

Article

Evolutionary Algorithms for Strength Prediction of Geopolymer Concrete

Bingzhang Huang ^{1,2}, Alireza Bahrami ^{3,*} , Muhammad Faisal Javed ⁴ , Iftikhar Azim ⁵ 
and Muhammad Ayyan Iqbal ⁶

- ¹ School of Civil Engineering and Architecture, Liuzhou Institute of Technology, Liuzhou 545004, China
 - ² Guangxi Prefabricated Building Life Cycle Management and Virtual Simulation Engineering Research Center, Liuzhou 545004, China
 - ³ Department of Building Engineering, Energy Systems and Sustainability Science, Faculty of Engineering and Sustainable Development, University of Gävle, 801 76 Gävle, Sweden
 - ⁴ Department of Civil Engineering, GIK Institute of Engineering Sciences and Technology, Topi, Swabi 23460, Pakistan
 - ⁵ Public Health Engineering Department, Government of Khyber Pakhtunkhwa, Peshawar 25000, Pakistan
 - ⁶ Department of Civil Engineering, University of Engineering and Technology, Lahore 39161, Pakistan
- * Correspondence: alireza.bahrami@hig.se

Abstract: Geopolymer concrete (GPC) serves as a sustainable substitute for conventional concrete by employing alternative cementitious materials such as fly ash (FA) instead of ordinary Portland cement (OPC), contributing to environmental and durability benefits. To increase the rate of utilization of FA in the construction industry, distinctive characteristics of two machine learning (ML) methods, namely, gene expression programming (GEP) and multi-expression programming (MEP), were utilized in this study to propose precise prediction models for the compressive strength and split tensile strength of GPC comprising FA as a binder. A comprehensive database was collated, which comprised 301 compressive strength and 96 split tensile strength results. Seven distinct input variables were employed for the modeling purpose, i.e., FA, sodium hydroxide, sodium silicate, water, superplasticizer, and fine and coarse aggregates contents. The performance of the developed models was assessed via numerous statistical metrics and absolute error plots. In addition, a parametric analysis of the finalized models was performed to validate the prediction ability and accuracy of the finalized models. The GEP-based prediction models exhibited better performance, accuracy, and generalization capability compared with the MEP-based models in this study. The GEP-based models demonstrated higher correlation coefficients (R) for predicting the compressive and split tensile strengths, with the values of 0.89 and 0.87, respectively, compared with the MEP-based models, which yielded the R values of 0.76 and 0.73, respectively. The mean absolute errors for the GEP- and MEP-based models for predicting the compressive strength were 5.09 MPa and 6.78 MPa, respectively, while those for the split tensile strengths were 0.42 MPa and 0.51 MPa, respectively. The finalized models offered simple mathematical formulations using the GEP and Python code-based formulations from MEP for predicting the compressive and tensile strengths of GPC. The developed models indicated practical application potential in optimizing geopolymer mix designs. This research work contributes to the ongoing efforts in advancing ML applications in the construction industry, highlighting the importance of sustainable materials for the future.

Keywords: geopolymer concrete; compressive strength; split tensile strength; prediction model; evolutionary algorithm



Citation: Huang, B.; Bahrami, A.; Javed, M.F.; Azim, I.; Iqbal, M.A. Evolutionary Algorithms for Strength Prediction of Geopolymer Concrete. *Buildings* **2024**, *14*, 1347. <https://doi.org/10.3390/buildings14051347>

Academic Editor: Bjorn Birgisson

Received: 10 March 2024

Revised: 8 April 2024

Accepted: 16 April 2024

Published: 9 May 2024



Copyright: © 2024 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

Concrete is a vital material widely used in modern construction, which consists of fine and coarse aggregates, cement, water, and admixtures [1]. Nevertheless, concerns over global warming have underscored the need to reduce carbon dioxide (CO₂) emissions,

primarily attributed to cement production. In fact, each ton of produced ordinary Portland cement (OPC) releases around one ton of CO₂ [2]. Additionally, the disposal of construction and demolition waste poses serious environmental issues [3,4]. Thus, the requirement of both green construction and sustainability in the construction industry demands new materials [5,6]. In response to these concerns, researchers have explored sustainable and eco-friendly solutions such as geopolymer cement. This type of cement is produced from raw materials containing aluminosilicate such as fly ash (FA), ground granulated blast furnace slag (GGBFS), and metakaolin, treated with alkali and alkali silicates [7]. Geopolymer concrete (GPC) reduces CO₂ emissions by 80% and is more cost-effective since it utilizes industrial and agricultural wastes. Its use also reduces the quantity of wastes sent to landfill sites, benefiting the natural ecosystem [8]. Furthermore, GPC has shown superior mechanical properties compared with conventional OPC-based concrete, including higher compressive strength (CS) and split tensile strength (ST), as well as better resistance to acid, fire, and high temperature. Geopolymerization, a fundamental chemical process integral to the formation of GPC, unfolds through distinct stages. Firstly, the aluminosilicate constituents are dissolved, resulting in the release of aluminate and silicate monomers such as Al(OH)₄ and Si(OH)₄. These monomers then undergo condensation, forming initial gels through sharing of oxygen atoms, resulting in mono cross-linked systems. In the final step, the initial gels undergo polycondensation, transforming into geopolymer gels. The process of geopolymerization is a vital step in the production of GPC, representing an eco-friendly alternative to conventional concrete [8].

FA, a byproduct of coal combustion, has been utilized for many years as a partial replacement for OPC [9]. Because of its aluminous and siliceous composition, FA can form a compound similar to OPC when mixed with water and lime. This makes it a suitable material for blended cement, mosaic tiles, and hollow blocks [10]. FA has lower embodied energy compared with other pozzolanic precursors such as metakaolin and GGBFS [11]. Moreover, FA has a good solidification effect on heavy metal pollutants, making it suitable to be used as an alternative cementitious material in concrete [12]. Amid the growing environmental concerns and the demand for sustainable construction materials, there has been a noteworthy rise in interest in FA-based GPC in the construction industry. GPC is produced by treating raw materials rich in aluminosilicates with alkali and alkali silicates. The process not only reduces CO₂ emissions but also exploits industrial or agricultural wastes. FA-based GPC ensures high strength, lower exploitation of natural resources, and low CO₂ emissions, making it an innovative and sustainable alternative to conventional concrete systems.

Optimizing the mix design of GPC proves to be a complex task, given the involvement of numerous parameters, including the types and concentrations of silicates, replacement material used for cement, admixtures, curing conditions, and curing time. Traditional experimental procedures for achieving optimal results are labor-intensive and time-consuming, requiring extensive laboratory-based experiments and significant resources. Therefore, the development of time- and cost-effective techniques to determine the correct proportions of constituents required for GPC formulations is necessary. These techniques can help streamline the optimization process and enhance the overall sustainability of the construction industry.

In recent years, machine learning (ML) techniques have been increasingly employed in civil engineering, among other fields, to drive advancements and contribute to societal progress. Traditional methods for predicting the mechanical properties of concrete relied on mathematical and statistical forecasting, along with non-linear prediction methods. However, the development of ML techniques has revolutionized the creation of accurate and reliable models for addressing civil engineering problems [13]. ML processes, rooted in natural phenomena, are implemented via various techniques such as the genetic algorithm (GA), genetic programming (GP), gene expression programming (GEP), multi-expression programming (MEP), adaptive neuro-fuzzy interface (ANFIS), fuzzy logic (FL), grey wolf optimization (GWO), random forest regression (RFR), artificial neural networks

(ANNs), and support vector machine (SVM) [14]. Leveraging the pattern recognition capabilities of ML, these techniques produce simplified models of intricate patterns, facilitating the optimization of the mix design of GPC [15,16]. ML-based approaches offer a time and cost-effective alternative by minimizing the dependence on extensive laboratory-based experiments, which typically involves substantial resources such as materials, time, and labor.

Several studies have utilized ML techniques to estimate the mechanical properties of various kinds of concrete. Khan et al. [17] proposed a GEP-based model to predict CS of FA-based GPC. The results were in a good agreement with the experimental investigations considered in the study. In addition, parametric analysis was performed to demonstrate that the developed model takes into account the underlying physical relationship in the considered system. Chu et al. [18] utilized GEP and MEP algorithms for the prediction of CS of FA-based GPC. It was concluded that the GEP-based model has a higher correlation coefficient (R) and minimal statistical errors compared with MEP. Similarly, Khan et al. [19] applied GEP and RFR algorithms to predict CS of FA-based GPC. It was reported that RFR outperformed GEP by giving a higher R value and minimal statistical errors, while GEP provided a simple empirical equation to estimate CS of GPC. Following this, Khan et al. [20] established numerous prediction models for the prediction of FA-based GPC by employing ANN, ANFIS, and GEP. The three models met the verification criterion proposed in the literature. However, the GEP-based model was considered ideal and robust because it provided a simple mathematical formulation and a higher generalization capability compared with others. Recently, Zhang et al. [21] proposed a hybrid RFR-GWO-XGBoost algorithm for predicting CS of GPC. The results were compared with stand-alone RFR and XGBoost models to display the supremacy of the proposed methodology. The GEP algorithm was utilized by Iqbal et al. [2] to estimate CS, ST, and elastic modulus of waste foundry sand (WFS)-based green concrete. GEP-based results were compared with linear and non-linear regression models to validate the proposed models. In another study, Iqbal et al. [16] applied MEP to predict ST and modulus of elasticity of WFS-based concrete. Both studies involved model validation and parametric studies to exhibit the accurate prediction of the systems under consideration. Meanwhile, the concrete strength comprising rice husk ash was evaluated by employing ANN in [22]. The Bayesian ANN technique was exploited to determine the strength of alkali-activated GPC comprising FA and bottom ash [8]. Shahmansouri et al. [23] incorporated natural zeolite and silica fume in ground GGBFS-based GPC to evaluate their effects on CS and developed an ANN prediction model for its mechanical properties. Peng and Unluer [24] assessed the performance of numerous ML algorithms for predicting CS of GPC incorporating waste glass powder and slag. Their results indicated that the support vector regression and random forest models outperformed the other algorithms applied in the study. It was further concluded that the addition of waste glass powder and slag improved CS of GPC. Ahmad et al. [25] utilized ANNs to develop a model for predicting the strength of GPC incorporating waste ceramic tiles and quarry dust as a partial replacement for fine aggregates. The ANN model illustrated better accuracy in predicting compared with traditional statistical models.

The review of the above-mentioned studies reveals that the modeling of CS was the primary focus of the published studies. These modeling efforts have largely overlooked ST of GPC, which is a crucial property that influences the performance, durability, and applicability of FA-based GPC in various construction scenarios. The majority of the studies have predominantly used ANNs, support vector regression models, XGBoost, and others. While these algorithms are more accurate than evolutionary algorithms, they do not provide simple mathematical formulations, which limit their utility for other researchers. Meanwhile, most of the aforementioned studies are restricted to smaller databases than what is typically used for a comprehensive analysis. It is worth noting that increasing the number of datasets improves the quality of ML-based models.

This research work seeks to address the above-mentioned gaps by developing prediction models utilizing the GEP and MEP algorithms to predict CS (f'_c) and ST (f'_{st}) of GPC

containing FA as a binder. For this purpose, a comprehensive database was collated from internationally published experimental results. Numerous combinations of input parameters were employed, and the results obtained from both algorithms were compared. The performance of the developed models was assessed using parametric and comprehensive statistical analyses. The accuracy and reliability of the models were validated with the experimental data. The significance of this study lies in its exploration of ML models, which have not been extensively utilized in the context of GPC. This study provides valuable insights into the applicability and accuracy of GEP and MEP for predicting the mechanical properties of GPC via simple mathematical formulations. The practical application potential of the developed models is evident in their ability to guide engineers and practitioners in selecting optimal GPC formulations, reducing material waste, and promoting the use of eco-friendly alternatives in the construction industry.

2. GP and Its Variants

Computer scientists draw inspiration from the power of natural evolution when developing automated problem solvers, i.e., algorithms. Nowadays, algorithms serve as central themes in modern problem-solving techniques [26]. An excellent example of the mimicking of evolutionary processes is GP, which is a branch of evolutionary algorithms. It was developed by Koza [27] to overcome the limitations of pattern recognition methods such as ANNs, FL, SVM, and ANFIS [28]. GP is the successor of GA developed by John Holland [29]. For the last two decades, GP and its variants, including GEP and MEP, have emerged as powerful techniques for modeling complex physical phenomena in civil engineering discipline. The following sections provide a detailed explanation of GEP and MEP.

2.1. GEP

As mentioned above, GEP belongs to the kin of evolutionary algorithms and is a direct descendent of GP. It was proposed by Ferreira [30], wherein individuals, i.e., candidate solutions, are encoded as linear strings of constant size known as a genome. The strings are later presented as non-linear entities with changing shapes and sizes called expression trees (ETs). GEP consists of a genotype–phenotype system in which a simple genome is stored and passes on genetic information, while a complex phenotype explores and acclimates to the environment, akin to a living organism. Models generated by GEP consist of multiple parse trees, owing to the multigenic nature of its genotype–phenotype system, which enables the assessment of complex programs comprising multiple sub-programs. The key difference between GEP and classical GP lies in the fixed-length string representation of candidate solutions generated by GEP, which are subsequently expressed as parse trees of varying sizes and shapes during the fitness evaluation. Chromosomes and ETs are two fundamental parameters of GEP, and the translation process involves decoding information from chromosomes to ETs based on a set of rules. ETs usually consist of a single chromosome, which comprises one or multiple genes. It should be noted that a gene comprises a head and a tail [31].

Like other evolutionary algorithms, the chromosomes of the individuals of the initial population are randomly generated via the functions and terminals deemed suitable to solve a problem in the GEP algorithm. These founder individuals are completely random and are yet to be toughened by the environment. Hence, they often prove to be inadequate solutions. To generate new individuals/candidate solutions, ETs undergo a selection process directed by their fitness utilizing roulette wheel selection, which ensures the replication and survival of the fittest individual to the next generation. These individuals are exposed to the same developmental process, i.e., expressing chromosomes as ETs, fitness evaluation, selection, and reproduction with modification. A flowchart of working of the GEP algorithm is depicted in Figure 1. The best individuals (i.e., those with the highest fitness scores) obtained from a generation are always kept for the next generation, often through a process known as elitism. This helps ensure that good solutions found in

previous generations are not lost and can continue to be improved upon. Genetic operators such as crossover, rotation, and mutation are employed by the GEP algorithm to initiate variations in the population by altering the chromosomes at the reproduction stage [32].

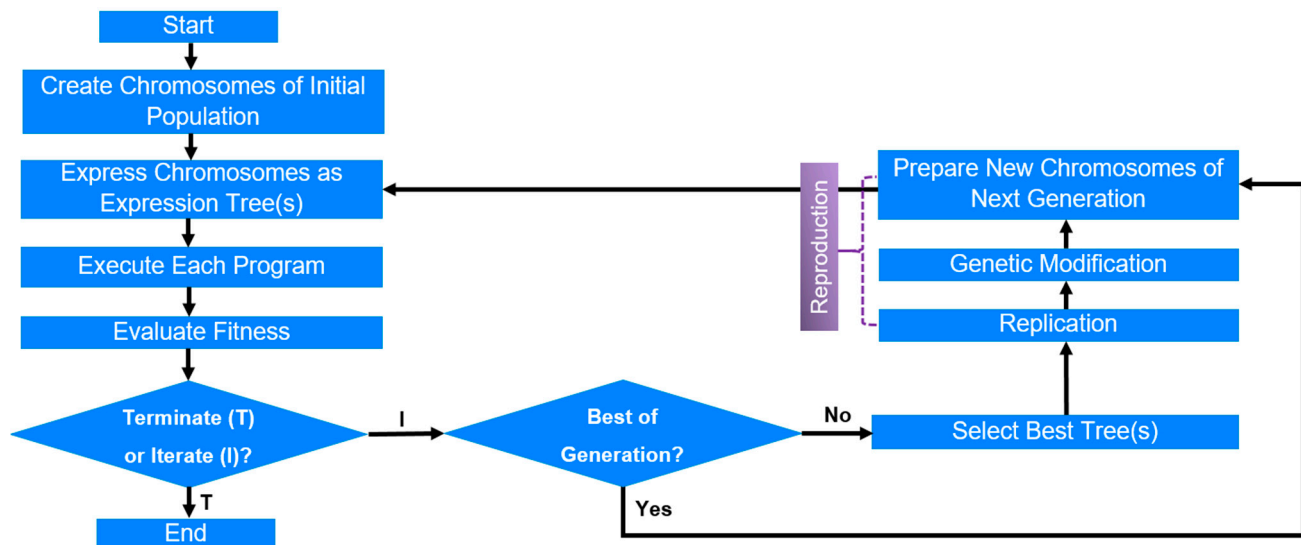


Figure 1. Flowchart of GEP algorithm.

2.2. MEP

MEP was proposed by Oltean and Dumitrescu [33]. It has gained substantial consideration in recent years as it offers a unique approach to problem-solving by utilizing a linear genome structure (similar to GEP) and multiple expressions or sub-programs. The latter gives MEP a unique capability to encode multiple solutions for a problem under consideration in a single chromosome. This ability renders this technique highly efficient, especially when the complexity of the problem is not known.

The algorithm process begins by generating a random population of chromosomes. Subsequently, a binary procedure is used to select two parents from this population. The parents undergo recombination, while the resulting offspring undergoes mutation. Finally, the least fit individuals are replaced with the newly generated offspring [34]. The process continues until a termination condition is reached. A flowchart of the MEP process is shown in Figure 2.

In MEP, the length of a chromosome is determined by a fixed number of genes per chromosome. Each gene within the chromosome encodes the elements existing in the function and terminal set. The output from MEP is in the form of linear string of instructions. These instructions are formed by combining functions (mathematical variables) and terminals (variables). In the chromosome structure, initial symbol represents a terminal symbol. A function gene, on the other hand, consists of pointers that the reference function arguments. The function parameters have indices corresponding to values lower than the position of the function within the chromosome. This can be better understood from the example given below considering the set of terminals as $T = \{Z_1, Z_2, Z_3, Z_4\}$ and the set of functions $F = \{+, \times, ^\wedge\}$.

0: Z_1 .

1: Z_2 .

2: $+ 0, 1$.

3: Z_3 .

4: $\times 2, 3$,

5: Z_4 .

6: $^{\wedge} 4, 5$.

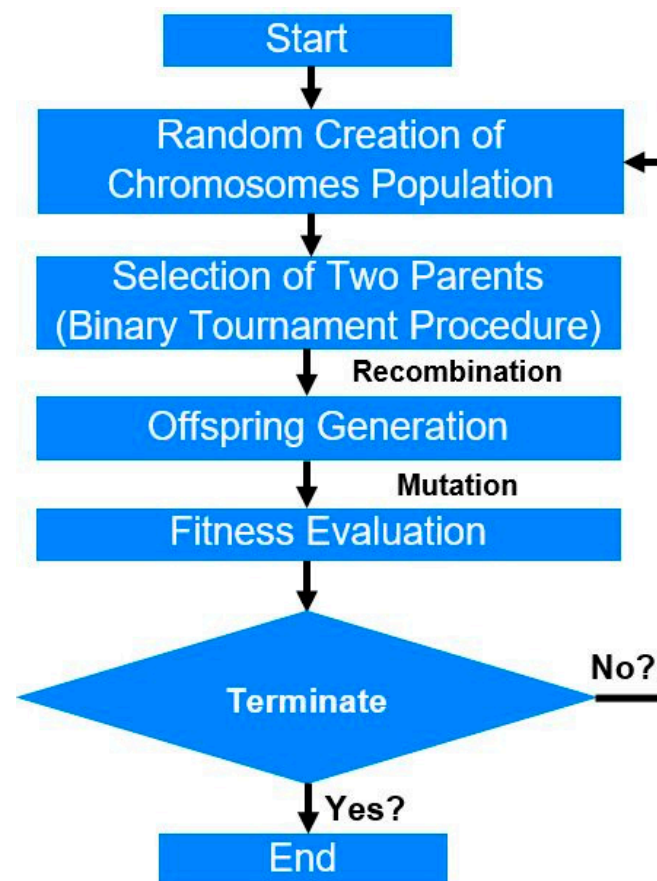


Figure 2. Flowchart of MEP algorithm [34].

Genes 0, 1, 3, and 5 encode simple expressions formed by a single terminal symbol. The expressions formed are given below:

$$G_0 = Z_1; G_1 = Z_2; G_3 = Z_3; G_5 = Z_4$$

From the above, we can interpret that gene 2 indicates the operation “+” applied to the operands situated at the positions 0 and 1 within the chromosome (Equation (1)). Likewise, genes 4 and 6 in the chromosome correspond to the operators “x” and “^” applied to the operands situated at the positions 2, 3, and 4, 5, respectively (Equations (2) and (3)). Thus, the expressions encoded by the genes are as follows:

$$G_2 = Z_1 + Z_2 \quad (1)$$

$$G_4 = (Z_1 + Z_2) Z_3 \quad (2)$$

$$G_6 = [(Z_1 + Z_2) Z_3]^{Z_4} \quad (3)$$

The MEP chromosome encodes multiple solutions or expressions (G_0, \dots, G_6), allowing for a diverse range of solutions within a single chromosome. This multi-expression representation in MEP results in a chromosome that can be regarded as a forest of trees, unlike in GEP, where a single tree structure is utilized. Figure 3 displays the forest of expressions encoded by the chromosome, as described earlier.

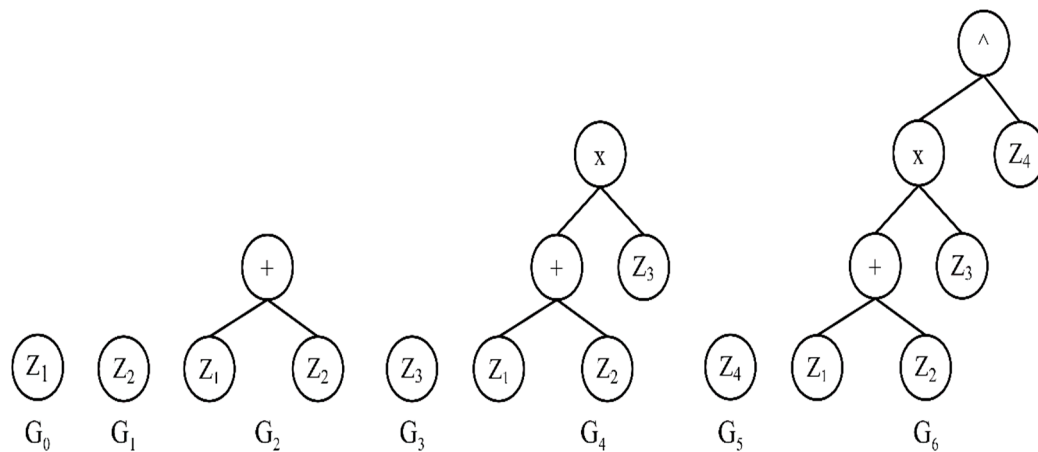


Figure 3. Forest trees serve as representation for expressions encoded within a MEP chromosome [34].

3. Research Methodology

3.1. Database Development and Data Curation

To predict CS and ST of GPC through GEP and MEP, a database was compiled from the published literature. After conducting a comprehensive literature review and performing initial trials, key input parameters having a substantial impact on f'_c of GPC were identified. Based on our preliminary findings, we determined that f'_c and f'_{st} are a function of the factors listed in Equations (4) and (5), respectively. The purpose of this study is to explore the effects of these input parameters on f'_c and f'_{st} of GPC and propose new models for their prediction.

$$f'_c = f(FA, C_{agg}, F_{agg}, NaSi, NaOH, SP, w) \quad (4)$$

$$f'_{st} = f(FA, C_{agg}, NaOH, w, GGBFS) \quad (5)$$

For the 28-day f'_c , 301 data points were collected from the literature [3,8,9,11,17,23,35–39]. The input parameters comprised the contents of fly ash (FA), fine aggregate (F_{agg}), coarse aggregate (C_{agg}), sodium hydroxide (NaOH), sodium silicate (NaSi), water content (w), and superplasticizer (SP). The f'_{st} model was developed using 96 data points, and an additional parameter, GGBFS, was included in this dataset, given its common usage as an alternative to cementitious materials in GPC. The input parameters were consistently recorded or converted in kg/m^3 units, where applicable. Table 1 summarizes the primary sources of the database. The database was constructed through a meticulous search on Google Scholar, employing keywords associated with FA-based GPC. This initial search yielded around 50 articles. Given the variety of binders available for GPC development, a detailed analysis was carried out to ensure the articles selected primarily focused on FA-based GPC. Criteria were established to narrow down the selection, with an emphasis on the number of the input variables, capping it at seven, as indicated in Equation (4). The data collection was systematically approached, ensuring each article provided data on at least five or six essential input variables such as FA, C_{agg} , F_{agg} , NaSi, NaOH, and w . SP was deemed as a secondary input.

Tables 2 and 3 provide the descriptive statistics for the input variables utilized in the f'_c and f'_{st} models' development. These statistics, including measures of central tendency and variability such as the standard deviation, as well as distribution shape indicators like skewness, offer deep insights into the variables' range and distribution characteristics. The range of these variables can act as a preliminary guide before applying the developed models. In addition, the even distribution of the input variables suggests that they are well-suited for training ML models, enhancing the models' ability to learn and predict accurately.

Table 1. Data sources used for modeling study.

Source	Theme of Article	Output Parameter
Aneja et al. [8]	Experiment and ANN models to predict strength of GPC	f'_c
Gomaa et al. [11]	Experiments on FA-based GPC	f'_c and f'_{st}
Shahmansouri et al. [23]	Experiment and ANNs on GPC	f'_c
Nguyen et al. [35]	Experimental and neural network approaches	f'_c
Naghizadeh et al. [36]	Experiments on FA geopolymers binder	f'_c
Hardjito et al. [37]	Development of low-calcium FA-based GPC	f'_c and f'_{st}
Rangan et al. [38]	FA-based GPC	f'_c and f'_{st}
Gomaa et al. [39]	Experiments on high-calcium alkali-activated FA mortar	f'_c
Amran et al. [9]	Review of FA-based GPC	f'_c and f'_{st}
Khan et al. [17]	ML models for CS of GPC	f'_c and f'_{st}

Table 2. Descriptive statistics of input variables used for f'_c modeling.

Parameter	FA	C_{agg}	F_{agg}	NaOH	NaSi	w	SP
Mean	417.77	995.49	726.64	59.03	128.28	18.06	4.37
Standard error	9.69	23.93	19.01	3.41	4.93	1.12	0.41
Median	408.00	1170.00	647.00	52.90	119.00	16.50	0.70
Mode	408.00	1170.00	630.00	41.00	103.00	0	0
Standard deviation	168.61	416.57	330.92	59.44	85.77	19.43	7.21
Sample variance	28,430.07	173,531.60	109,505.13	3533.21	7356.41	377.57	51.94
Kurtosis	37.90	0.99	9.58	50.96	25.10	3.28	3.86
Skewness	5.65	−1.58	2.69	6.30	3.77	1.29	2.14
Range	1368.00	1591.00	2085.00	600.00	800.00	113.60	28.00
Minimum	232.00	0	315.00	0	0	0	0
Maximum	1600.00	1591.00	2400.00	600.00	800.00	113.60	28.00

Table 3. Descriptive statistics of input variables utilized for f'_{st} modeling.

Parameter	C_{agg}	FA	GGBFS	NaOH	w
Mean	987.05	358.29	101.74	77.84	35.75
Standard error	48.24	39.41	8.87	12.76	5.93
Median	1142.60	286.00	110.00	54.73	0
Mode	1143.00	286.00	0	54.73	0
Standard deviation	472.65	386.13	86.90	125.06	58.08
Sample variance	223,395.17	149,095.50	7551.93	15,639.88	3373.56
Kurtosis	−0.41	6.54	−1.29	14.82	0.09
Skewness	−0.58	2.79	0.17	3.56	1.31
Range	1870.00	1600.00	270.00	800.00	175.00
Minimum	0	0	0	0	0
Maximum	1870.00	1600.00	270.00	800.00	175.00

The data were split into a training set comprising 80% of the samples and a validation set consisting of the remaining 20% of the samples. For the f'_c model, 239 data points were utilized for training and the remaining 62 points were utilized for validation. Similarly, for the f'_{st} model, 79 and 17 data points were allocated for training and validation, respectively. It should be noted that the validation data comprised a combination of data used for validation during training (10%) to meet the performance criteria, and the manual verification of the unseen data (10%) after training was completed by the algorithm. Additionally, the data were randomly arranged for both the models to maintain objectivity and ensure reliable results [40,41]. The interdependency of variables in a model is a critical concern, as

it can lead to difficulties in accurately interpreting results and, consequently, to suboptimal the performance of the model. This challenge, often referred to as the “multicollinearity problem”, arises when the input variables are not independent [15]. To ensure the development of a reliable model, it is recommended that the correlation between any two input variables should not exceed 0.80 [32]. As a result, R values for every possible pair of the input variables were computed, as detailed in Tables 4 and 5. The findings exhibited that both positive and negative correlations between various parameters are well below the 0.80 threshold, effectively reducing the potential impact of multicollinearity.

Table 4. R values among input variables used for f'_c modeling.

	FA	C_{agg}	F_{agg}	$NaOH$	$NaSi$	w	SP
FA	1	−0.40	0.76	0.68	0.63	0.17	0.06
C_{agg}	−0.40	1	−0.72	−0.05	−0.42	−0.04	−0.22
F_{agg}	0.76	−0.72	1	0.39	0.70	0.26	0.16
$NaOH$	0.68	−0.05	0.39	1	0.28	−0.07	0.22
$NaSi$	0.63	−0.42	0.70	0.28	1	0.20	0.05
w	0.17	−0.04	0.26	−0.07	0.20	1	−0.02
SP	0.06	−0.22	0.16	0.22	0.05	−0.02	1

Table 5. R values among input variables utilized for f'_{st} modeling.

	C_{agg}	FA	$GGBFS$	$NaOH$	w
C_{agg}	1	−0.66	−0.05	−0.42	−0.10
FA	−0.66	1	−0.41	0.80	0.54
$GGBFS$	−0.05	−0.41	1	−0.29	−0.62
$NaOH$	−0.42	0.80	−0.29	1	0.43
w	−0.10	0.54	−0.62	0.43	1

3.2. GEP's Optimal Parameter Settings

Table 6 provides a detailed overview of the specific parameters employed in the GEP algorithm, which were finalized after conducting 47 different trials. The conducted trials encompassed adjustments of several parameters, as outlined in the table, including the number of chromosomes, ranging from 30 to 450 across different models, and the number of genes, ranging from 0 to 6 with a step size of 1. It should be noted that the models for CS and ST were denoted as MG-CS and MG-ST, respectively. Additionally, the head size, determining the ultimate complexity of the models or formulations, was set at 12 for the CS model and 8 for the ST model. The addition “+” function was selected as the linking function in both the models to ensure simplicity in the final equations. Different mathematical operators and function sets were utilized to achieve the desired accuracy. The number of generations in the trial models was kept at an optimum value to allow the algorithm to evolve properly.

Table 6. GEP algorithm's parameter settings for both models.

Parameter	MG-CS	MG-ST
Chromosomes	450	350
Genes	06	03
Head sizes	12	08
Linking function	+	+
Mathematical operators	+, −, ×, ÷, √	+, −, ×, ÷, √
Function set	Average, square root, tangent, sine, cosine	Average, square root
Constants per gene	10	10
Number of generations	1000	1000

3.3. MEP Optimal Parmeter Settings

Table 7 summarizes the settings of the models for CS and ST that were finalized after running numerous trials. The size and number of the subpopulation are crucial parameters that determine the overall accuracy and complexity of the models. If these parameters have larger values, a model will take more time to converge and yield accurate results. Nonetheless, there is a risk of overfitting and poor performance on unseen data. It should be noted that the models for CS and ST are denoted as MM-CS and MM-ST, respectively. For the MM-CS model, the number of generations was determined by observing the fitness function, and it was found that no substantial improvement occurred beyond 1000 generations, which was considered the optimal value. On the other hand, the MM-ST model did not provide substantial improvement in R beyond 500 generations. Hence, 500 was chosen as the optimal value. In both the models, the mutation rate was set to 0.01, and the crossover rate was set to 0.90. These rates were chosen to ensure that offspring undergo mutation and crossover operations during the modeling process. Moreover, the code length was determined as 40 for both the models. However, the final models were simplified employing basic mathematical rules. Table 7 lists the selected settings for each model, which were determined by testing different combinations of these parameters on the training data.

Table 7. MEP algorithm's parameter settings for both models.

Parameter	MM-CS	MM-ST
Number of subpopulations	40	40
Size of subpopulations	300	250
Code lengths	30	20
Crossover probability	0.90	0.90
Mathematical operators	+, −, ×, ÷, √	+, −, ×, ÷
Mutation probability	0.010	0.01
Tournament sizes	4	3
Operators	0.50	0.50
Variables	0.50	0.50
Number of generations	1000	500

4. Performance Evaluation Criteria for Models

The inclusion of statistical error measures is crucial to evaluate the accuracy and effectiveness of empirical models. These measures ensure that the models are reliable for predicting the properties of GPC. In this research work, error measures such as the mean absolute error (MAE), root mean square error (RMSE), R , and relative root mean square error (RRMSE) were considered, as used commonly in the literature [42]. To determine the accuracy of the proposed models, a statistical study utilizing these measures was conducted, and a performance indicator (ρ) that considered both R and RRMSE was employed [15]. The mathematical formulas for these errors and the acceptable range for an accurate model are demonstrated in Table 8.

Table 8. Summary of statistical parameters and recommended criteria for an accurate empirical model.

Parameter	Expression	Criteria
R	$\frac{\sum_{i=1}^n (e_i - \bar{e}_i)(m_i - \bar{m}_i)}{\sqrt{\sum_{i=1}^n (e_i - \bar{e}_i)^2 \sum_{i=1}^n (m_i - \bar{m}_i)^2}}$	>0.80
MAE	$\frac{\sum_{i=1}^n e_i - m_i }{n}$	Minimum
RMSE	$\sqrt{\frac{\sum_{i=1}^n (e_i - m_i)^2}{n}}$	Minimum
RRMSE	$\frac{1}{ \bar{e} } \sqrt{\frac{\sum_{i=1}^n (e_i - m_i)^2}{n}}$	0–0.10 (Excellent) or 0.11–0.20 (Good)
ρ	$\rho = \frac{RRMSE}{1+R}$	<0.20

In Table 8, n represents the total number of samples, e_i denotes the i^{th} model output, and m_i designates the corresponding experimental output. The mean values of the experimental and model outputs are denoted by \bar{e}_i and \bar{m}_i , respectively. The accuracy of a model is directly proportional to R and inversely proportional to MAE, RMSE, and RRMSE. The value of ρ ranges from 0 to $+\infty$, and as it approaches 0, the accuracy increases. Likewise, the R value ranges from 0 to 1, and an R value of 0.80 is considered indicative of a good correlation between the inputs and output.

5. Results and Discussion

In this section, the modeling results of the GEP and MEP algorithms are separately presented and discussed. We begin by discussing the results obtained from the GEP algorithm, followed by an in-depth analysis of the MEP results.

5.1. Modeling Results of GEP

5.1.1. MG-CS

To ensure reliability of the model, it is advisable that the ratio of data points to input variables be greater than three [43]. Alternatively, reducing the (Kolmogorov) complexity of the data points could facilitate faster network convergence, particularly if trained with a smaller dataset [44]. As a result, in this study, the model had a ratio of 43, displaying a satisfactory sample size. A total of 47 trials were conducted to optimize the model's accuracy and simplify the formulation, with the GEP algorithm generating ETs, as illustrated in Figure 4. The variables used in ETs are defined in Table 9 and were decoded to develop an empirical equation for the f'_c prediction based on the given input. The final simplified equations for predicting CS are given in the following.

$$f'_c \text{ (MPa)} = A + B + C + D + E + F \quad (6)$$

where

$$A = \frac{\tan\left(\frac{FA + NaOH - 4.8}{3}\right)^2 + \sqrt{F_{agg}} + \tan 9.5FA}{3} \quad (7)$$

$$B = \tan\left(SP + \frac{w^5 + FA + F_{agg} + \tan C_{agg}}{3}\right)^{0.25} \quad (8)$$

$$C = \tan\left(\sin\frac{1.28 + \frac{C_{agg}}{FA} + FA - 5.59F_{agg}}{4}\right)^5 \quad (9)$$

$$D = \left(FA + (\tan 2.37C_{agg} + 4)^{0.66}\right)^{0.33} \quad (10)$$

$$E = \frac{\tan F_{agg} - 0.54NaOH + \tan NaOH + 17 - SP}{12} + 68 \quad (11)$$

$$F = \left(\frac{FA + 2NaOH + 10C_{agg} \times SP + NaSi^2 - 35115}{9}\right)^{0.33} \quad (12)$$

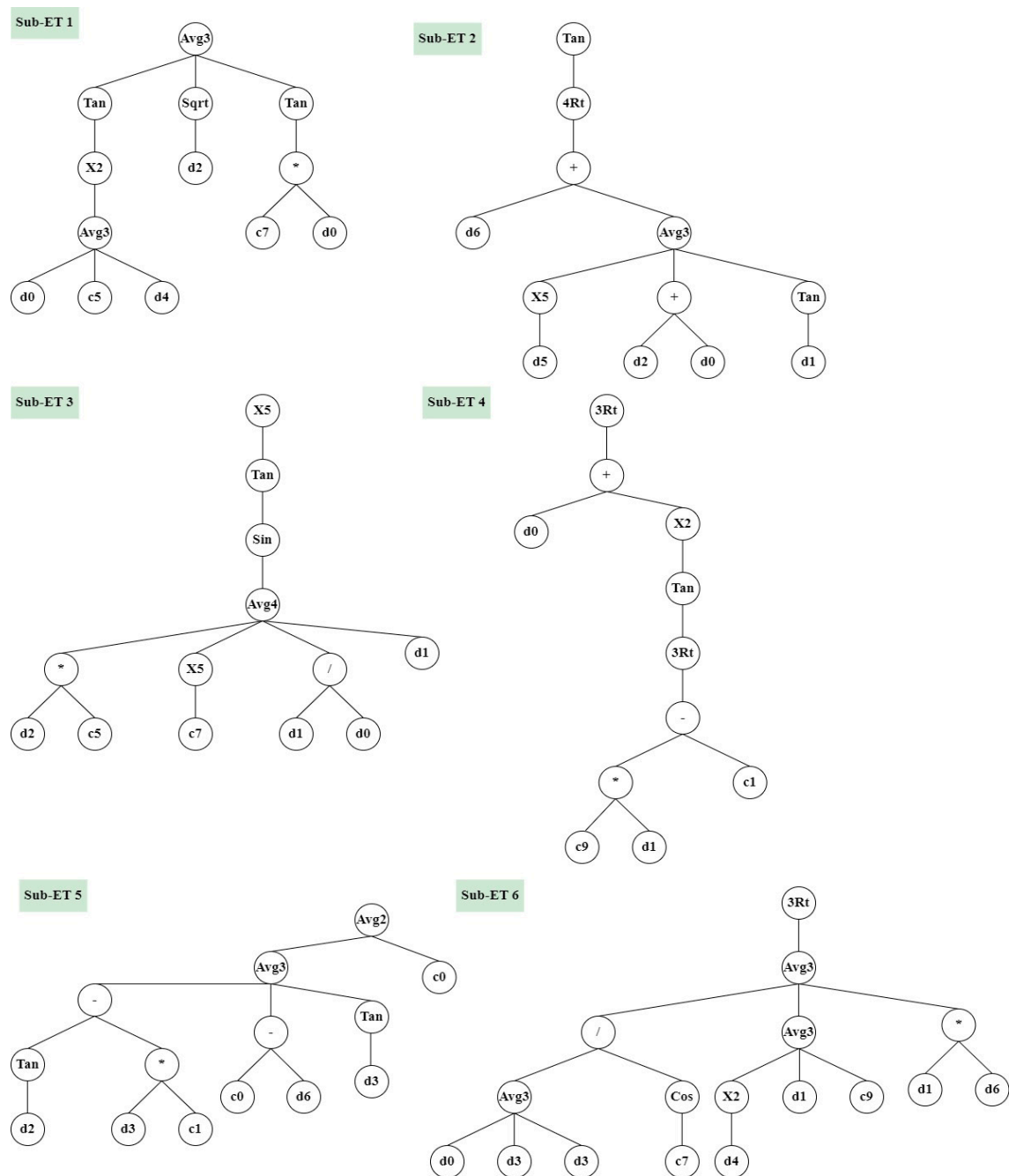


Figure 4. ETs of MG-CS model. Note: Signs '-', '*', and '/' refer to minus, multiplication, and division, respectively.

Table 9. Description of input variables presented in ETs of MG-CS model.

Variable	d0	d1	d2	d3	d4	d5	d6
Corresponding input	<i>FA</i>	<i>C_{agg}</i>	<i>F_{agg}</i>	<i>NaOH</i>	<i>NaSi</i>	<i>w</i>	<i>SP</i>

The modeled and experimental results are indicated in Figure 5, along with the linear fit regression trend lines for both the training and validation sets. The accuracy of the developed model can be reliably assessed by the slope of the regression line. It can be observed that the slope of the regression line was approximately 0.80 and 0.75 for the training and validation sets, respectively. Moreover, *R* for both sets was considerably high,

with the values of 0.89 and 0.83 for the training and validation sets, respectively. According to Table 8, an R value greater than 0.80 demonstrates a highly accurate model. These results pointed out that the model performed well not only on the training data but also on the validation or unseen data.

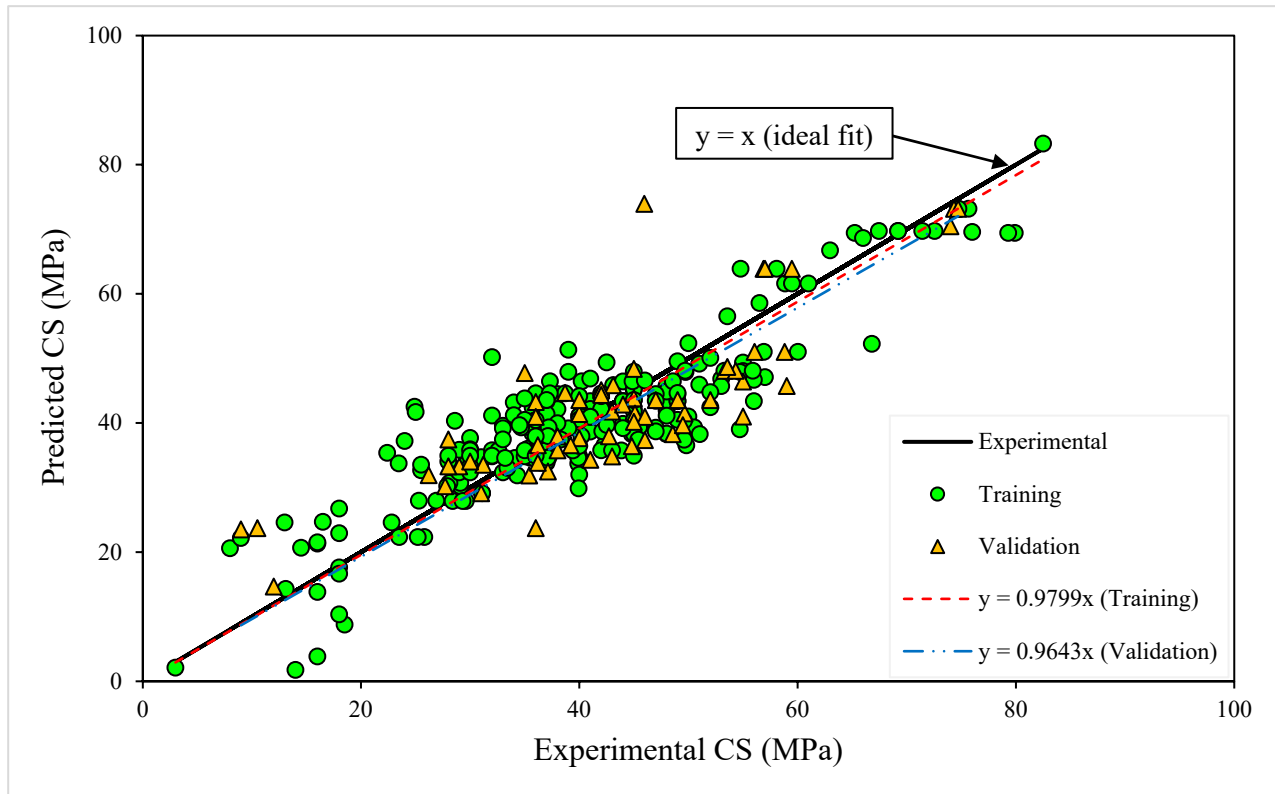


Figure 5. Comparison of experimental and predicted values of MG-CS model.

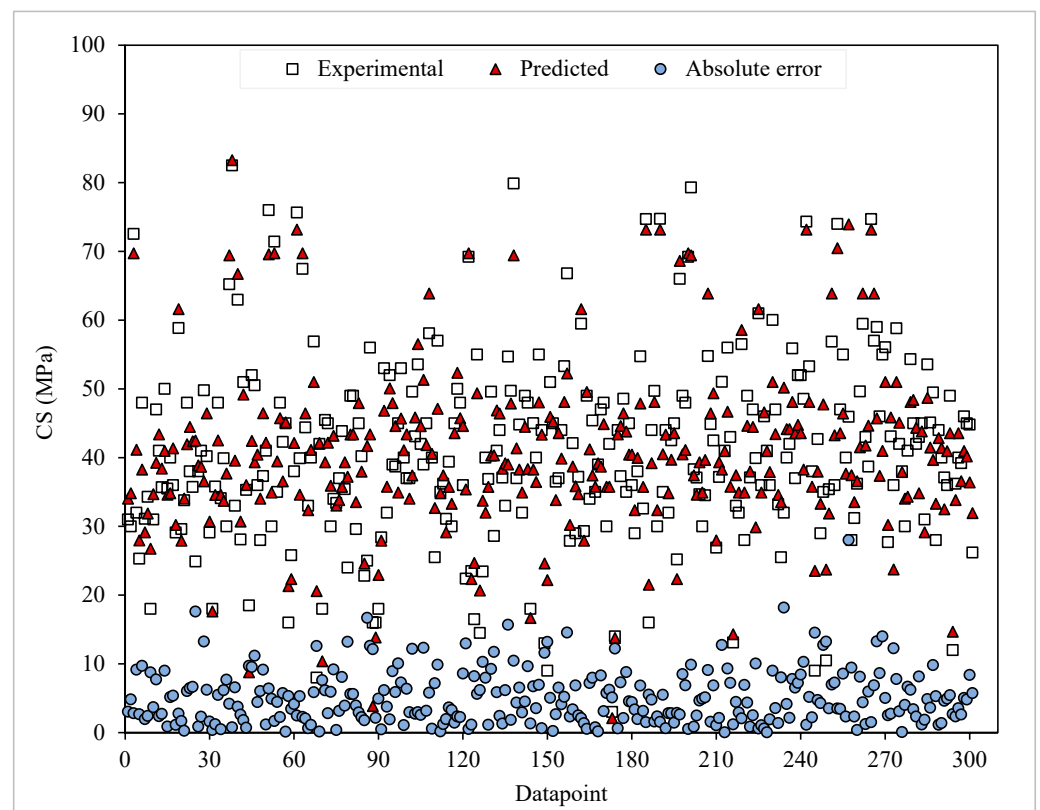
The performance of the model on the training and validation data can be analyzed by looking at various metrics summarized in Table 10. For the training data, the R value was found to be 0.89, exhibiting a strong linear relationship between the actual and predicted values. MAE and RMSE were 4.88 and 6.16, respectively. RRMSE was calculated to be 0.15. These results illustrated that the model performed well on the training data with low error and high accuracy.

On the other hand, for the validation data, the R value was resulted as 0.83, displaying a slightly weaker linear relationship than the training data. MAE and RMSE were 5.82 and 7.39, respectively, which were higher than the training set. RRMSE was found to be 0.17. These results showed that the model had higher error and lower accuracy on the validation data compared with the training data. RMSE increased by approximately 20% in the validation data, signifying that the model had a marginally higher error rate in predicting unseen data. Despite this, the RRMSE values indicate that the model still had good accuracy in predicting both sets of the data, with values below 0.20 for both sets. Based on the performance index, it can be concluded that the GEP model performed similarly for both the training and validation data. The performance index is a comprehensive measure that considers both the accuracy and complexity of the model. A lower value of the performance index demonstrates a better-performing model. In this case, the performance index was 0.08 for the training data and 0.09 for the validation data, which exhibited that the model performed similarly on both sets. This suggests that the model is generalizable and can be used to make predictions on unseen data with reasonable accuracy.

Table 10. Statistical indicators for training and validation sets of MG-CS model.

Model	Dataset	<i>R</i>	MAE	RMSE	RRMSE	ρ
MG-CS	Training	0.89	4.88	6.16	0.15	0.08
	Validation	0.83	5.82	7.39	0.17	0.09

The accuracy of the predicted results is further visualized through the absolute error plot in Figure 6, which compares the experimental and predicted data points. The absolute difference between the two sets was depicted to be noticeably low, with more than 80% of the datasets having an absolute difference of less than 1.50 MPa. Only 5 out of 301 data points gave an error greater than 10 MPa. These results further reinforced the claim that the GEP model could accurately predict CS of GPC.

**Figure 6.** Absolute errors between experimental and predicted values of MG-CS model.

5.1.2. MG-ST

In a similar fashion to the CS modeling, ST of GPC was modeled utilizing the GEP approach. The model considered input parameters such as *FA*, *GGBFS*, *NaOH*, *C_{agg}*, and water content. The resulting formulation obtained from the GEP modeling is expressed in Equation (13):

$$f'_{st}(\text{MPa}) = \frac{2C_{agg} + 2.58w}{C_{agg} + FA} + \frac{GGBFS}{w + NaOH - GGBFS + 15} + \frac{4GGBFS - 0.84}{4} \quad (13)$$

Figure 7 presents a comparison of the experimental and modeled results for ST of GPC, alongside the regression trend lines for the training and validation datasets. The *R* values for both the datasets were remarkably high, with the values of 0.87 and 0.82 for the training and validation datasets, respectively. These results illustrated that the model had a good performance not only on the training data but also on unseen data in the form of the validation dataset.

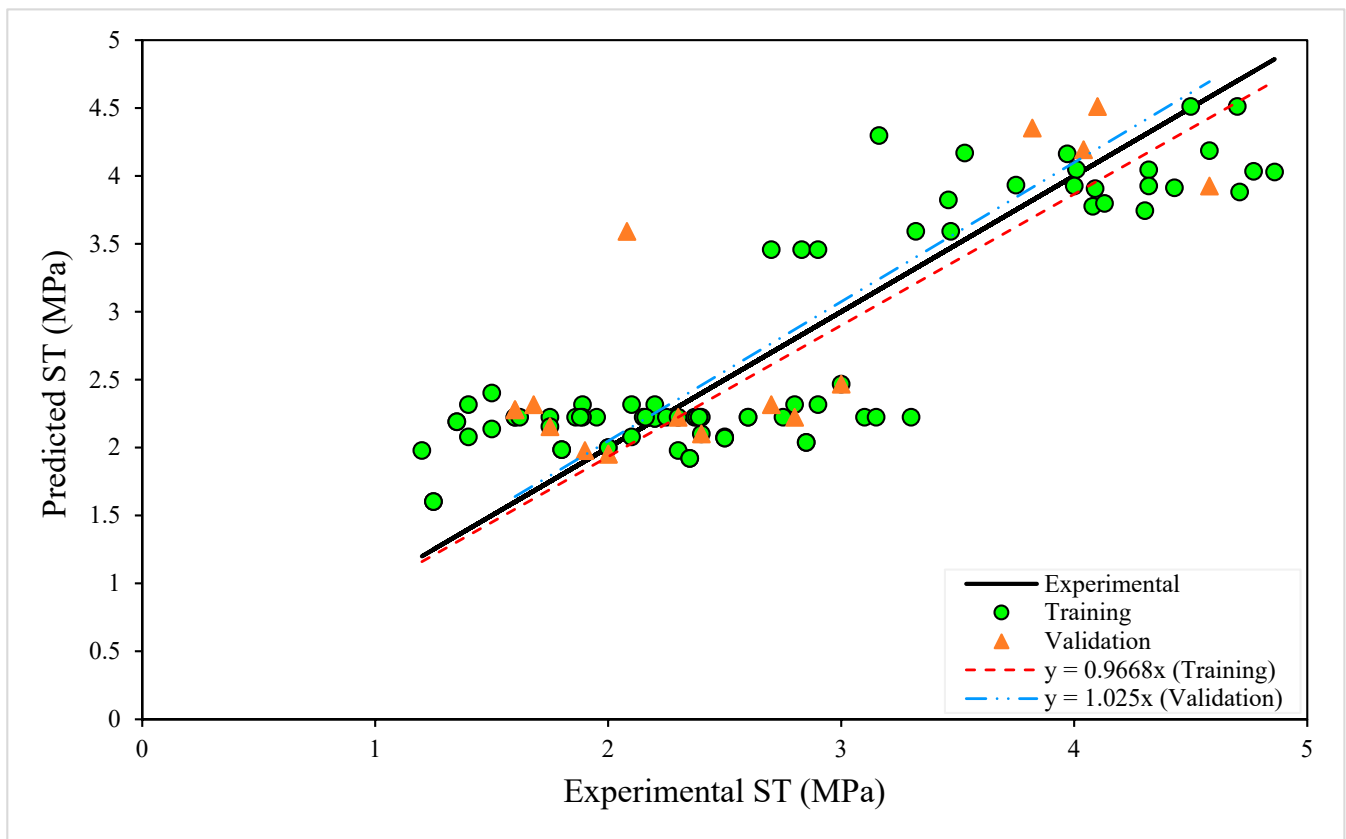


Figure 7. Comparison of experimental and predicted values of MG-ST model.

To evaluate the accuracy of the predicted results for the MG-ST model, an absolute error plot was generated in Figure 8, which compares the experimental and predicted data points. The plot depicts that the maximum error observed in the data was 1.50 MPa, and MAE was found to be 0.42 MPa, which was relatively small. These results underscored the dependability of the GEP model in providing precise predictions for ST of GPC.

The performance of the ST model was also assessed via the same metrics as the CS model. The results of both the training and validation datasets are provided in Table 11. For the training dataset, the R value was 0.87, implying a strong positive relationship between the predicted and actual values. MAE was 0.42, which means that on average, the predicted value was 0.42 MPa away from the actual value. RMSE was 0.51 MPa, showing an average deviation of 0.51 MPa between the predicted and actual values. RRMSE was 0.19, which was relatively low, suggesting that the model had a low relative error. ρ was 0.10, which was within the acceptable range of 0.10 or less.

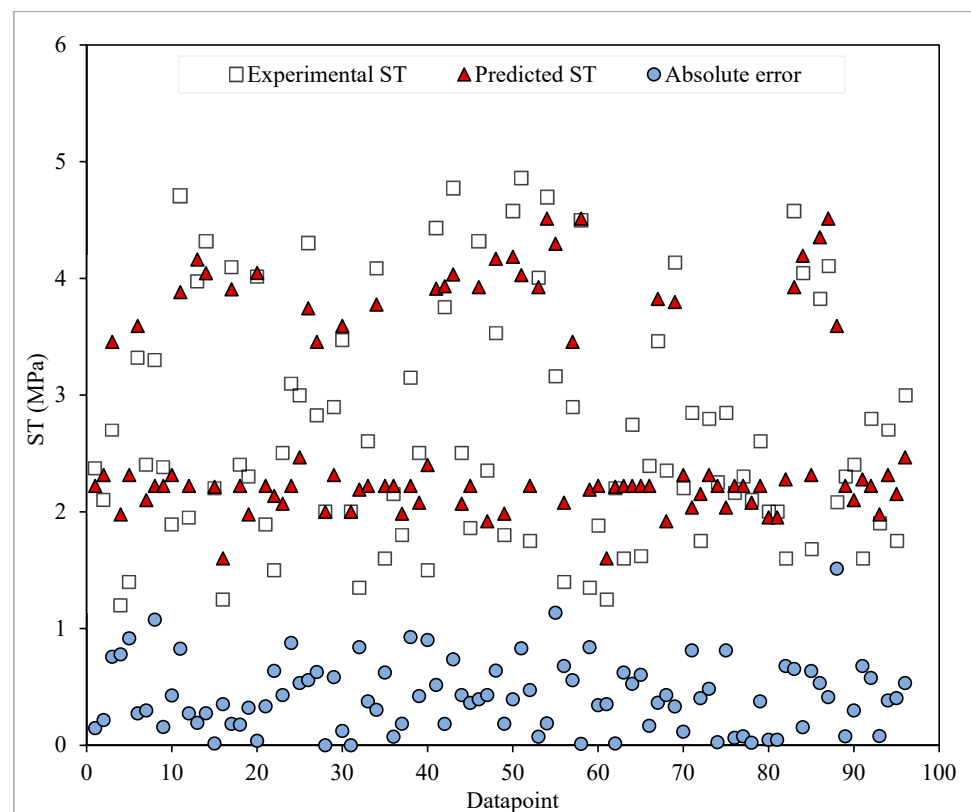


Figure 8. Absolute errors between experimental and predicted values of MG-ST model.

Table 11. Statistical measures for training and validation sets of MG-ST model.

Model	Dataset	<i>R</i>	MAE	RMSE	RRMSE	ρ
MG-ST	Training	0.87	0.42	0.51	0.19	0.10
	Validation	0.82	0.45	0.57	0.22	0.12

For the validation dataset, the *R* value was 0.82, which was a little lower than that of the training dataset, but still pointed out a strong positive relationship between the predicted and actual values. MAE was 0.45 MPa, indicating that on average, the predicted value was 0.45 MPa away from the actual value. RMSE was 0.57 MPa, which was slightly higher than that of the training dataset. RRMSE was 0.22, which was higher than that of the training dataset, demonstrating that the relative error was a bit higher. The performance index was 0.12, which was higher than that of the training dataset. Compared with the training data, the validation set of the ST model demonstrated an increase of approximately 13.10% in RMSE and 8.10% in MAE. However, both the training and validation datasets exhibited similar performance for the ST model, as evidenced by the comparable values of the performance index, which was considered the best indicator of the overall performance. The performance index for both datasets was within the acceptable range of 0.10 or less, suggesting that the ST model was suitable for predicting ST of GPC.

5.2. Modeling Results of MEP

5.2.1. MM-CS

Figure 9 displays a comparative analysis of the prediction results acquired from the MM-CS model when compared with the experimental data. The datasets utilized for the GEP modeling were also employed for the MEP modeling. The statistical parameters for the training and validation datasets are presented in Table 12, allowing for a comprehensive evaluation. The findings in Figure 9 revealed a weak correlation compared with the MG-CS

model (Figure 5) as evident from a comparison of the slope of the regression lines (1 for an ideal model). Furthermore, the formulations for predicting CS using MEP are listed in Table A1 of Appendix A.

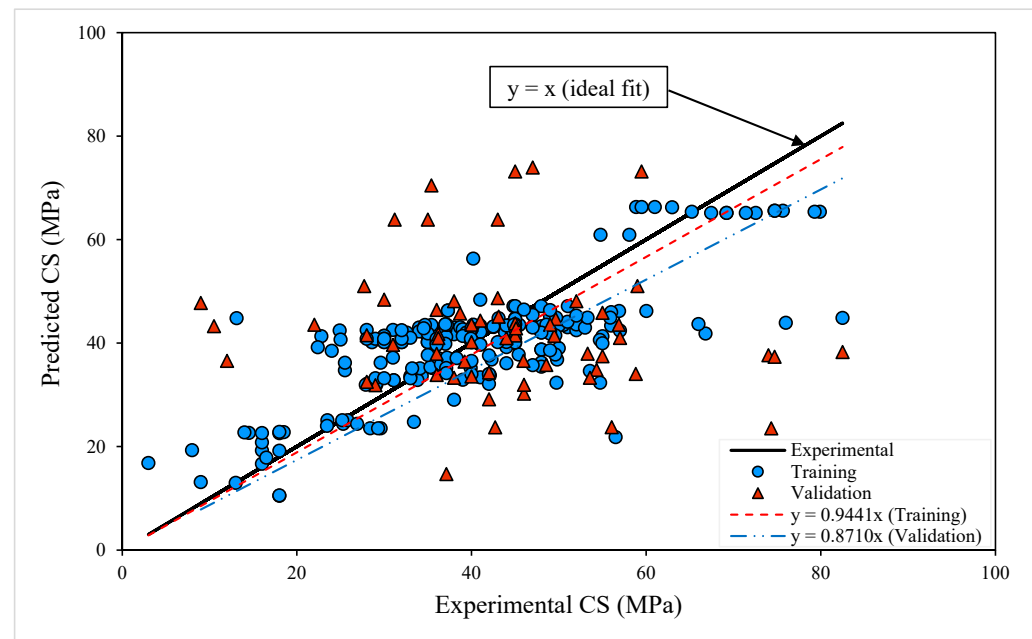


Figure 9. Comparison of experimental and predicted values of MM-CS model.

Table 12. Statistical indicators for training and validation sets of MM-CS model.

Model	Dataset	<i>R</i>	MAE	RMSE	RRMSE	ρ
MM-CS	Training	0.76	6.72	8.92	0.22	0.13
	Validation	0.75	6.97	9.45	0.22	0.12

Figure 10 depicts a graphical representation of the absolute errors for the MM-CS results. MAE was 6.78 MPa, which was higher compared with 5.09 MPa for the MG-CS model. In addition, the maximum absolute error noted was >37 MPa, while it was <28 MPa for the MG-CS model. The aforementioned discussion demonstrates the accuracy of the proposed GEP formulation.

As discussed above, Table 12 shows the statistical indicators for the training and validation sets of the MM-CS model. The analysis of the statistical parameters indicates that the MM-CS model exhibited the highest accuracy compared with MG-CS, as evidenced by its lowest *R* values for the training and validation sets (Table 10). It is also important to note that the values of other parameters were comparatively low for the MG-CS model, suggesting better performance of the model compared with MM-CS. Additionally, the values of MAE and RMSE were close to each other in both sets, displaying the good generalization and high predictability of the MG-CS model (Table 10). On the other hand, the values of ρ were close to zero for the model MG-CS, while the values of RRMSE were <0.20, revealing that the model could be termed as “good”.

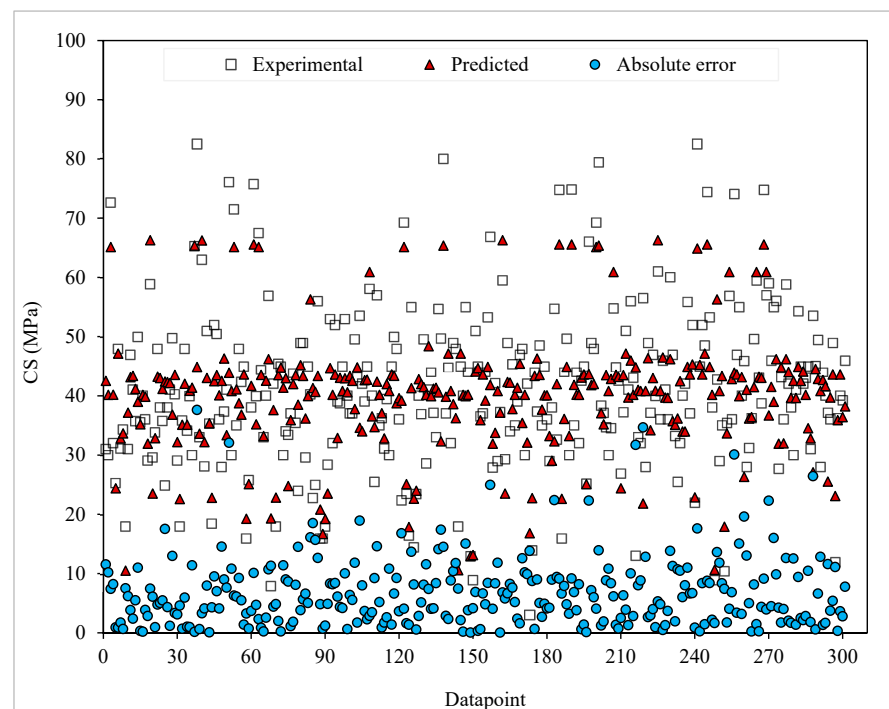


Figure 10. Absolute errors between experimental and predicted values of MM-CS model.

5.2.2. MM-ST

The results of the MM-ST model (both the training and validation sets) against the experimental results are illustrated in Figure 11. The figure also depicts the slope of regression lines for both sets. The slopes for the training and validation sets were 0.96 and 0.99, respectively, demonstrating a very good correlation between the experimental and model values. The values of the slope of the regression lines were comparable to the MG-ST model (Figure 7); however, it cannot be considered a sole criterion to assess the model performance. The absolute errors of the MM-ST model are shown in Figure 12, providing additional insights. Also, the formulations for predicting ST using MEP are presented in Table A2 of Appendix A.

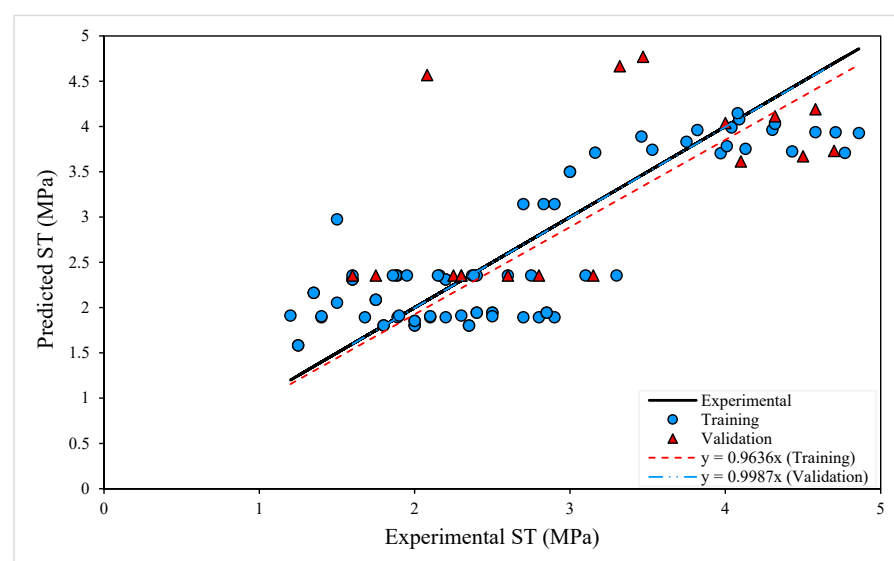


Figure 11. Comparison of experimental and predicted values of MM-ST model.

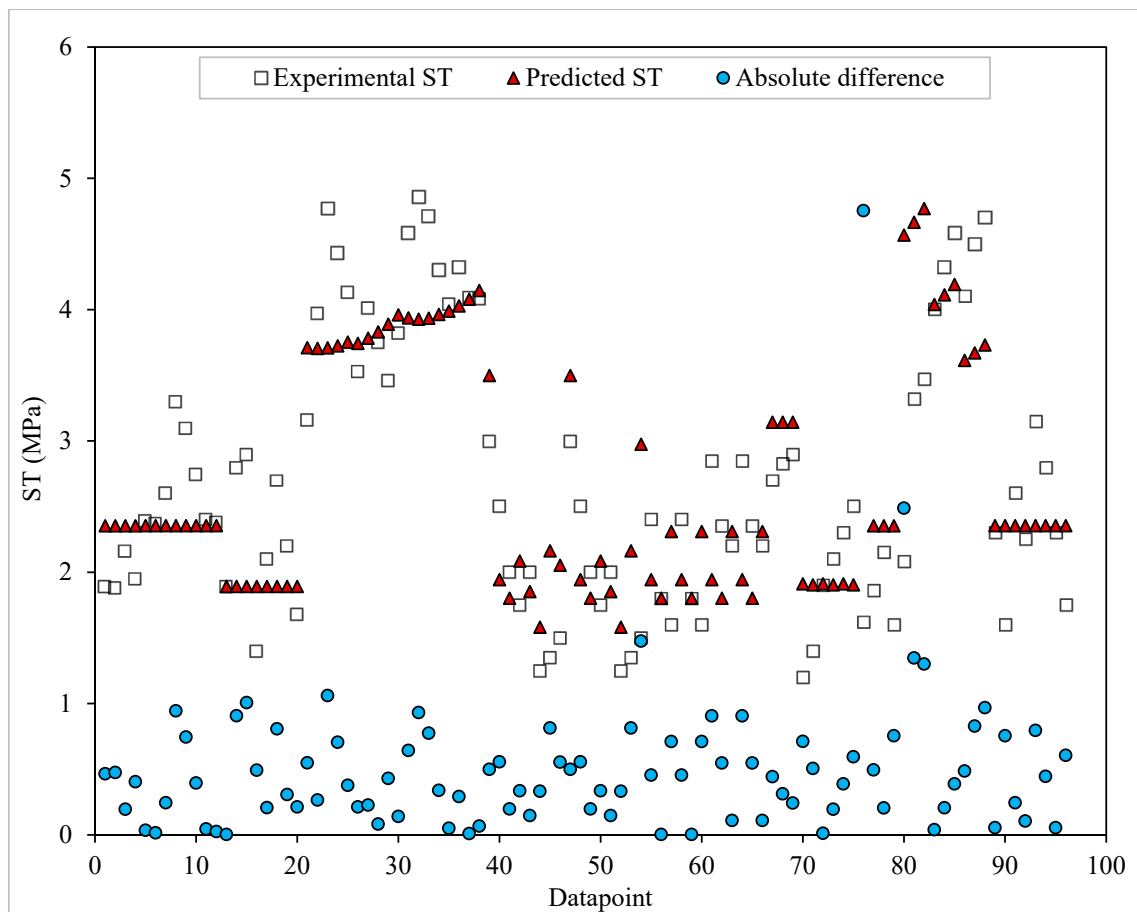


Figure 12. Absolute errors between experimental and predicted values of MM-ST model.

Figure 12 indicates that the experimental and predicted values are close to each other. MAE was 0.51 MPa, while the maximum value was 4.75 MPa. For the MG-ST model, MAE was 0.42 MPa, while the maximum value was 1.08 MPa. Furthermore, the cumulative sum of the absolute difference values was 49.07 MPa for MM-ST, whereas for MG-ST, it was 40.77 MPa. These findings exhibited the excellent performance of the GEP model.

Table 13 displays the values of the various statistical parameters chosen for the analysis. The R values for the training and validation sets were 0.73 and 0.70, respectively, both falling below the recommended criterion of 0.80, as stated in Table 8. The R values for both sets were on the lower side compared with the values of MG-ST (Table 11). Moreover, the values of the remaining parameters, i.e., MAE, RRMSE, RMSE, and ρ , were very high compared with MG-ST, elaborating the poor performance of the MEP algorithm in this case. Furthermore, the values of the parameters for both sets were not close to each other as well as compared with the MG-ST model.

Table 13. Statistical parameters for training and validation sets of MM-ST model.

Model	Dataset	R	MAE	RMSE	RRMSE	ρ
MM-ST	Training	0.73	0.48	0.75	0.29	0.17
	Validation	0.70	0.65	0.89	0.28	0.18

6. Parametric Analysis

Based on the mentioned analysis, the GEP-based models are finalized for predicting CS and ST of GPC. In this regard, the GEP-based prediction models developed were further validated through a parametric analysis utilizing a Python code. The average values of

all the input parameters were fixed, and the effect of varying one of the inputs on the mechanical properties was plotted, as depicted in Figures 13 and 14.

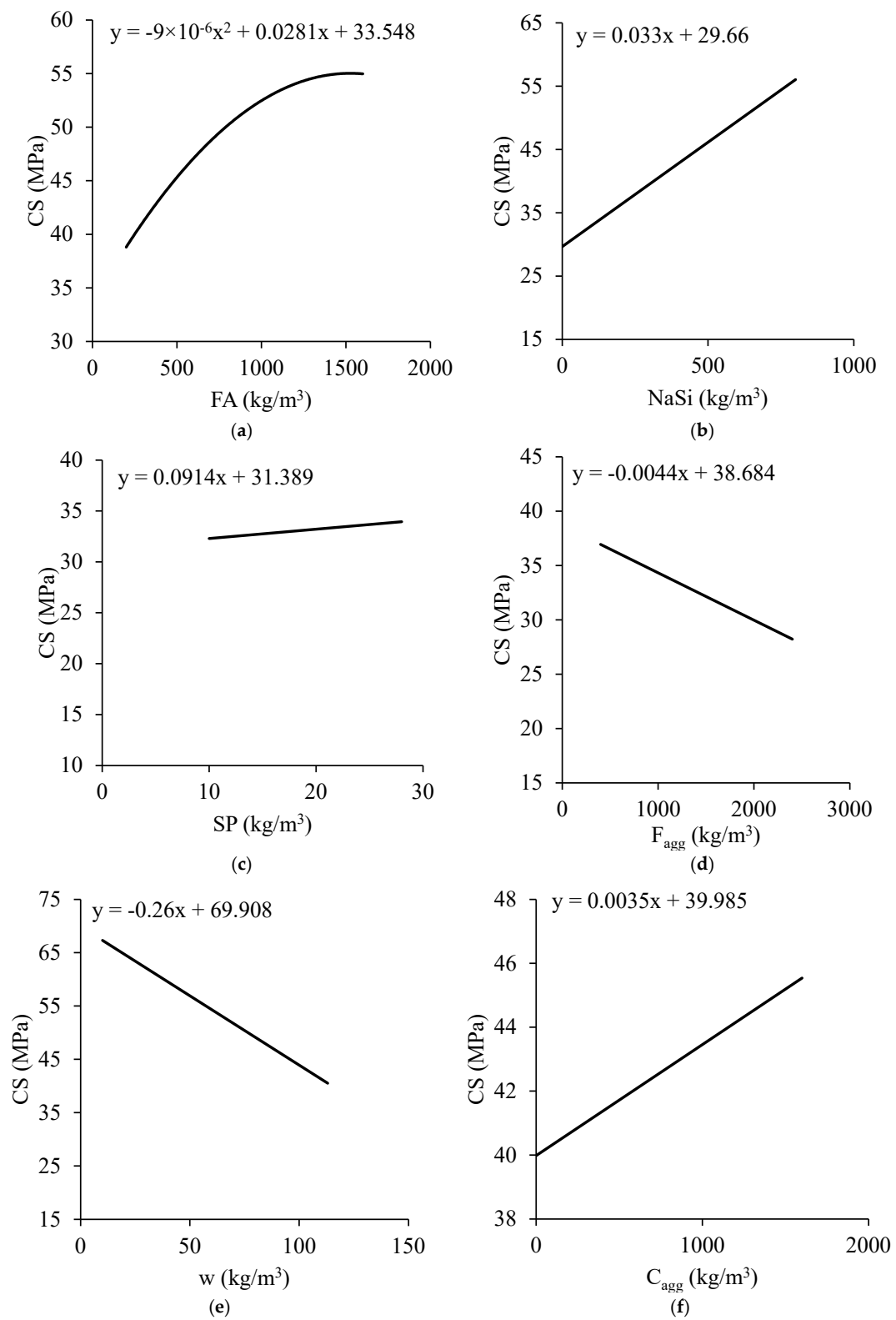


Figure 13. Parametric analysis indicating variations in CS with: (a) FA, (b) NaSi, (c) SP, (d) F_{agg} , (e) w , and (f) C_{agg} .

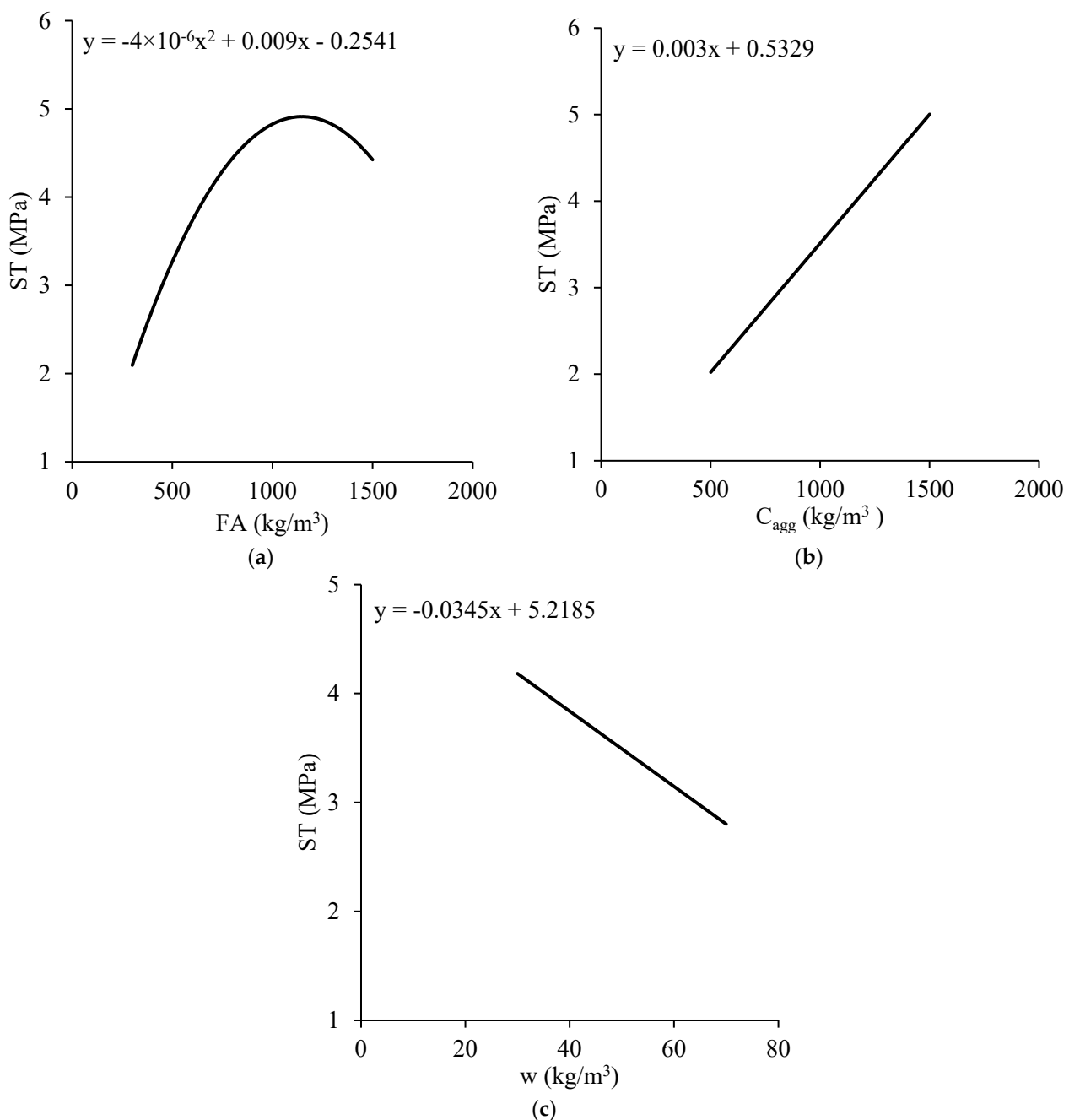


Figure 14. Parametric analysis showing variations in ST with: (a) FA, (b) C_{agg} , and (c) w .

Generally, an increase in the FA content results in an increase in CS because of the pozzolanic reaction. This reaction leads to the formation of more calcium silicate hydrate (CSH) gel. The trend observed in this study illustrated that CS of GPC increased initially with increasing the FA content, but after reaching an optimum point, it began to decrease, as depicted in Figure 13a. This decrease could be attributed to the reduced workability of the mix resulting from an increase in the FA content, leading to the improper compaction and weaker interfacial bonding between aggregates and paste. The activator, which is typically *NaSi*, plays a crucial role in the development of the strength in GPC. As the amount of activator was increased, there was a noticeable enhancement in the strength of GPC, as shown in Figure 13b. This was because the activator helped initiate the reaction between the alkaline solution and FA. This reaction led to the formation of a geopolymer gel that bound the particles together. Therefore, a higher amount of activator could facilitate a

more complete reaction, resulting in higher strength. However, it is important to note that beyond a certain point, increasing the amount of activator might not lead to further improvements in the strength and might even have a negative impact. The effect of *SP* on CS of GPC was also investigated. The results indicated that the addition of *SP* did not have a substantial effect on CS of GPC, as expressed in Figure 13c. This suggests that the use of *SP* in GPC mixtures may not be necessary and only impacts the workability of concrete.

The F_{agg} content in GPC was observed to have a significant impact on its CS. It was seen that as the amount of F_{agg} was increased, the strength of GPC decreased linearly, as exhibited in Figure 13d. This trend can be attributed to the fact that the F_{agg} components had a higher water absorption capacity and a lower specific gravity compared with the other components in the mix [45]. As a result, an increase in the F_{agg} content led to higher water demand and, subsequently, a weaker interfacial transition zone (ITZ) between aggregates and geopolymer matrix [46]. This weaker ITZ resulted in a lower CS of GPC. The impact of the water content on the strength of GPC displayed in Figure 13e implies that an increase in the water content led to a decrease in the strength of GPC. This can be ascribed to the fact that excess water content in the mix reduced the strength of the cementitious matrix and increased the porosity of GPC, which in turn reduced its strength.

The trend of C_{agg} on CS of GPC is plotted in Figure 13f, indicating that the strength of GPC increased linearly with an increase in the content of C_{agg} . This is likely owing to the fact that increasing the amount of C_{agg} led to better particle packing and improved interlocking, which resulted in higher strength. However, it should be mentioned that beyond a certain point, an increase in the C_{agg} content might lead to a decrease in the strength because of the reduced workability and increased void content. Therefore, careful optimization of the amount of C_{agg} is necessary to achieve the highest possible strength.

The combined impact of FA , C_{agg} , and water content on ST of GPC is also noteworthy, as summarized in Figure 14a–c. Similar to CS, ST initially increased with an increase in the FA and C_{agg} contents up to a certain optimal level. However, ST decreased as the water content increased beyond a particular level. These trends highlight the importance of balancing the quantities of FA , C_{agg} , and water content in order to achieve the desired ST strength in GPC. It is essential to carefully consider the optimal proportions of these ingredients during the mixture design stage to achieve the desired strength properties of GPC. The GEP model demonstrated a high degree of accuracy in capturing the correlation between the input parameters and mechanical properties of GPC. The regression trend lines and absolute error plot depict that the GEP model's predicted results were in close agreement with the experimental data. These findings suggest that the GEP models can be a reliable tool for predicting CS and ST of GPC, which can help optimize the material's composition and design more durable structures.

7. Conclusions

This article presented a novel approach to developing accurate and reliable models for CS and ST of GPC utilizing the GEP and MEP algorithms. A large database was collated from the published literature, which contained the input parameters used in the development of the models. The performance of the developed models was evaluated employing various statistical measures such as R , MAE, RMSE, and ρ . Also, the absolute error analysis was conducted to calculate the mean, maximum, and minimum errors between the experimental and predicted values. The main conclusions drawn from the above study are listed below.

The main conclusions for the CS model are in the following:

- In the training phase, the MG-CS model exhibited superior reliability and accuracy with an R value of 0.89, noticeably outperforming the MEP-based model (R of 0.76). The GEP model also illustrated lower MAE (4.88), RMSE (6.16), RRMSE (0.15), and ρ (0.08) values, indicating its robust predictive capabilities.

- During the validation phase, the MG-CS model maintained a high R value of 0.83, with corresponding lower MAE, RMSE, RRMSE, and ρ values of 5.82, 7.39, 0.17, and 0.09, respectively, further validating its predictive accuracy and reliability.
- The absolute difference between the experimental and predicted sets was demonstrated to be considerably low, with more than 80% of the datasets having an absolute difference of less than 1.50 MPa in case of GEP. MAE for the GEP-based model was 5.09 MPa for both sets, which outperformed the MEP-based model (MAE of 6.78 MPa). These results further reinforced the claim that the GEP model provided sufficiently accurate predictions for CS of GPC.

The main conclusions for the ST models are in the following:

- In the training phase, the MG-ST model displayed superior performance with an R value of 0.87, surpassing the MEP-based models (R of 0.73). Similar to the CS model, MG-ST provided lower MAE (0.42), RMSE (0.51), RRMSE (0.19), and ρ (0.10) values, illustrating its efficacy in predicting ST.
- During validation, the MG-ST model maintained a high R value of 0.82, with corresponding MAE, RMSE, RRMSE, and ρ values of 0.45, 0.57, 0.22, and 0.12, respectively, underscoring its reliability and accuracy in predicting ST.
- Based on the absolute difference between the experimental and predicted sets, MAE for the GEP-based model was 0.42 MPa for both sets, surpassing the MEP-based model's MAE of 0.51 MPa, which reinforced the assertion that the GEP model exhibited accurate predictions for ST of GPC.

This study also involved deriving empirical equations for the MG-CS and MG-ST models. A parametric analysis of the proposed empirical equations was done specifically for both models, demonstrating that the models effectively accounted for the system being studied. Moreover, the suggested empirical equations made a solid and accurate basis for enhancing the application of ML methods in predicting f'_c and f'_{st} of GPC using a simple scientific calculator. Overall, this study provides a valuable contribution to sustainable construction by reducing reliance on conventional cement-based concrete and promoting the application of industrial waste materials in the production of GPC.

Author Contributions: B.H.: conceptualization, methodology, validation, formal analysis, investigation, writing—original draft preparation, and writing—review and editing. A.B.: conceptualization, methodology, validation, formal analysis, investigation, resources, writing—original draft preparation, and writing—review and editing. M.F.J.: conceptualization, validation, formal analysis, investigation, writing—original draft preparation, and writing—review and editing. I.A.: validation, investigation, writing—original draft preparation, and writing—review and editing. M.A.I.: validation, formal analysis, writing—original draft preparation, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are thankful for the financial support from the Guangxi Key R&D Plan Project [Grant No. AB22080083] and the Guangxi Key R&D Plan Project [Grant No. AB23026065].

Data Availability Statement: Data will be made available on request.

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

Appendix A

Table A1. MEP-based model used to calculate f'_c . Value of each “prg” is calculated individually utilizing defined input variables. f'_c is equal to final prg, i.e., prg(28).

$f'_c = \text{prg}(28)$
where, $x[0] = \text{FA}$; $x[1] = \text{Cagg}$; $x[2] = \text{Fagg}$; $x[3] = \text{NaOH}$; $x[4] = \text{NaSi}$; $x[5] = w$; $x[6] = \text{SP}$
$\text{prg}[0] = x[2]$; $\text{prg}[1] = \text{prg}[0] + \text{prg}[0]$; $\text{prg}[2] = x[0]$; $\text{prg}[3] = \text{prg}[0] + \text{prg}[1]$;

Table A1. Cont.

prg[4] = sqrt(prg[0]);
prg[5] = prg[3] + prg[2];
prg[6] = x[3];
prg[7] = x[6];
prg[8] = sqrt(prg[4]);
prg[9] = sqrt(prg[0]);
prg[10] = x[5];
prg[11] = prg[6] + prg[10];
prg[12] = x[1];
prg[13] = x[4];
prg[14] = prg[6] + prg[13];
prg[15] = prg[2] − prg[13];
prg[16] = prg[13] − prg[15];
prg[17] = prg[12] − prg[14];
prg[18] = prg[16] − prg[17];
prg[19] = prg[15]/prg[8];
prg[20] = prg[17] − prg[1];
prg[21] = x[5];
prg[22] = prg[11]/prg[9];
prg[23] = prg[7] − prg[22];
prg[24] = prg[21] * prg[18];
prg[25] = prg[23] − prg[22];
prg[26] = prg[19] + prg[24];
prg[27] = prg[20]/prg[26];
prg[28] = prg[25] + prg[27];

Table A2. MEP-based model used to calculate f'_{st} . f'_{st} is equal to final “prg”, i.e., prg(20).

$f'_{st} = \text{prg}(20)$
where, x[0] = Cagg; x[1] = Fagg; x[2] = FA; x[3] = w; x[4] = GGBFS; x[5] = NaOH; x[6] = NaSi
prg[0] = x[1];
prg[1] = x[2];
prg[2] = x[0];
prg[3] = prg[2] + prg[1];
prg[4] = x[7];
prg[5] = prg[0] + prg[4];
prg[6] = prg[5] + prg[3];
prg[7] = x[4];
prg[8] = x[5];
prg[9] = x[3];
prg[10] = prg[7] − prg[4];
prg[11] = prg[7]/prg[3];
prg[12] = prg[10] + prg[0];
prg[13] = prg[5] − prg[8];
prg[14] = prg[9]/prg[13];
prg[15] = prg[6]/prg[12];
prg[16] = prg[12] + prg[12];
prg[17] = prg[10]/prg[16];
prg[18] = prg[15] + prg[11];
prg[19] = prg[18] − prg[17];
prg(20) = prg[19] − prg[14];

References

- He, H.; Qiao, H.; Sun, T.; Yang, H.; He, C. Research progress in mechanisms, influence factors and improvement routes of chloride binding for cement composites. *J. Build. Eng.* **2024**, *86*, 108978. [\[CrossRef\]](#)
- Iqbal, M.F.; Liu, Q.; Azim, I.; Zhu, X.; Yang, J.; Javed, M.F.; Rauf, M. Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *J. Hazard. Mater.* **2020**, *384*, 121322. [\[CrossRef\]](#) [\[PubMed\]](#)
- Imtiaz, L.; Kashif-ur-Rehman, S.; Alaloul, W.S.; Nazir, K.; Javed, M.F.; Aslam, F.; Musarat, M.A. Life cycle impact assessment of recycled aggregate concrete, geopolymer concrete, and recycled aggregate-based geopolymer concrete. *Sustainability* **2021**, *13*, 13515. [\[CrossRef\]](#)
- Li, Z.; Gao, M.; Lei, Z.; Tong, L.; Sun, J.; Wang, Y.; Wang, X.; Jiang, X. Ternary cementless composite based on red mud, ultra-fine fly ash, and GGBS: Synergistic utilization and geopolymerization mechanism. *Case Stud. Constr. Mater.* **2023**, *19*, e02410. [\[CrossRef\]](#)
- Wang, X.; Li, L.; Xiang, Y.; Wu, Y.; Wei, M. The influence of basalt fiber on the mechanical performance of concrete-filled steel tube short columns under axial compression. *Front. Mater.* **2024**, *10*, 1332269. [\[CrossRef\]](#)
- Singh, A.; Wang, Y.; Zhou, Y.; Sun, J.; Xu, X.; Li, Y.; Liu, Z.; Chen, J.; Wang, X. Utilization of antimony tailings in fiber-reinforced 3D printed concrete: A sustainable approach for construction materials. *Constr. Build. Mater.* **2023**, *408*, 133689. [\[CrossRef\]](#)
- Singh, N.B.; Kumar, M.; Rai, S. Geopolymer cement and concrete: Properties. *Mater. Today Proc.* **2020**, *29*, 743–748. [\[CrossRef\]](#)
- Aneja, S.; Sharma, A.; Gupta, R.; Yoo, D.-Y. Bayesian regularized artificial neural network model to predict strength characteristics of fly-ash and bottom-ash based geopolymer concrete. *Materials* **2021**, *14*, 1729. [\[CrossRef\]](#) [\[PubMed\]](#)

9. Amran, M.; Debbarma, S.; Ozbakkaloglu, T. Fly ash-based eco-friendly geopolymers concrete: A critical review of the long-term durability properties. *Constr. Build. Mater.* **2021**, *270*, 121857. [\[CrossRef\]](#)
10. Rossow, M. *Fly Ash Facts for Highway Engineers*; Continuing Education and Development: New York, NY, USA, 2003.
11. Gomaa, E.; Sargon, S.; Kashosi, C.; Gheni, A.; ElGawady, M.A. Mechanical properties of high early strength class C fly Ash-based alkali activated concrete. *Transp. Res. Rec.* **2020**, *2674*, 430–443. [\[CrossRef\]](#)
12. Bai, B.; Chen, J.; Bai, F.; Nie, Q.; Jia, X. Corrosion effect of acid/alkali on cementitious red mud-fly ash materials containing heavy metal residues. *Environ. Technol. Innov.* **2024**, *33*, 103485. [\[CrossRef\]](#)
13. Long, X.; Mao, M.; Su, T.; Su, Y.; Tian, M. Machine learning method to predict dynamic compressive response of concrete-like material at high strain rates. *Def. Technol.* **2023**, *23*, 100–111. [\[CrossRef\]](#)
14. She, A.; Wang, L.; Peng, Y.; Li, J. Structural reliability analysis based on improved wolf pack algorithm AK-SS. *Structures* **2023**, *57*, 105289. [\[CrossRef\]](#)
15. Azim, I.; Yang, J.; Javed, M.F.; Iqbal, M.F.; Mahmood, Z.; Wang, F.; Liu, Q.-F. Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. *Structures* **2020**, *25*, 212–228. [\[CrossRef\]](#)
16. 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. [\[CrossRef\]](#)
17. Khan, M.A.; Zafar, A.; Akbar, A.; Javed, M.F.; Mosavi, A. Application of gene expression programming (GEP) for the prediction of compressive strength of geopolymers concrete. *Materials* **2021**, *14*, 1106. [\[CrossRef\]](#)
18. Chu, H.-H.; Khan, M.A.; Javed, M.; Zafar, A.; 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 geopolymers concrete. *Ain Shams Eng. J.* **2021**, *12*, 3603–3617. [\[CrossRef\]](#)
19. Khan, M.A.; Memon, S.A.; Farooq, F.; Javed, M.F.; Aslam, F.; Alyousef, R. Compressive strength of fly-ash-based geopolymers concrete by gene expression programming and random forest. *Adv. Civ. Eng.* **2021**, *2021*, 6618407. [\[CrossRef\]](#)
20. Khan, M.A.; Zafar, A.; Farooq, F.; Javed, M.F.; Alyousef, R.; Alabduljabbar, H. Geopolymers concrete compressive strength via artificial neural network, adaptive neuro fuzzy interface system, and gene expression programming with K-fold cross validation. *Front. Mater.* **2021**, *8*, 621163. [\[CrossRef\]](#)
21. Zhang, J.; Wang, R.; Lu, Y.; Huang, J. Prediction of compressive strength of geopolymers concrete landscape design: Application of the novel hybrid RF-GWO-XGBoost algorithm. *Buildings* **2024**, *14*, 591. [\[CrossRef\]](#)
22. Getahun, M.A.; Shitote, S.M.; Abiero Gariy, Z.C. Artificial neural network based modelling approach for strength prediction of concrete incorporating agricultural and construction wastes. *Constr. Build. Mater.* **2018**, *190*, 517–525. [\[CrossRef\]](#)
23. Shahmansouri, A.A.; Yazdani, M.; Ghanbari, S.; Bengar, H.A.; Jafari, A.; Ghatte, H.F. Artificial neural network model to predict the compressive strength of eco-friendly geopolymers concrete incorporating silica fume and natural zeolite. *J. Clean. Prod.* **2021**, *279*, 123697. [\[CrossRef\]](#)
24. Peng, Y.; Unluer, C. Analyzing the mechanical performance of fly ash-based geopolymers concrete with different machine learning techniques. *Constr. Build. Mater.* **2022**, *316*, 125785. [\[CrossRef\]](#)
25. Ahmad, A.; Ahmad, W.; Chaiyasarn, K.; Ostrowski, K.A.; Aslam, F.; Zajdel, P.; Joyklad, P. Prediction of geopolymers concrete compressive strength using novel machine learning algorithms. *Polymers* **2021**, *13*, 3389. [\[CrossRef\]](#) [\[PubMed\]](#)
26. He, H.; Shuang, E.; Ai, L.; Wang, X.; Yao, J.; He, C.; Cheng, B. Exploiting machine learning for controlled synthesis of carbon dots-based corrosion inhibitors. *J. Clean. Prod.* **2023**, *419*, 138210. [\[CrossRef\]](#)
27. Koza, J.R. *Genetic Programming II: Automatic Discovery of Reusable Programs*; MIT Press: Cambridge, UK, 1994.
28. Gandomi, A.H.; Faramarzifar, A.; Rezaee, P.G.; Asghari, A.; Talatahari, S. New design equations for elastic modulus of concrete using multi expression programming. *J. Civ. Eng. Manag.* **2015**, *21*, 761–774. [\[CrossRef\]](#)
29. Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, USA, 1975.
30. Ferreira, C. *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006.
31. Azim, I.; Yang, J.; Iqbal, M.F.; Javed, M.F.; Nazar, S.; Wang, F.; Liu, Q.-F. Semi-analytical model for compressive arch action capacity of RC frame structures. *Structures* **2020**, *27*, 1231–1245. [\[CrossRef\]](#)
32. Azim, I.; Yang, J.; Iqbal, M.F.; Mahmood, Z.; Javed, M.F.; Wang, F.; Liu, Q.-F. Prediction of catenary action capacity of RC beam-column substructures under a missing column scenario using evolutionary algorithm. *KSCE J. Civ. Eng.* **2021**, *25*, 891–905. [\[CrossRef\]](#)
33. Oltean, M.; Dumitrescu, D. *Multi Expression Programming*; Technical Report; Babes-Bolyai University: Cluj-Napoca, Romania, 2002; pp. 1–28.
34. Pang, Y.; Azim, I.; Rauf, M.; Iqbal, M.F.; Ge, X.; Ashraf, M.; Tariq, M.A.U.R.; Ng, A.W.M. Prediction of bidirectional shear strength of rectangular RC columns subjected to multidirectional earthquake actions for collapse prevention. *Sustainability* **2022**, *14*, 6801. [\[CrossRef\]](#)
35. Nguyen, K.T.; Nguyen, Q.D.; Le, T.A.; Shin, J.; Lee, K. Analyzing the compressive strength of green fly ash based geopolymers concrete using experiment and machine learning approaches. *Constr. Build. Mater.* **2020**, *247*, 118581. [\[CrossRef\]](#)
36. Naghizadeh, A.; Ekelu, S.O. Behaviour of fly ash geopolymers binders under exposure to alkaline media. *Asian J. Civ. Eng.* **2019**, *20*, 785–798. [\[CrossRef\]](#)

37. Hardjito, D.; Rangan, B.V. *Development and Properties of Low-Calcium Fly Ash-Based Geopolymer Concrete*; Curtin University of Technology: Bentley, WA, Australia, 2005.
38. Rangan, B.V. *Fly Ash-Based Geopolymer Concrete*; Curtin University of Technology: Bentley, WA, Australia, 2008.
39. Gomaa, E.; Sargon, S.; Kashosi, C.; ElGawady, M. Fresh properties and compressive strength of high calcium alkali activated fly ash mortar. *J. King Saud Univ.-Eng. Sci.* **2017**, *29*, 356–364. [[CrossRef](#)]
40. Ashraf, M.; Iqbal, M.F.; Rauf, M.; Ulhaq, A.; Muhammad, H.; Liu, Q.-F. Developing a sustainable concrete incorporating bentonite clay and silica fume: Mechanical and durability performance. *J. Clean. Prod.* **2022**, *337*, 130315. [[CrossRef](#)]
41. Liu, Q.; Iqbal, M.; Yang, J.; Lu, X.-Y.; Zhang, P.; Rauf, M. Prediction of chloride diffusivity in concrete using artificial neural network: Modelling and performance evaluation. *Constr. Build. Mater.* **2021**, *268*, 121082. [[CrossRef](#)]
42. Wang, H.; Zhang, X.; Jiang, S. A laboratory and field universal estimation method for tire–pavement interaction noise (TPIN) based on 3D image technology. *Sustainability* **2022**, *14*, 12066. [[CrossRef](#)]
43. Frank, I.E.; Todeschini, R. *The Data Analysis Handbook*; Elsevier: Amsterdam, The Netherlands, 1994; Volume 14.
44. Kabir, H.; Garg, N. Machine learning enabled orthogonal camera goniometry for accurate and robust contact angle measurements. *Sci. Rep.* **2023**, *13*, 1497. [[CrossRef](#)]
45. Rauf, M.; Khaliq, W.; Khushnood, R.A.; Ahmed, I. Comparative performance of different bacteria immobilized in natural fibers for self-healing in concrete. *Constr. Build. Mater.* **2020**, *258*, 119578. [[CrossRef](#)]
46. Meng, Z.; Liu, Q.; She, W.; Cai, Y.; Yang, J.; Iqbal, M.F. Electrochemical deposition method for load-induced crack repair of reinforced concrete structures: A numerical study. *Eng. Struct.* **2021**, *246*, 112903. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.