.47
	R	0.9986	0.9986	0.9842	0.9907	0.9909	0.9936
	RRMSE	0.0833	0.0273	0.2824	0.2213	0.2023	0.1812
	PI	0.0416	0.361	0.1423	0.1111	0.1016	0.0908
	OF	0.2300		0.1200		0.090	

### 3.3.2. ANFIS

The ANFIS model also delivers a good result based on performance criteria. Along with high values of  $R$  exceeding 0.90, lower error statistics were recorded. Unlike ANN and GEP, the magnitude of errors (MAE, RMSE and RSE) is higher. The percent error histogram and distribution of MAPE is shown in Figure 8, indicating that 46% and 27.3% ANFIS predicted values for  $N_{st}$  and  $N_{lg}$ , respectively, have percent error less than 10%.

In comparison with ANN models, it gives more high-error values. However, as presented in Figure 8, the absolute percent error runs near the axis showing that the overall performance is satisfactory. As stated in Table 3, the  $MAE_{testing}$  and  $RMSE_{testing}$  of ANFIS models is (55.6)% and (71.7)% greater than that of ANN models in the case of  $N_{st}$  and (60.8)% and (63.2)% greater in the case of  $N_{lg}$ , respectively. Like the ANN model, the  $RSE_{testing}$  for ANFIS models is also near to 0. Thus, based on the above facts, the prediction of bearing capacity of short and long CFST columns can also be obtained through these reliable and accurate models.

### 3.3.3. GEP Model

The GEP model provides better results than the ANFIS model but is worse than the ANN model based on  $R$  or  $R^2$  and the magnitude of error statistics. However, as shown in Figure 9, around 42% and 27% (nearly equal to the results deduced for the ANFIS model) of the GEP predicted values for  $N_{st}$  and  $N_{lg}$  have percent error below 10%, which are less than the ANFIS model. In the GEP model, the RMSE and RSE values for the testing set are lower than the ANFIS model by 22.6% and 36% for  $N_{st}$ , respectively, and 44.6% and 69.3% lower for  $N_{lg}$ , respectively.

### 3.4. Comparison of Models Using External Testing Criteria

The literature recommended an RRMSE of between 0 and 0.1 for an excellent model, and for a good model between 0.1 and 0.2. As shown in Table 3, following RRMSE, the performance of the three proposed models for both classes can be categorized as ANN followed by GEP and ANFIS. Gandomi and Roke [36] also classified the machine learning models as Good, if the  $PI$  and  $OF$  are less than 0.2 ( $OF$  nearly equal to 0 denotes an ideal model). The  $PI$  and  $OF$  values of GEP models for the  $N_{st}$  and  $N_{lg}$  fall within the prescribed limit and show an outstanding performance. However, the  $PI$  in the training and testing stage of the ANFIS model for  $N_{lg}$  are equal to 0.2990 and 0.2652. Its  $OF$  value is also equal to

0.2700. In the ANN model for  $N_{st}$ , the  $PI$  in the testing stage is 0.361, which is considerably higher than the prescribed limit. Consequently, the  $OF$  for the model is 0.2300. Thus, the performance of the ANN and ANFIS model based on  $PI$  and  $OF$  is ambiguous and marked as satisfactory, although the ANFIS model for short and ANN model for long CFST columns perform well. Lewis [94] suggested that the model can be categorized as either excellent prediction ( $MAPE \leq 10\%$ ), good prediction ( $10\% < MAPE \leq 20\%$ ), acceptable prediction ( $20\% < MAPE \leq 50\%$ ) and inaccurate prediction ( $50\% < MAPE$ ) [79,95]. In accordance with the mentioned criteria of model categorization based on MAPE, the ANN provides excellent forecasting results for  $N_{st}$ , while all other models including  $N_{lg}$ -ANN falls in the “acceptable prediction” category.

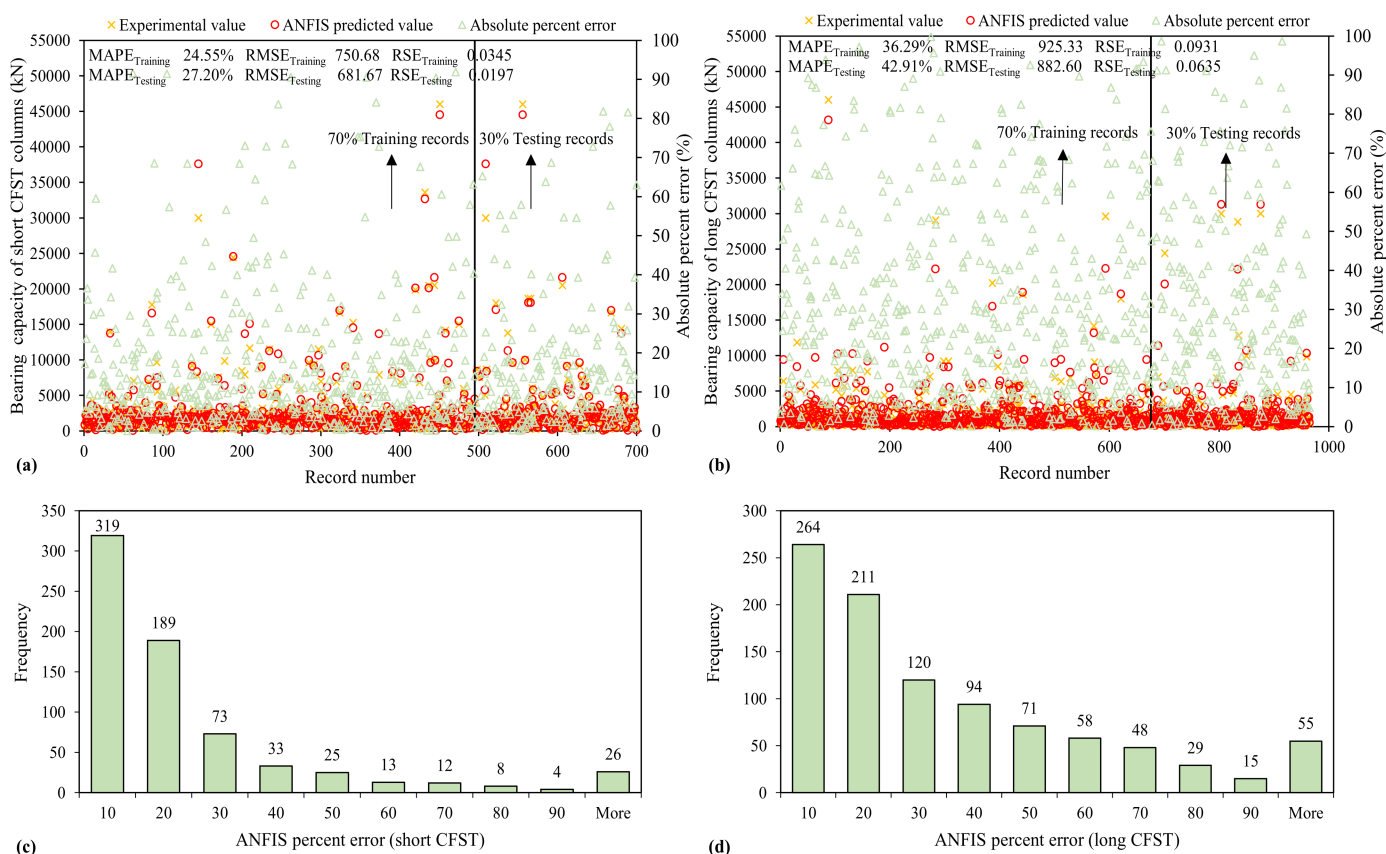
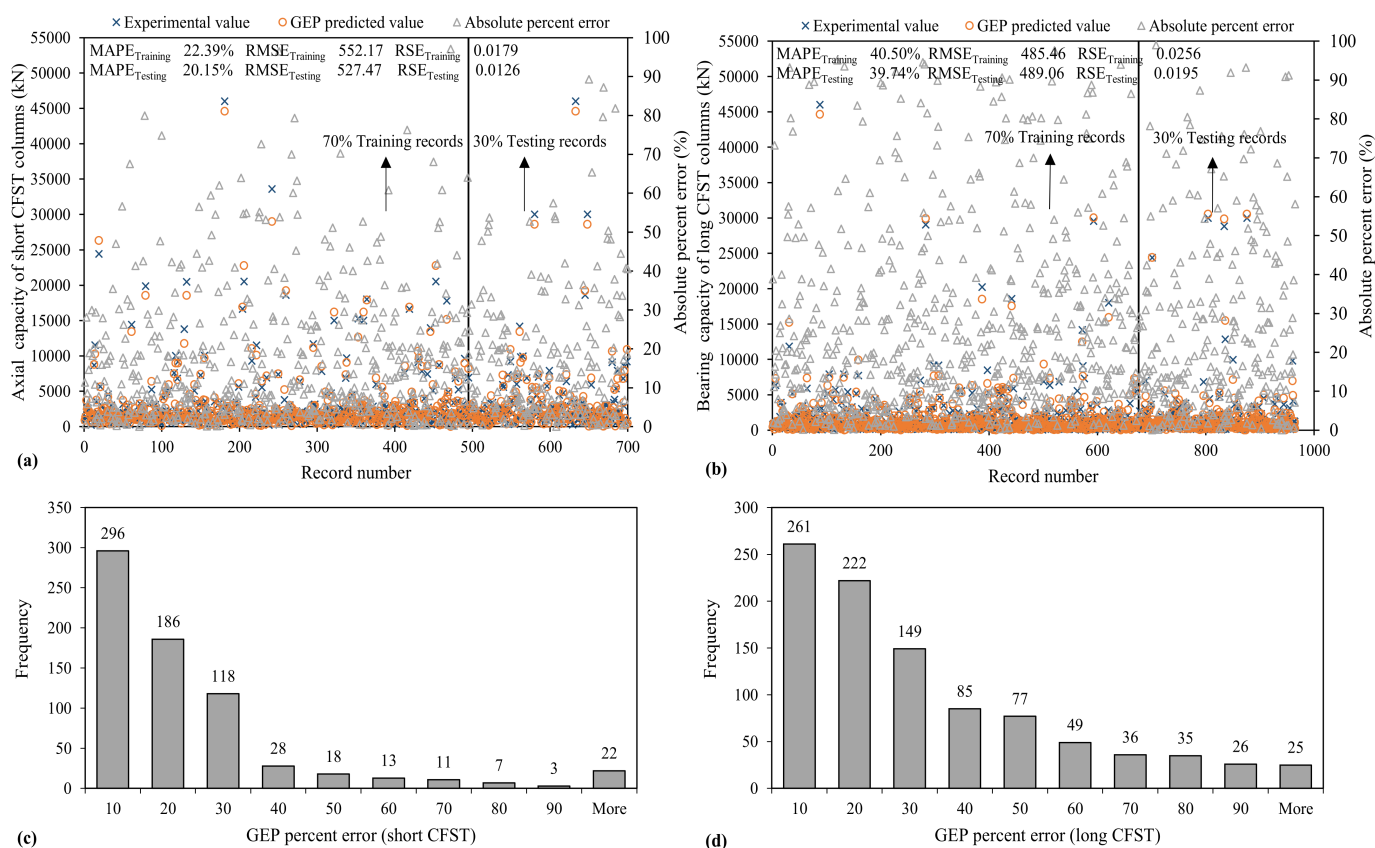


Figure 8. Variation of mean absolute error and error histograms of bearing capacity established using ANFIS algorithm (a,c) short CFST (b,d) long CFST columns.

The GEP suggested models for forecasting bearing capacity surpasses the other two models since this method provides simplified expressions, presented as Equations (23) and (24). With these formulas, the overall time needed for both the  $N_{st}$  and  $N_{lg}$  tests utilizing the corresponding GEP models is significantly quicker considering the time needed by the traditional test technique [41]. Thus, the suggested mathematical formulas provide a feasible fast method for finding  $N_{st}$  and  $N_{lg}$ .

Furthermore, several other external testing criteria for the GEP model were suggested in the literature and are presented in Table 4. Mollahasani, Alavi [96] stated that at least one of the regression slopes lines ( $k'/k$ ) moving through the origin must reach one. The results of performance indicators ( $m$  and  $n$ ) must not be above 0.1. The criterion of third external testing presented in this research was given by [97], which implies an  $Rm > 0.5$  and in this research, this is fulfilled by both GEP models. Likewise, the square correlation coefficient amongst predicted and actual data ( $R_0^2$ ), and amongst experimental and actual data ( $R_0'^2$ ) must be nearer to 1 [76]. It can be observed in Table 4 that the suggested GEP

models satisfy all the necessary criteria, thus indicating the better level of accurateness of both the models.



**Figure 9.** Variation of mean absolute error and error histograms of bearing capacity established using GEP algorithm (a,c) short CFST (b,d) long CFST columns.

**Table 4.** Assessment of GEP model using external testing criterion suggested in the literature.

Equation	Condition	GEP Model	
		Long	Short
$k = \sum_{i=1}^n \frac{(q_i \times p_i)}{q_i^2}$	$0.85 < k < 1.15$	0.989	1.00
$k' = \sum_{i=1}^n \frac{(q_i \times p_i)}{p_i^2}$	$0.85 < k' < 1.15$	0.995	0.974
$R_m = R^2 (1 - \sqrt{ R^2 - R_0^2 })$	$0.5 < R_m$	0.847	0.877
$R_0^2 = 1 - \frac{\sum_{i=1}^n (p_i - q_i^0)^2}{\sum_{i=1}^n (p_i - p_i^0)^2}$	$R_0^2 \cong 1$	0.999	0.999
$R_0'^2 = 1 - \frac{\sum_{i=1}^n (q_i - p_i^0)^2}{\sum_{i=1}^n (q_i - q_i^0)^2}$	$R_0'^2 \cong 1$	0.979	0.958
$q_i^0 = k \times p_i$			
$p_i^0 = k \times q_i$			

### 3.5. Sensitivity and Parametric Study of GEP Models

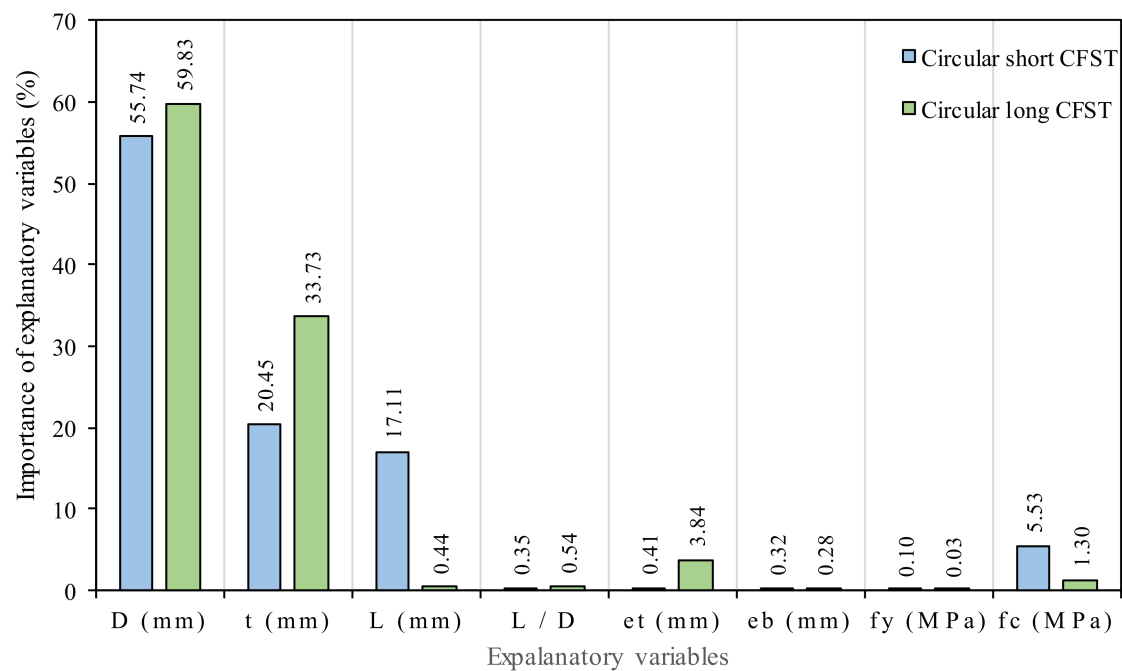
Sensitivity analysis shows how an output of a proposed model is sensitive to a certain change in its input parameters. It ranks parameters by capturing the behavior of the model in response to the changes in a particular parameter and indicates the effectiveness of each parameter [54,72,98]. The comparative impact of the input variables included in this research on the short and long circular CFST columns is evaluated by executing sensitivity analysis (SA) on the GEP models utilizing Equations (24) and (25):

$$N_i = f_{max}(y_i) - f_{min}(y_i), \tag{25}$$



$$SA (\%) = \frac{N_i}{\sum_{j=1}^n N_j} * 100. \quad (26)$$

Here,  $f_{max}(y_i)$  is the maximum Model projected output for the  $i$ th input parameter and  $f_{min}(y_i)$  is the minimum model projected output for the  $i$ th input parameter, while the rest of the input parameters are kept as 1. The relative importance of input variables on the bearing capacity of short and long circular CFST columns is graphically displayed in Figure 10, which shows that the diameter, thickness, and length are the most sensitive parameters in short columns with a value of relative contribution of 55.7%, 20.45% and 17.1%, respectively, while in the case of slender columns, only diameter and thickness are sensitive parameters with a value of relative contribution of 59.8% and 33.7%, respectively. The sensitivity analysis also reveals that the influence of strength of steel ( $f_y$ ), and eccentricity at bottom ( $e_b$ ) parameters is negligible for both models with relative contribution of 0.09%, 0.03%, 0.3% for short and 0.03%, 0.3%, 0.5% for long CFST columns, respectively. The impact of length ( $L$ ) in long columns and eccentricity at the top ( $e_t$ ) in short columns is also less, with a relative contribution of 0.44% and 0.41%, respectively. According to [99], the bearing capacity of CFST columns is mainly governed by the diameter and thickness of the steel tube.



**Figure 10.** Relative importance of input variables on the bearing capacity of short and long circular CFST columns.

Secondly, parametric analysis is performed to verify the efficiency of input variables and the strength of the GEP model. This is accomplished by changing one parameter with a particular increment while keeping all other parameters constant at their mean values. The main characteristics of the input parameters of CFST columns are material and geometric properties. Figure 11 indicates the prediction ability of GEP models for simulation of  $N_{st}$  and  $N_{lg}$  with varying input variables, i.e.,  $D$ ,  $t$ ,  $L$ ,  $L/D$ ,  $e_t$ ,  $e_b$ ,  $f_y$ , and  $f_c$ . It is well understood that  $D$  and  $t$  are important factors controlling the bearing capacity of CFST columns. The capacity of both short and long columns follows a degree polynomial curve with variation diameter and thickness of the tube. Similar trends were also noticed for the compressive strength and yield strength of the steel tube. Increasing the compressive strength of concrete will divert the failure control mode to the yield strength of the steel tube and vice versa. The strength of the concrete inside the CFST is responsible for the stiffness of the CFST columns [100]. Stiffness rises along with concrete compressive strength, yet

columns fracture owing to concrete crushing and brittle behavior when filled with high strength concrete. However, regardless of the length to diameter ratio, an increase in concrete core strength enhances the strength of filled columns to a greater extent. Linear decreasing pattern was observed for variation in the length or length to diameter ratio of the steel tubes with the prescribed limits. The effect of the eccentricity at the top or bottom face of the column also adversely affecting the capacity of CFST columns. Compared to zero eccentricity, in the case of eccentric loading the contact stresses will not be distributed non-uniformly, causing outward buckling [101]. Moreover, changes in the diameter and thickness of the steel tube greatly influence the bearing capacity of CFST columns for  $N_{lg}$  and  $N_{st}$ , as observed and stated by many researchers in the past [6].

Hence the results of the current study are similar to those found in past research studies accumulated in the database [6]. The parametric analysis also effectively captures the input parameters for the prediction of  $N_{lg}$  and  $N_{st}$ .

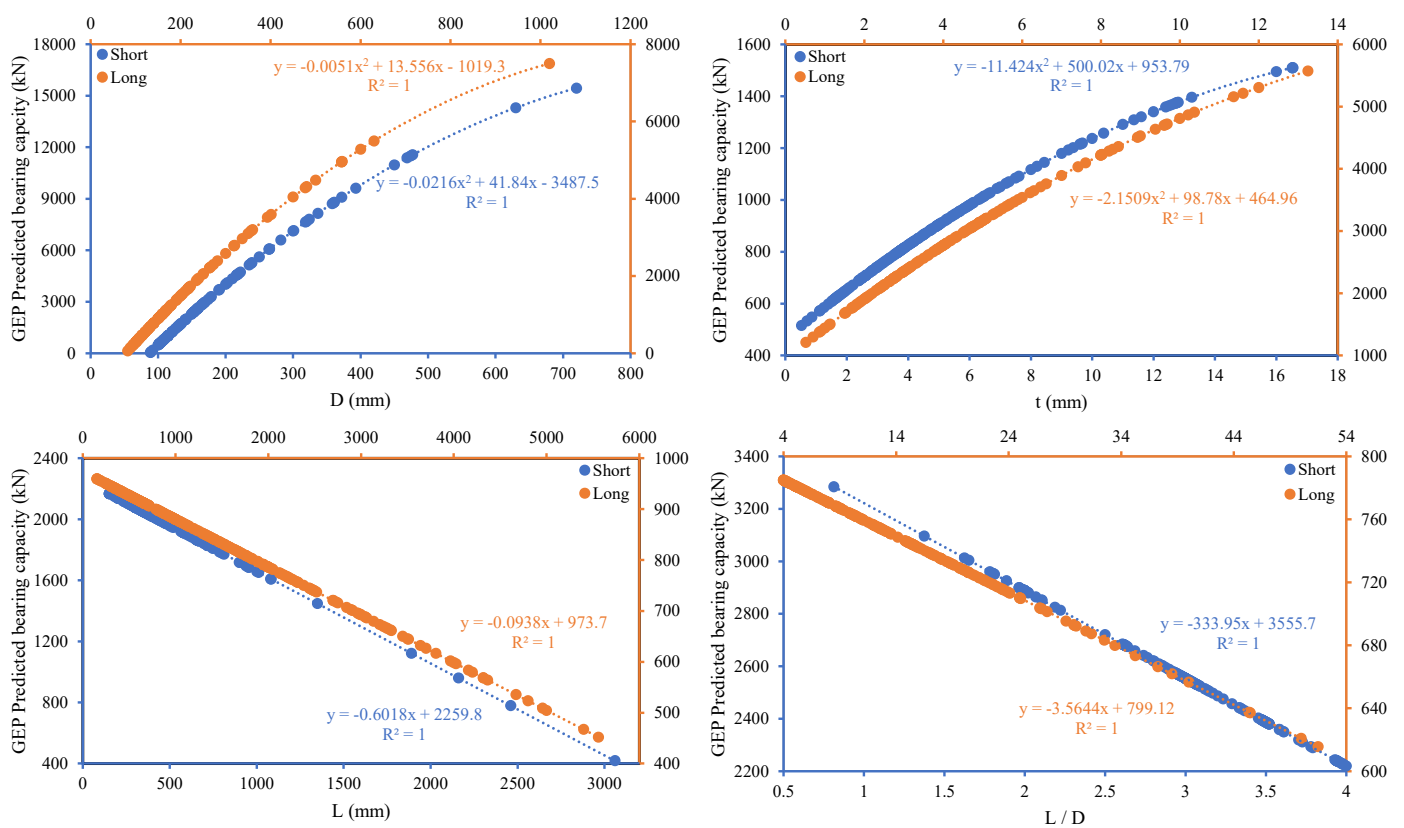
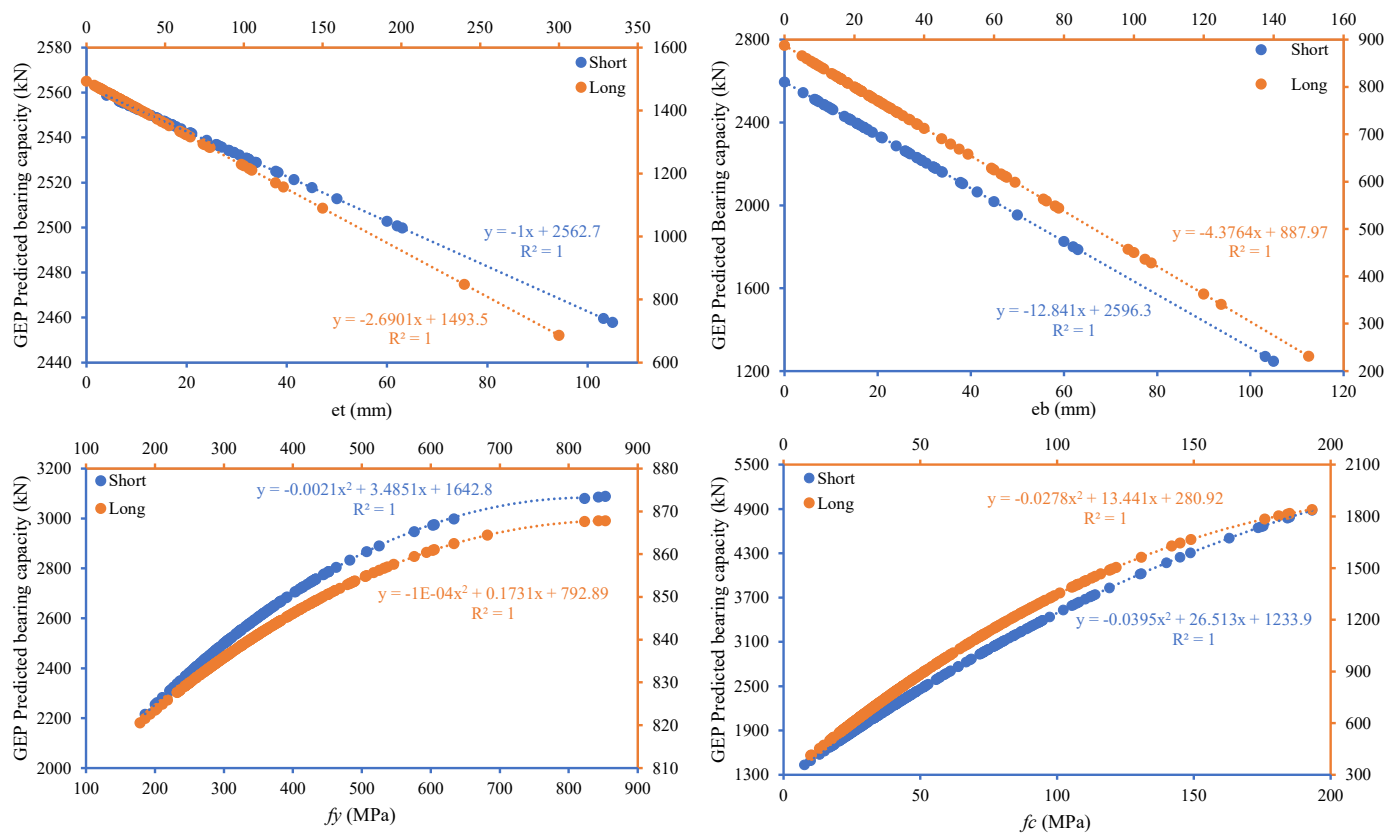


Figure 11. Cont.



**Figure 11.** Summarized parametric study for formulation of bearing capacity of short and long circular CFST columns using Gene expression programming (GEP) in reference to the input variables ( $D$ : diameter of tube,  $t$ : thickness of tube,  $L$ : length of tube,  $L/D$ : length to diameter ratio,  $e_t$  and  $e_b$ : eccentricity at top and bottom surface,  $f_y$ : yield strength of tube,  $f_c$ : compressive strength of concrete).

#### 4. Conclusions

In this study, prediction models for the bearing capacity of circular CFST short ( $N_{st}$ ) and long columns ( $N_{lg}$ ) were developed through ANN and ANFIS and GEP. Two databases were extracted from the literature by collecting 702 datasets of short and 965 datapoints of long circular CFST columns. The conclusion is drawn below.

1. The GEP model can efficiently predict  $N_{st}$  and  $N_{lg}$  with high accuracy and best performance. Moreover, the bearing capacity prediction model from GEP is better than the ANFIS and ANN models. The diversity of the GEP technique can be seen from the simplified formulation, with higher accuracy and correlation among the experimental and predicted data with the consideration of linear and non-linear data.
2. The statistical indicators used to evaluate the performance of the model were mean absolute error (MAE), root square error (RSE), root means square error (RMSE), correlation coefficient ( $R$ ), relative root mean square error (RRMSE), performance index ( $PI$ ) and objective function ( $OF$ ). The  $PI$  of the predicted  $N_{st}$  by GEP, ANN and ANFIS for training are 0.0416, 0.1423, and 0.1016, respectively, and for  $N_{lg}$  these values are 0.1169, 0.2990 and 0.1542, respectively. Corresponding  $OF$  values are 0.2300, 0.1200, and 0.090 for  $N_{st}$ , and 0.1000, 0.2700, and 0.1500 for  $N_{lg}$ . The superiority of the GEP method to the other techniques can be seen from the fact that the GEP technique provides suitable connections based on the practical experimental work and does not undertake prior solutions. In reference to MAPE indicator, the ANN provides excellent forecasting results for  $N_{st}$ , while all other models including  $N_{lg}$ -ANN fall in the "acceptable prediction" category.
3. Sensitivity analysis was performed and the following input importance with increasing pattern was observed for  $N_{st}$ :  $D$  (55.45) >  $T$  (20.45) >  $L$  (17.109) >  $f_c$  (5.526) >  $e_t$

(0.41) >  $L/D$  (0.34) >  $e_b$  (0.32) >  $f_y$  (0.096); whereas, in the case of  $N_{lg}$ , it followed the order:  $D$  (59.83) >  $T$  (33.73) >  $e_t$  (3.844) >  $f_c$  (1.302) >  $L/D$  (0.541) >  $L$  (0.443) >  $e_b$  (0.282) >  $f_y$  (0.033). Parametric analysis showed a trend similar to the findings in previous literature. The effect of input parameters on the bearing capacity of circular short ( $N_{st}$ ) and long ( $N_{lg}$ ) CFST columns was studied. Thus, it can be concluded from this research that artificial intelligence techniques can be effectively employed to solve various complex engineering problems, especially in structural and material engineering. A simple, reliable, and accurate model can be developed which can perform better on unseen data.

- The overall comparison shows that the most reliable and accurate technique for developing prediction models is GEP. The prediction models developed through the GEP technique are simpler than ANN and ANFIS models. It is, therefore, suggested that the developed GEP equations (Equations (22) and (23)) are used in routine design for circular short ( $N_{st}$ ) and long ( $N_{lg}$ ) CFST columns with eccentric loading using simple geometric and material properties. These models can replace tedious, time consuming and costly experimental work for finding the bearing capacity of CFST columns.

It is recommended to study tother AI techniques such as Random Forest Regression (RFR) and Multi-Expression Programming (MEP), etc., on the same data, and the models should be compared for accuracy, reliability, and ability to correlate the predicted data with the experimental data.

**Author Contributions:** Conceptualization, S.K., M.A.K. and M.F.J.; methodology, S.K. and M.A.K.; software, S.K. and M.A.K.; validation, S.K., A.Z. and M.F.J.; formal analysis, A.Z.; investigation, M.A.M. and N.I.V.; resources, A.Z. and M.F.J.; writing—original draft preparation, S.K.; writing—review and editing, M.F.J. and F.A.; supervision, A.Z.; project administration, M.F.J.; funding acquisition, M.A.M., N.I.V., F.A. and A.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation as part of the World-class Research Center program: Advanced Digital Technologies (contract No. 075-15-2020-934 dated 17.11.2020)

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Available on reasonable request from correspondence authors.

**Acknowledgments:** The authors are thankful to National University of Science and Technology (NUST), Islamabad, Pakistan and COMSATS University Islamabad (CU), Islamabad, Pakistan for providing a tremendous research environment. The authors also pay gratitude to CECOS University of IT and Emerging Science, Peshawar, Pakistan.

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

## References

- Romero, M.L.; Espinós, A.; Lapuebla-Ferri, A.; Albero, V.; Hospitaler, A. Recent developments and fire design provisions for CFST columns and slim-floor beams. *J. Constr. Steel Res.* **2020**, *172*, 106159. [[CrossRef](#)]
- Suizi, J.; Wanlin, C.; Zibin, L.; Wei, D.; Yingnan, S. Experimental study on a prefabricated lightweight concrete-filled steel tubular framework composite slab structure subjected to reversed cyclic loading. *Appl. Sci.* **2019**, *9*, 1264. [[CrossRef](#)]
- Ayough, P.; Ibrahim, Z.; Sulong, N.R.; Hsiao, P.-C. The effects of cross-sectional shapes on the axial performance of concrete-filled steel tube columns. *J. Constr. Steel Res.* **2021**, *176*, 106424. [[CrossRef](#)]
- Ibañez, C.; Hernández-Figueirido, D.; Piquer, A. Shape effect on axially loaded high strength CFST stub columns. *J. Constr. Steel Res.* **2018**, *147*, 247–256. [[CrossRef](#)]
- Phan, D.H.H.; Patel, V.I.; Al Abadi, H.; Thai, H.-T. Analysis and design of eccentrically compressed ultra-high-strength slender CFST circular columns. *Structures* **2020**, *27*, 2481–2499. [[CrossRef](#)]
- Thai, S.; Thai, H.-T.; Uy, B.; Ngo, T. Concrete-filled steel tubular columns: Test database, design and calibration. *J. Constr. Steel Res.* **2019**, *157*, 161–181. [[CrossRef](#)]

7. Chinese, S. *Technical Code for Concrete Filled Steel Tubular Structures*; GB 50936; Ministry of Housing and Urban Rural Construction of the People's Republic of China: Beijing, China, 2014.
8. AIJ. *Recommendations for Design and Construction of Concrete Filled Steel Tubular Structures*; AIJ: Tokyo, Japan, 1997.
9. ANSI/AISC 360-05. *Specification for Structural Steel Buildings*; American Institute of Steel Construction: Chicago, IL, USA, 2016; p. 586.
10. Uy, B.; Hicks, S.J.; Kang, W.-H.; Thai, H.-T.; Aslani, F. The New Australia/New Zealand Standard on Composite Steel-Concrete Buildings. In *Proceedings of the 8th International Conference on Composite Construction in Steel and Concrete*, Jackson, WY, USA, 30 July–2 August 2017. ASNZS2327.
11. European Committee for Standardization. *1-1; Eurocode 4: Design of Composite Steel and Concrete Structures—Part 1-1: General Rules and Rules for Buildings*; European Committee for Standardization: Brussels, Belgium, 2004.
12. Liew, J.R. *Design Guide for Concrete Filled Tubular Members with High Strength Materials to Eurocode 4*; Research Publishing: Singapore, 2015; ISBN 978-981-09-3267-1.
13. Khan, M.; Uy, B.; Tao, Z.; Mashiri, F. Behaviour and design of short high-strength steel welded box and concrete-filled tube (CFT) sections. *Eng. Struct.* **2017**, *147*, 458–472. [[CrossRef](#)]
14. Mursi, M.; Uy, B. Strength of slender concrete filled high strength steel box columns. *J. Constr. Steel Res.* **2004**, *60*, 1825–1848. [[CrossRef](#)]
15. Vatulia, G.; Orel, Y.; Rezenenko, M.; Panchenko, N. Using statistical methods to determine the load-bearing capacity of rectangular CFST columns. *MATEC Web Conf.* **2018**, *234*, 04002. [[CrossRef](#)]
16. Le, T.-T. Practical machine learning-based prediction model for axial capacity of square CFST columns. *Mech. Adv. Mater. Struct.* **2020**, 1–16. [[CrossRef](#)]
17. Mai, S.H.; Seghier, M.; Nguyen, P.L.; Jafari-Asl, J.; Thai, D.-K. A hybrid model for predicting the axial compression capacity of square concrete-filled steel tubular columns. *Eng. Comput.* **2020**, 1–18. [[CrossRef](#)]
18. Xu, J.; Wang, Y.; Ren, R.; Wu, Z.; Ozbakkaloglu, T. Performance evaluation of recycled aggregate concrete-filled steel tubes under different loading conditions: Database analysis and modelling. *J. Build. Eng.* **2020**, *30*, 101308. [[CrossRef](#)]
19. Nguyen, H.Q.; Ly, H.-B.; Tran, V.Q.; Nguyen, T.-A.; Le, T.-T.; Pham, B.T. Optimization of Artificial Intelligence System by Evolutionary Algorithm for Prediction of Axial Capacity of Rectangular Concrete Filled Steel Tubes under Compression. *Materials* **2020**, *13*, 1205. [[CrossRef](#)] [[PubMed](#)]
20. Ly, H.-B.; Pham, B.T.; Le, L.M.; Le, 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](#)]
21. Luat, N.-V.; Shin, J.; Lee, K. Hybrid BART-based models optimized by nature-inspired metaheuristics to predict ultimate axial capacity of CCFST columns. *Eng. Comput.* **2020**, 1–30.
22. Dao, D.V.; Ly, H.-B.; Vu, H.-L.T.; Le, T.-T.; Pham, B.T. Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete. *Materials* **2020**, *13*, 1072. [[CrossRef](#)] [[PubMed](#)]
23. Chahnasir, E.S.; Zandi, Y.; Shariati, M.; Dehghani, E.; Toghrol, A.; Mohamad, E.T.; Shariati, A.; Safa, M.; Wakil, K.; Khorami, M. Application of support vector machine with firefly algorithm for investigation of the factors affecting the shear strength of angle shear connectors. *Smart Struct. Syst.* **2018**, *22*, 413–424.
24. Han, Q.; Gui, C.; Xu, J.; Lacidogna, G.J.C.; Materials, B. A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Constr. Build. Mater.* **2019**, *226*, 734–742. [[CrossRef](#)]
25. Khan, M.A.; Memon, S.A.; Farooq, F.; Javed, M.F.; Aslam, F.; Alyousef, R. Compressive strength of fly-ash-based geopolymer concrete by gene expression programming and random forest. *Adv. Civ. Eng.* **2021**, *2021*, 6618407. [[CrossRef](#)]
26. Shariati, M.; Mafipour, M.S.; Haido, J.H.; Yousif, S.T.; Toghrol, A.; Trung, N.T.; Shariati, A. Identification of the most influencing parameters on the properties of corroded concrete beams using an Adaptive Neuro-Fuzzy Inference System (ANFIS). *Smart Struct. Syst.* **2020**, *34*, 155.
27. Singh, V.; Bano, S.; Yadav, A.K.; Ahmad, S. Feasibility of artificial neural network in civil engineering. *IJTSRD* **2019**, *3*, 724–728. [[CrossRef](#)]
28. Hajihassani, M.; Armaghani, D.J.; Kalatehjari, R.J.G.; Engineering, G. Applications of particle swarm optimization in geotechnical engineering: A comprehensive review. *Geotech. Geol. Eng.* **2018**, *36*, 705–722. [[CrossRef](#)]
29. Abdollahzadeh, G.; Jahani, E.; Kashir, Z. Genetic programming based formulation to predict compressive strength of high strength concrete. *Civ. Eng. Infrastruct. J.* **2017**, *50*, 207–219.
30. Ali Khan, M.; Zafar, A.; Akbar, A.; Javed, M.F.; Mosavi, A.J.M. Application of Gene Expression Programming (GEP) for the prediction of compressive strength of geopolymer concrete. *Materials* **2021**, *14*, 1106. [[CrossRef](#)] [[PubMed](#)]
31. Nguyen, Q.H.; Ly, H.-B.; Tran, V.Q.; Nguyen, T.-A.; Phan, V.-H.; Le, T.-T.; Pham, B. A novel hybrid model based on a feedforward neural network and one step secant algorithm for prediction of load-bearing capacity of rectangular concrete-filled steel tube columns. *Molecules* **2020**, *25*, 3486. [[CrossRef](#)]
32. Al-Khaleefi, A.M.; Terro, M.J.; Alex, A.P.; Wang, Y. Prediction of fire resistance of concrete filled tubular steel columns using neural networks. *Fire Saf. J.* **2002**, *37*, 339–352. [[CrossRef](#)]
33. Zarringol, M.; Thai, H.-T.; Thai, S.; Patel, V. Application of ANN to the design of CFST columns. *Structures* **2020**, *28*, 2203–2220. [[CrossRef](#)]



34. Balasubramanian, S.; Jegan, J.; Sundarraja, M. ANFIS-Based Accurate Estimation of the Confinement Effect for Concrete-Filled Steel Tubular (CFST). *Int. J. Fuzzy Syst.* **2020**, *22*, 1760–1771. [[CrossRef](#)]
35. Basarir, H.; Elchalakani, M.; Karrech, A. The prediction of ultimate pure bending moment of concrete-filled steel tubes by adaptive neuro-fuzzy inference system (ANFIS). *Neural Comput. Appl.* **2019**, *31*, 1239–1252. [[CrossRef](#)]
36. Gandomi, A.H.; Roke, D.A. Assessment of artificial neural network and genetic programming as predictive tools. *Adv. Eng. Softw.* **2015**, *88*, 63–72. [[CrossRef](#)]
37. Javed, M.F.; Farooq, F.; Memon, S.A.; Akbar, A.; Khan, M.A.; Aslam, F.; Alyousef, R.; Alabduljabbar, H.; Rehman, S.K.U. New prediction model for the ultimate axial capacity of concrete-filled steel tubes: An evolutionary approach. *Crystals* **2020**, *10*, 741. [[CrossRef](#)]
38. Ipek, S.; Güneyisi, E. Ultimate axial strength of concrete-filled double skin steel tubular column sections. *Adv. Civ. Eng.* **2019**, *2019*, 6493037. [[CrossRef](#)]
39. Güneyisi, E.M.; Gültekin, A.; Mermerdaş, K. Ultimate capacity prediction of axially loaded CFST short columns. *Int. J. Steel Struct.* **2016**, *16*, 99–114. [[CrossRef](#)]
40. Zhang, Q.; Barri, K.; Jiao, P.; Salehi, H.; Alavi, A.H. Genetic programming in civil engineering: Advent, applications and future trends. *Artif. Intell. Rev.* **2021**, *54*, 1863–1885. [[CrossRef](#)]
41. 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](#)] [[PubMed](#)]
42. Venkatesh, K.; Bind, Y.K. ANN and Neuro-Fuzzy Modeling for Shear Strength Characterization of Soils. *Proc. Natl. Acad. Sci. India Sect. A Phys. Sci.* **2020**, 1–7. [[CrossRef](#)]
43. Sada, S.; Ikpeseni, S. Evaluation of ANN and ANFIS modeling ability in the prediction of AISI 1050 steel machining performance. *Heliyon* **2021**, *7*, e06136. [[CrossRef](#)]
44. Kourgialas, N.N.; Dokou, Z.; Karatzas, G.P. Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *J. Environ. Manag.* **2015**, *154*, 86–101. [[CrossRef](#)]
45. Das, S.K. 10 Artificial Neural Networks in Geotechnical Engineering: Modeling and Application Issues. *Metaheuristics Water Geotech. Transp. Eng.* **2013**, *45*, 231–267.
46. Koçak, Y.; Şiray, G.Ü. New activation functions for single layer feedforward neural network. *Expert Syst. Appl.* **2021**, *164*, 113977. [[CrossRef](#)]
47. Xu, B.; Huang, R.; Li, M. Revise saturated activation functions. *arXiv* **2016**, arXiv:1602.05980. preprint.
48. Naresh Babu, K.; Edla, D.R. New algebraic activation function for multi-layered feed forward neural networks. *IETE J. Res.* **2017**, *63*, 71–79. [[CrossRef](#)]
49. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941. preprint.
50. Cai, C.; Xu, Y.; Ke, D.; Su, K. Deep neural networks with multistate activation functions. *Comput. Intell. Neurosci.* **2015**, *2015*, 721367. [[CrossRef](#)]
51. Tang, Y.-J.; Zhang, Q.-Y.; Lin, W. Artificial neural network based spectrum sensing method for cognitive radio. In Proceedings of the 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Chengdu, China, 23–25 September 2010; pp. 1–4.
52. Tahani, M.; Vakili, M.; Khosrojerdi, S. Experimental evaluation and ANN modeling of thermal conductivity of graphene oxide nanoplatelets/deionized water nanofluid. *Int. Commun. Heat Mass Transf.* **2016**, *76*, 358–365. [[CrossRef](#)]
53. Dorofki, M.; Elshafie, A.H.; Jaafar, O.; Karim, O.A.; Mastura, S. Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data. *Int. Proc. Chem. Biol. Environ. Eng.* **2012**, *33*, 39–44.
54. Hanandeh, S.; Ardah, A.; Abu-Farsakh, M. Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula. *Transp. Geotech.* **2020**, *24*, 100358. [[CrossRef](#)]
55. Alavi, A.H.; Gandomi, A.H. A robust data mining approach for formulation of geotechnical engineering systems. *Eng. Comput. Int. J. Comput.-Aided Eng.* **2011**, *28*, 242–274.
56. Nosratabadi, S.; Mosavi, A.; Duan, P.; Ghamisi, P.; Filip, F.; Band, S.S.; Reuter, U.; Gama, J.; Gandomi, A.H. Data science in economics: Comprehensive review of advanced machine learning and deep learning methods. *Mathematics* **2020**, *8*, 1799. [[CrossRef](#)]
57. Shahin, M.A. Artificial intelligence in geotechnical engineering: Applications, modeling aspects, and future directions. In *Metaheuristics in Water, Geotechnical and Transport Engineering*; Curtin University: Perth, Australia, 2013; Volume 169204.
58. Golafshani, E.M.; Behnood, A.; Arashpour, M. Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. *Constr. Build. Mater.* **2020**, *232*, 117266. [[CrossRef](#)]
59. Islam, M.R.; Jaafar, W.Z.W.; Hin, L.S.; Osman, N.; Hossain, A.; Mohd, N.S. Development of an intelligent system based on ANFIS model for predicting soil erosion. *Environ. Earth Sci.* **2018**, *77*, 186. [[CrossRef](#)]
60. Gao, W. A comprehensive review on identification of the geomaterial constitutive model using the computational intelligence method. *Adv. Eng. Inform.* **2018**, *38*, 420–440. [[CrossRef](#)]
61. Shishegaran, A.; Boushehri, A.N.; Ismail, A.F. Gene expression programming for process parameter optimization during ultrafiltration of surfactant wastewater using hydrophilic polyethersulfone membrane. *J. Environ. Manag.* **2020**, *264*, 110444. [[CrossRef](#)]

62. Ferreira, C. Gene expression programming in problem solving. In *Soft Computing and Industry*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 635–653.
63. Wang, M.; Wan, W. A new empirical formula for evaluating uniaxial compressive strength using the Schmidt hammer test. *Int. J. Rock Mech. Min. Sci.* **2019**, *123*, 104094. [[CrossRef](#)]
64. Soleimani, S.; Rajaei, S.; Jiao, P.; Sabz, A.; Soheilinia, S. New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-genetic programming. *Measurement* **2018**, *113*, 99–107. [[CrossRef](#)]
65. Ferreira, C. Mutation, Transposition, and Recombination: An Analysis of the Evolutionary Dynamics. In Proceedings of the 6th Joint Conference on Information Sciences, Research Triangle Park, Raleigh, NC, USA, 8–13 March 2002; pp. 614–617.
66. Armaghani, D.J.; Safari, V.; Fahimifar, A.; Monjezi, M.; Mohammadi, M.A. Uniaxial compressive strength prediction through a new technique based on gene expression programming. *Neural Comput. Appl.* **2018**, *30*, 3523–3532. [[CrossRef](#)]
67. Vyas, R.; Goel, P.; Tambe, S.S. Genetic programming applications in chemical sciences and engineering. In *Handbook of Genetic Programming Applications*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 99–140.
68. Mazari, M.; Rodriguez, D.D. Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. *J. Traffic Transp. Eng. (Engl. Ed.)* **2016**, *3*, 448–455. [[CrossRef](#)]
69. Lam, D.; Goode, C. Concrete Filled Steel Tube Columns-Test compared with Eurocode4. In Proceedings of the International Conference on Composite Construction in Steel and Concrete 2008, Devil's Thumb Ranch, CO, USA, 20–24 July 2008.
70. Mansur, M.A.; Islam, M.M. Interpretation of concrete strength for nonstandard specimens. *J. Mater. Civ. Eng.* **2002**, *14*, 151–155. [[CrossRef](#)]
71. Ahmad, M.R.; Chen, B.; Dai, J.-G.; Kazmi, S.M.S.; Munir, M. Evolutionary artificial intelligence approach for performance prediction of bio-composites. *Constr. Build. Mater.* **2021**, *290*, 123254. [[CrossRef](#)]
72. Maeda, T. *How to Rationally Compare the Performances of Different Machine Learning Models?* PeerJ Preprints: London, UK, 2018; pp. 2167–9843.
73. Jalal, M.; Grasley, Z.; Nassir, N.; Jalal, H. Strength and dynamic elasticity modulus of rubberized concrete designed with ANFIS modeling and ultrasonic technique. *Constr. Build. Mater.* **2020**, *240*, 117920. [[CrossRef](#)]
74. Ferreira, C. *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 21.
75. Alavi, A.H.; Gandomi, A.H.; Nejad, H.C.; Mollahasani, A.; Rashed, A. Design equations for prediction of pressuremeter soil deformation moduli utilizing expression programming systems. *Neural Comput. Appl.* **2013**, *23*, 1771–1786. [[CrossRef](#)]
76. Ferreira, C. Genetic representation and genetic neutrality in gene expression programming. *Adv. Complex Syst.* **2002**, *5*, 389–408. [[CrossRef](#)]
77. 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](#)]
78. Çanakçı, H.; Baykasoğlu, A.; Güllü, H. Prediction of compressive and tensile strength of Gaziantep basalts via neural networks and gene expression programming. *Neural Comput. Appl.* **2009**, *18*, 1031. [[CrossRef](#)]
79. Ağbulut, Ü.; Gürel, A.E.; Biçen, Y. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110114. [[CrossRef](#)]
80. Alade, I.O.; Bagudu, A.; Oyehan, T.A.; Abd Rahman, M.A.; Saleh, T.A.; Olatunji, S.O. Estimating the refractive index of oxygenated and deoxygenated hemoglobin using genetic algorithm-support vector regression model. *Comput. Methods Programs Biomed.* **2018**, *163*, 135–142. [[CrossRef](#)]
81. Zhang, W.; Zhang, R.; Wu, C.; Goh, A.T.C.; Lacasse, S.; Liu, Z.; Liu, H. State-of-the-art review of soft computing applications in underground excavations. *Geosci. Front.* **2020**, *11*, 1095–1106. [[CrossRef](#)]
82. Shahin, M.A. Use of evolutionary computing for modelling some complex problems in geotechnical engineering. *Geomech. Geoengin.* **2015**, *10*, 109–125. [[CrossRef](#)]
83. Gandomi, A.H.; Alavi, A.H.; Mirzahosseini, M.R.; Nejad, F.M. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *J. Mater. Civ. Eng.* **2011**, *23*, 248–263. [[CrossRef](#)]
84. Emamgholizadeh, S.; Bahman, K.; Bateni, S.M.; Ghorbani, H.; Marofpoor, I.; Nielson, J.R. Estimation of soil dispersivity using soft computing approaches. *Neural Comput. Appl.* **2017**, *28*, 207–216. [[CrossRef](#)]
85. Chu, H.-H.; Khan, M.A.; Javed, M.; Zafar, A.; Khan, M.I.; Alabduljabbar, H.; Qayyum, S. Sustainable use of fly-ash: Use of gene-expression programming (GEP) and multi-expression programming (MEP) for forecasting the compressive strength geopolymer concrete. *Ain Shams Eng. J.* **2021**, *12*, 3603–3617. [[CrossRef](#)]
86. Khan, M.A.; Shah, M.I.; Javed, M.F.; Khan, M.I.; Rasheed, S.; El-Shorbagy, M.; El-Zahar, E.R.; Malik, M. Application of random forest for modelling of surface water salinity. *Ain Shams Eng. J.* **2021**, in press.
87. Erzin, Y. Artificial neural networks approach for swell pressure versus soil suction behaviour. *Can. Geotech. J.* **2007**, *44*, 1215–1223. [[CrossRef](#)]
88. Aslam, F.; Elkotb, M.A.; Iqtidar, A.; Khan, M.A.; Javed, M.F.; Usanova, K.I.; Khan, M.I.; Alamri, S.; Musarat, M.A. Compressive strength prediction of rice husk ash using multiphysics genetic expression programming. *Ain Shams Eng. J.* **2021**, in press. [[CrossRef](#)]
89. Gholampour, A.; Gandomi, A.H.; Ozbakkaloglu, T. New formulations for mechanical properties of recycled aggregate concrete using gene expression programming. *Constr. Build. Mater.* **2017**, *130*, 122–145. [[CrossRef](#)]

90. Mohammadzadeh, S.; Kazemi, S.-F.; Mosavi, A.; Nasseralshariati, E.; Tah, J.H. Prediction of compression index of fine-grained soils using a gene expression programming model. *Infrastructures* **2019**, *4*, 26. [[CrossRef](#)]
91. Frank, I.E.; Todeschini, R. *The Data Analysis Handbook*; Elsevier: Amsterdam, The Netherlands, 1994.
92. Nguyen, T.; Kashani, A.; Ngo, T.; Bordas, S. Deep neural network with high-order neuron for the prediction of foamed concrete strength. *Comput. Civ. Infrastruct. Eng.* **2019**, *34*, 316–332. [[CrossRef](#)]
93. Iqbal, M.F.; Javed, M.F.; Rauf, M.; Azim, I.; Ashraf, M.; Yang, J.; Liu, Q.-F. Sustainable utilization of foundry waste: Forecasting mechanical properties of foundry sand based concrete using multi-expression programming. *Sci. Total Environ.* **2021**, *780*, 146524. [[CrossRef](#)]
94. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*; Butterworth-Heinemann: Oxford, UK, 1982.
95. Chen, R.J.C.; Bloomfield, P.; Cabbage, F.W. Comparing Forecasting Models in Tourism. *J. Hosp. Tour. Res.* **2008**, *32*, 3–21. [[CrossRef](#)]
96. Mollahasani, A.; Alavi, A.H.; Gandomi, A.H. Empirical modeling of plate load test moduli of soil via gene expression programming. *Comput. Geotech.* **2011**, *38*, 281–286. [[CrossRef](#)]
97. Roy, P.P.; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **2008**, *27*, 302–313. [[CrossRef](#)]
98. Trucchia, A.; Frunzo, L. Surrogate based Global Sensitivity Analysis of ADM1-based Anaerobic Digestion Model. *J. Environ. Manag.* **2021**, *282*, 111456. [[CrossRef](#)] [[PubMed](#)]
99. Wang, W.; Ma, H.; Li, Z.; Tang, Z. Size effect in circular concrete-filled steel tubes with different diameter-to-thickness ratios under axial compression. *Eng. Struct.* **2017**, *151*, 554–567. [[CrossRef](#)]
100. Yadav, R.; Chen, B. Parametric study on the axial behaviour of concrete filled steel tube (CFST) columns. *Am. J. Appl. Sci. Res.* **2017**, *3*, 37–41. [[CrossRef](#)]
101. Cai, J.; Pan, J.; Lu, C.; Li, X. Nonlinear analysis of circular concrete-filled steel tube columns under eccentric loading. *Mag. Concr. Res.* **2020**, *72*, 292–303. [[CrossRef](#)]