



OPEN Innovative approaches for predicting the elastic modulus and tensile strength of green concrete using soft computing methods

Elaheh Moshaver¹, Ali Seyedkazemi¹✉ & Saba Jahangir²

In this study, interpretable and semi-interpretable soft computing techniques, including Group Method of Data Handling (GMDH), Gene Expression Programming (GEP), and Response Surface Methodology (RSM), were employed to develop predictive relationships for estimating the elastic modulus (E) and splitting tensile strength (STS) of concrete containing waste foundry sand (WFS). A sensitivity analysis was subsequently conducted to evaluate the influence of various parameters on these mechanical properties. The input variables considered were the ratio of waste foundry sand to cement (WFS/C), the ratio of waste foundry sand to fine aggregate (WFS/FA), the ratio of fine aggregate to total aggregate (FA/TA), the ratio of water to cement (W/C), the ratio of coarse aggregate to cement (CA/C), the ratio of superplasticizer to cement (100SP/C) and the age of the sample. The results revealed that the GMDH model achieved the highest correlation coefficient (R) among all methods in predicting STS, exhibiting the lowest root mean square error (RMSE = 0.533) and mean absolute error (MAE = 0.434). Thus, the GMDH model demonstrated superior performance in predicting STS based on all statistical indicators. For predicting E, the RSM method provided the most accurate results, with the highest $R = 0.978$ and the lowest errors, RMSE = 1.372 and MAE = 1.088. Sensitivity analysis indicated that CA/C had the most significant effect on STS, while W/C had the greatest influence on E. Other parameters, WFS/C, CA/C, and FA/TA, showed relatively minor impacts on E.

Keywords Green concrete, Waste foundry sand, Mechanical parameters, Soft computing methods, Sensitivity analysis

The widespread use of concrete in civil engineering has brought growing attention to the challenges of sourcing its raw materials. The use of waste materials and industrial by-products plays an important role in producing green and sustainable concrete^{1–8}. Therefore, civil engineers aim to partially replace conventional concrete components with waste materials and industrial by-products. Many researchers have studied the use of fly ash, rice husk ash, tire rubber, recycled aggregates, and waste foundry sand for producing green concrete. Among these materials, the use of WFS in producing green concrete appears to be a promising option because of the continuous growth of the foundry industry. The cost of landfill disposal for WFS is relatively high, ranging from about USD 135 to 675 per ton. Moreover, the presence of heavy metals such as cadmium, zinc, and lead in, which can raise environmental concerns when released into the environment. Using WFS as a partial replacement for fine aggregates in concrete production can help reduce environmental issues and economic costs. Laboratory studies have reported that the optimal replacement level of WFS as fine aggregate is around 10–20%^{9–12}.

Determining the mechanical properties of concrete through laboratory testing is expensive, time-consuming, and requires extensive experimental procedures for each mix design. Therefore, civil engineers are looking for methods to estimate the mechanical properties of different types of material without performing physical tests^{13–20}. Soft computing techniques can predict these properties by learning the hidden relationships within datasets^{21–26}. Several studies have applied different soft computing approaches to predict the mechanical parameters of green concrete^{27–35}. Javed et al.³⁶ used machine learning methods, including support vector regression (SVR), decision tree (DT), and AdaBoost regressor (AR), to predict the mechanical properties of green concrete. The results of this study showed that the SVR method was the most suitable method for predicting the mechanical properties of green concrete. In addition to the appropriate accuracy, these methods were black box methods, meaning they did not provide explicit mathematical relationships for predicting mechanical parameters. Instead, their

¹Department of Civil Engineering, Am.C., Islamic Azad University, Amol, Iran. ²Department of Architecture, Am.C., Islamic Azad University, Amol, Iran. ✉email: ali.seyedkazemi@iau.ac.ir

application requires access to the implemented software code and the corresponding dataset. Iqbal et al.³⁷ used the GEP method to predict the mechanical parameters of concrete containing WFS. The results showed that the R-value obtained from the GEP model for the elastic modulus and splitting tensile strength were 0.812 and 0.818, respectively. Jakubowski and Tomczak³⁸ predicted the self-healing process of concrete using a convolutional neural network (CNN). Their model could reasonably accurately predict the crack width at different stages of self-healing. In evaluating the model, the performance criteria included an MAE of 14.1 μm and an RMSE of 26.7 μm . Kumar et al.³⁹ predicted the compressive strength of lightweight concrete using soft computing methods. All the models used in this study were black-box methods. The results showed that the Gaussian Process Regression (GPR) and Support Vector Machine Regression (SVMR) methods predicted the compressive strength of lightweight concrete with an R-value of 0.9740 and 0.9777, respectively. Shahrokhishahraki et al.⁴⁰ used various machine learning methods to predict the compressive strength of structural concrete. The results showed that the Elastic Net algorithm performed best in determining the optimal mixing design to achieve maximum compressive strength. This mixing design resulted in a 10% reduction in cement consumption while maintaining the compressive strength of the concrete.

Research significant

The utilization of recycled materials and industrial by-products, such as waste foundry sand, as a partial replacement for natural aggregates represents an effective step toward the development of sustainable and eco-friendly concretes. Although numerous studies have been conducted, accurately predicting WFS-containing concrete mechanical properties remains a key challenge in civil engineering. Most previous studies have employed soft computing approaches with a black-box nature. Although these models often exhibit satisfactory predictive accuracy, their complex and non-transparent internal structures make it difficult to interpret the relationships among variables and to apply the results in engineering design. Nevertheless, an approach that provides both high prediction accuracy and interpretability of relationships has not yet been fully achieved.

In the present study, three modeling techniques with varying levels of interpretability, including RSM, GMDH, and GEP methods, have been employed to develop accurate and reliable relationships for predicting E and STS of WFS-based concrete. The selection of these three methods was based on their ability to balance high predictive accuracy with the ability to interpret mathematical relationships. This ensures that while making accurate predictions, it is also possible to analyze the influence of input variables.

The main innovation of this research lies in a comprehensive analytical framework that combines interpretable and semi-transparent methods. This framework enables both a comparative evaluation of model performance in predicting the mechanical behavior of WFS-based concrete and the development of robust empirical relationships applicable to engineering design. Furthermore, the sensitivity analysis based on the developed models enables the identification of the most influential variables affecting E and STS. It leads to a deeper understanding of the interaction between the mixture composition and the mechanical behavior of WFS-based concrete. This approach can serve as a practical step toward optimizing the mix design and advancing the development of sustainable concrete in civil engineering.

Data collection

This study considers the splitting tensile strength (STS) and elastic modulus (E) of concrete as output variables. To model these properties, seven input variables have been selected, including the ratio of waste foundry sand to cement (WFS/C), the ratio of waste foundry sand to fine aggregate (WFS/FA), the ratio of fine aggregate to total aggregate (FA/TA), the ratio of water to cement (W/C), the ratio of coarse aggregate to cement (CA/C), the ratio of superplasticizer to cement (1000SP/C) and the age of the sample. These parameters were chosen because of their significant influence on the mechanical properties of concrete and their critical role in designing the optimal mix. A brief description of the rationale behind selecting each input parameter is provided in the following section.

Waste foundry sand to cement ratio (WFS/C)

This ratio represents the amount of waste foundry sand used as a cement substitute. An increase in this ratio may reduce the elastic modulus and splitting tensile strength, because WFS generally lacks the mechanical properties of cement. This input variable was included due to the importance of waste management and in reducing cement consumption.

Waste foundry sand to fine aggregate ratio (WFS/FA)

This ratio indicates the proportion of waste sand used to replace part of the fine aggregate. An increasing ratio may influence the mechanical properties of concrete by affecting the bond quality between the cement paste and the aggregates. This input parameter was selected to examine the effect of using alternative materials as fine aggregates.

Fine aggregate to total aggregate ratio (FA/TA)

This ratio represents the distribution of fine and coarse aggregate within the concrete mix. Variations in this ratio can influence the compressibility and density of the mix and, consequently, affect mechanical properties such as splitting tensile strength. Moreover, a higher proportion of fine aggregates reduces porosity and increases the elastic modulus by making the mix more compact. This input parameter was included to optimize the concrete mix design.

Variable	Mean	StD	Variance	CoV	Min	Max
WFS/C	0.3657	0.3173	0.1007	86.76	0	1.28
W/C	0.46826	0.04765	0.00227	10.18	0.4	0.56
CA/C	2.8401	0.5680	0.3227	20.00	0.93	4.2
FA/TA	0.31442	0.09573	0.00916	30.44	0.1	0.69
WFS/FA	0.3178	0.3804	0.1447	119.71	0	2.33
1000SP/C	2.616	5.057	25.576	193.30	0	17.8
Age	53.51	86.98	7565.42	162.54	3.00	365
STS (MPa)	3.2050	0.7835	0.6138	24.45	1.69	4.9

Table 1. Dataset used in the current study for STS.

Variable	Mean	StD	Variance	CoV	Min	Max
WFS/C	0.2916	0.2962	0.0877	101.59	0.0000	1.2800
W/C	0.45548	0.04768	0.00227	10.47	0.40000	0.55000
CA/C	2.6896	0.5184	0.2687	19.27	1.4900	3.3000
FA/TA	0.32959	0.08843	0.00782	26.83	0.10000	0.50000
WFS/FA	0.2381	0.3214	0.1033	134.98	0.0000	2.3300
1000SP/C	2.974	5.022	25.220	168.87	0.000	15.860
Age	64.37	102.31	10468.08	158.95	3.00	365.00
E (GPa)	30.455	6.736	45.372	22.12	18.400	46.650

Table 2. Dataset used in the current study for E.

Water to cement ratio (W/C)

A higher W/C ratio generally reduces the elastic modulus as excess water increases the porosity of the concrete matrix. Conversely, a lower W/C ratio tends to enhance the splitting tensile strength due to improved bonding between the cement paste and aggregates. However, with excessive W/C, the splitting tensile strength decreases. This input is selected based on concrete mix design standards.

Coarse aggregate to cement ratio (CA/C)

This ratio reflects the influence of coarse aggregates on the strength and elastic modulus of concrete. Increasing this ratio enhances the elastic modulus but may adversely affect the workability of the mixture. A higher CA/C ratio can also reduce the splitting tensile strength because coarse aggregates tend to create discontinuities within the cementitious matrix. This input was selected due to the role of coarse aggregates in stress distribution and cracking reduction.

Superplasticizer to cement ratio (1000SP/C)

The use of superplasticizers improves the workability of concrete while reducing the water-to-cement ratio. An increase in this ratio generally enhances the splitting tensile strength and elastic modulus by reducing porosity and strengthening internal bonds. This input parameter was selected because of the important role of superplasticizers in concrete quality.

Age of the sample (age)

The age of concrete directly affects its mechanical properties. As the age increases, the splitting tensile strength and elastic modulus of concrete typically improve due to continued cement hydration. This input was selected because of the natural effect of time on the properties of concrete.

In this study, two separate laboratory data sets were used to analyze the mechanical properties of concrete: 146 data sets related to elastic modulus and 242 data sets related to splitting tensile strength^{41–44}. Prior to modeling, data preprocessing was carried out to ensure consistency and reliability. Outliers were identified and removed using the Modified Z-score method with a threshold value of 3.5, applied to all input variables. After data cleaning, 242 valid data points remained for STS and 146 for E. In the following sections, the effect of input variables on each of these outputs is examined separately, and the modeling results are presented. Tables 1 and 2 provide statistical information related to STS and E input and output variables, respectively. Additionally, Figs. 1 and 2 illustrate the histograms of the input and output data for STS and E and their distribution curves.

Modeling

GMDH

A three-layer hidden architecture with up to 20 neurons in each layer was used to predict the E and STS using the GMDH method. The network structure was automatically optimized to achieve the best model performance. In the model training process, 80% of the data was used for model development, and the remaining 20% was

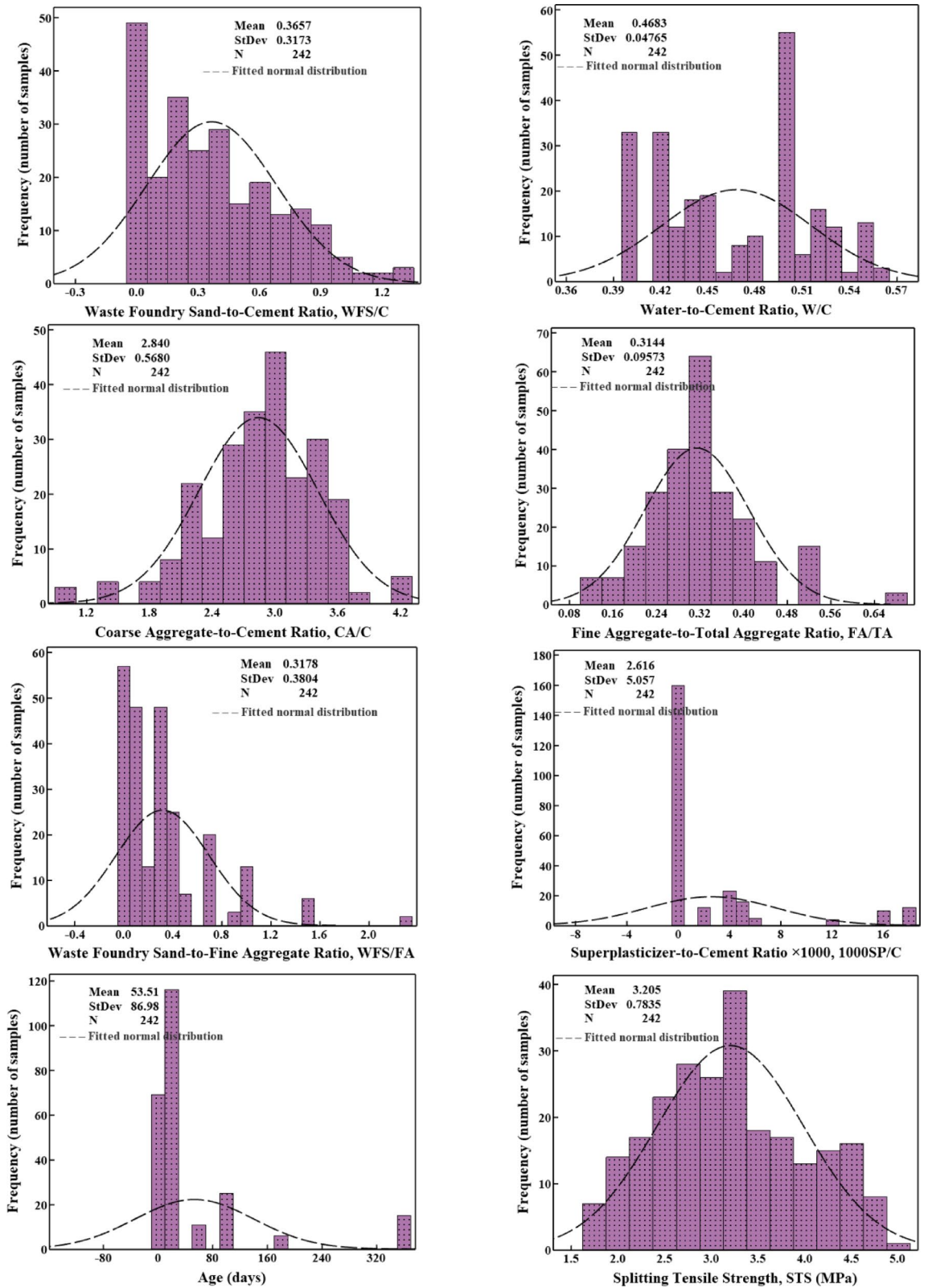


Fig. 1. Histogram and fitted normal distribution of the input and output data related to splitting tensile strength (STS).

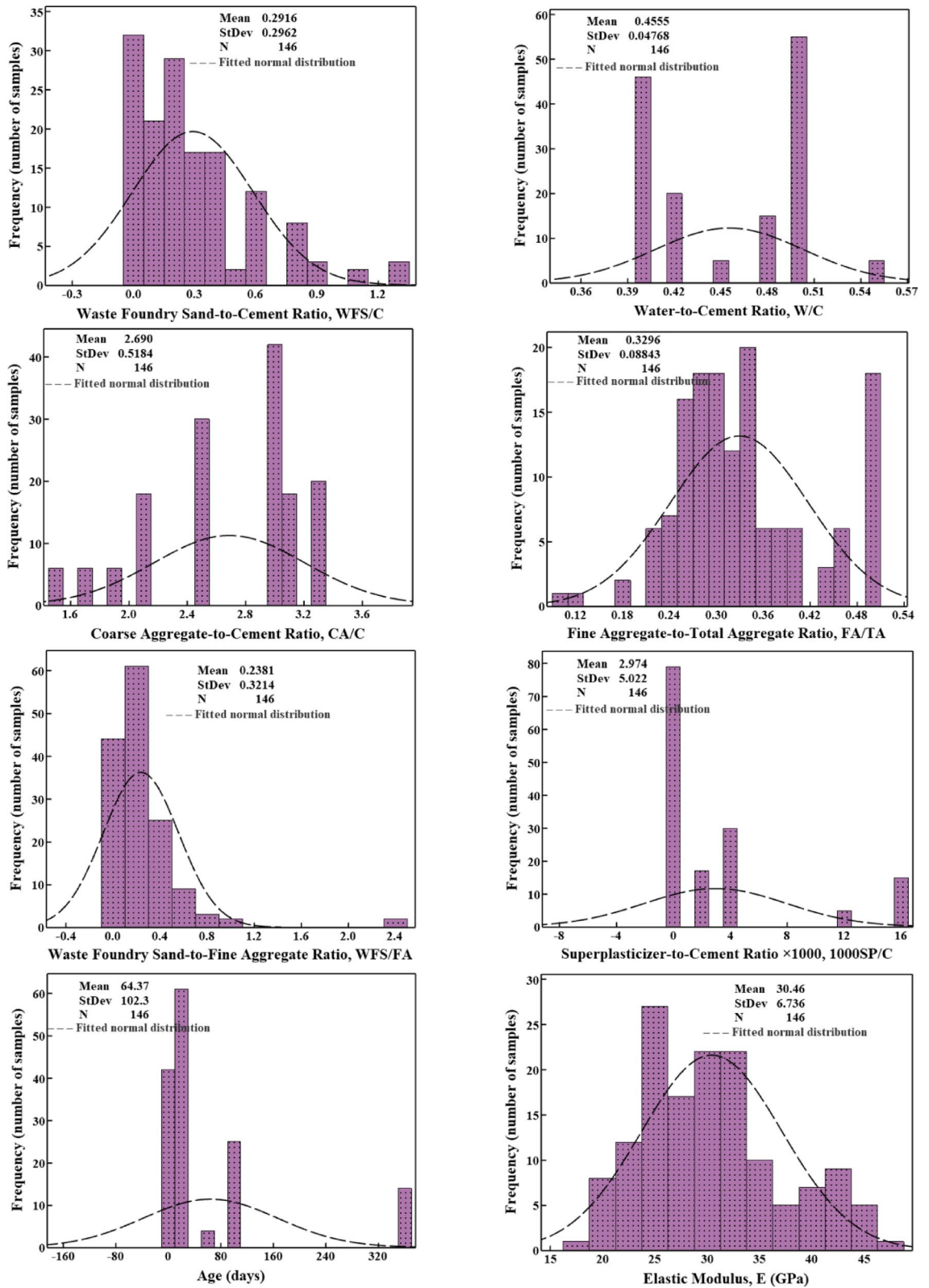
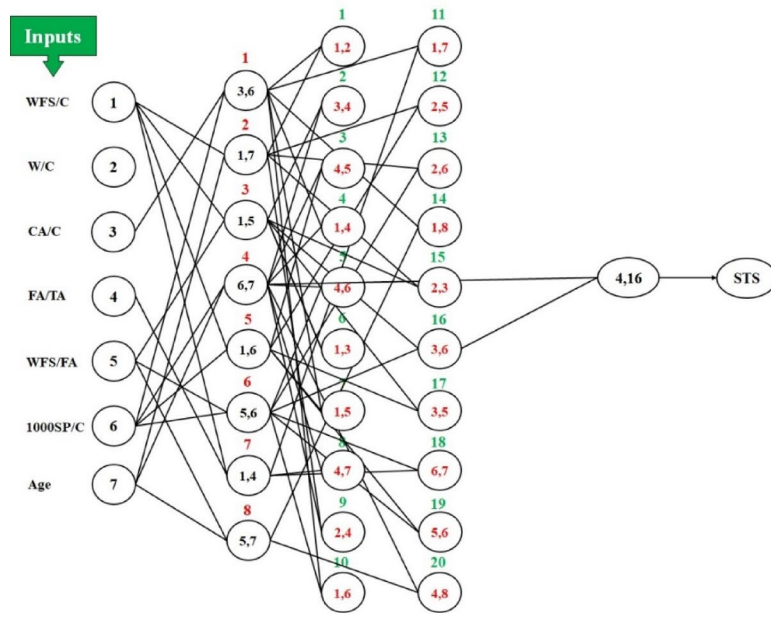


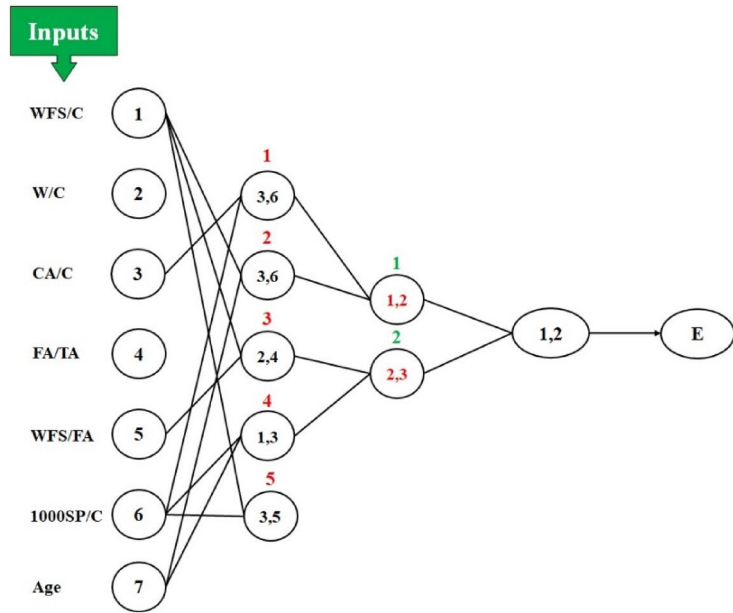
Fig. 2. Histogram and fitted normal distribution of the input and output data related to elastic modulus (E).

reserved for validation. This approach allowed for the assessment of the predictive accuracy of the model under various conditions. Figure 3 illustrates the structure of GMDH networks used for STS and E prediction.

Equations 1 and 2 present the polynomial expressions derived from the GMDH method for predicting STS and E, respectively.



(a)



(b)

Fig. 3. Structure of GMDH neural network for predicting (a) STS and (b) E.

$$\left\{ \begin{array}{l} y_1 = 3.473 + 0.121 \times CA/C + 0.66 \times 1000SP/C - 0.091 \times CA/C^2 \\ \quad - 6.63 \times 10^{-3} \times 1000SP/C^2 - 0.17 \times CA/C \times 1000SP/C \\ y_2 = 3.427 - 0.889 \times WFS/C + 0.165 \times WFS/FA + 0.215 \times WFS/C^2 \\ \quad - 5.51 \times 10^{-3} \times WFS/FA^2 - 0.192 \times WFS/C \times WFS/FA \\ y_3 = 2.928 - 0.053 \times 1000SP/C + 0.008 \times Age + 0.004 \times 1000SP/C^2 \\ \quad - 1.52 \times 10^{-5} \times Age^2 - 8.92 \times 10^{-5} \times 1000SP/C \times Age \\ y_4 = 3.333 - 0.712 \times WFS/FA - 0.012 \times 1000SP/C + 0.182 \\ \quad \times WFS/FA^2 + 1.17 \times 10^{-3} \times 1000SP/C^2 - 0.016 \times WFS/FA \times 1000SP/C \end{array} \right. \quad (1)$$

$$\left\{ \begin{array}{l} Y_1 = -33.908 - 2.212 \times y_1 + 22.624 \times y_3 + 0.434 \\ \quad \times y_1^2 - 3.317 \times y_3^2 + 0.147 \times y_1 \times y_3 \\ Y_2 = 1.903 + 0.690 \times y_2 - 0.991 \times y_4 - 0.28 \\ \quad \times y_2^2 - 0.095 \times y_4^2 + 0.595 \times y_2 \times y_4 \end{array} \right.$$

$$STS = -13.191 + 0.507 \times Y_1 + 8.675 \times Y_2 \\ - 0.352 \times Y_1^2 - 1.803 \times Y_2^2 + 0.903 \times Y_1 \times Y_2$$

$$\left\{ \begin{array}{l} y_1 = 353.89 - 1311.69 \times W/C - 5.076 \times 1000SP/C + 1309.14 \\ \quad \times W/C^2 + 0.052 \times 1000SP/C^2 + 8.859 \times W/C \times 1000SP/C \\ y_2 = -27.11 + 64.051 \times CA/C - 5.818 \times 1000SP/C \\ \quad - 15.254 \times CA/C^2 + 0.095 \times 1000SP/C^2 + 1.506 \times CA/C \times 1000SP/C \\ y_3 = 236.64 - 725.085 \times W/C - 171.24 \times FA/TA + 697.37 \\ \quad \times W/C^2 + 234.43 \times FA/TA^2 + 45.94 \times W/C \times FA/TA \end{array} \right. \quad (2)$$

$$\left\{ \begin{array}{l} Y_1 = -35.558 + 5.353 \times y_1 - 2.105 \times y_2 \\ \quad - 0.038 \times y_1^2 + 0.084 \times y_2^2 - 0.08 \times y_1 \times y_2 \\ Y_2 = 12.313 + 0.846 \times y_2 - 0.638 \times y_3 \\ \quad + 0.081 \times y_2^2 + 0.104 \times y_3^2 - 0.173 \times y_2 \times y_3 \end{array} \right.$$

$$E = 1.391 + 4.86 \times Y_1 - 3.653 \times Y_2 - 0.174 \\ \times Y_1^2 - 0.029 \times Y_2^2 + 0.204 \times Y_1 \times Y_2$$

RSM

In this section, the RSM method was employed to model and predict STS and E. As a practical statistical approach, RSM enables the analysis of variable interactions and the development of mathematical relationships among them. The final regression equations were derived to describe the relationships between input and output variables and to assess the significance of the RSM model parameters. Equations 3 and 4 present the mathematical relationship between the input variables with STS and E, respectively.

$$STS = 4.0817 - 0.642 \times WFS/C - 1.526 \times W/C \\ - 0.120 \times CA/C + 0.1844 \times WFS/FA \\ + 0.7824 \times 1000SP/C - 4.334 \times 10^{-3} \times Age \\ + 0.0205 \times WFS/C \times Age - 0.6672 \times W/C \times 1000SP/C \\ + 0.0775 \times W/C \times Age - 0.1392 \times CA/C \\ \times 1000SP/C - 8.2998 \times 10^{-3} \times CA/C \times Age \\ - 0.0158 \times WFS/FA \times Age \\ - 1.73690507445996 \times 10^{-5} \times Age^2 \quad (3)$$

$$\begin{aligned}
E = & 2299.95 - 45.9427 \times WFS/C4321.21 \times W/C - 604.924 \times CA/C \\
& - 2041.52 \times FA/TA - 435.29 \times WFS/FA - 59.6731 \times 1000SP/C \\
& - 0.388477 \times Age - 287.085 \times WFS/C \times W/C + 96.8323 \times WFS/C \times CA/C \\
& - 167.266 \times WFS/C \times FA/TA + 197.768 \times WFS/C \times WFS/FA \\
& + 0.340552 \times WFS/C \times Age - 959.453 \times W/C \times CA/C + 207.258 \times W/C \times WFS/FA \\
& + 232.165 \times W/C \times 1000SP/C - 0.762129 \times W/C \times Age + 562.072 \times CA/C \times FA/TA \\
& + 41.872 \times CA/C \times WFS/FA - 17.2479 \times CA/C \times 1000SP/C + 0.112909 \times CA/C \times Age \\
& + 516.433 \times FA/TA \times WFS/FA + 1.47503 \times FA/TA \\
& \times Age - 141.834 \times WFS/C^2 + 7168.61 \times W/C^2 \\
& + 154.541 \times CA/C^2 + 572.575 \times FA/TA^2 + -16.6941 \\
& \times WFS/FA^2 - 1.1044 \times 10^{-4} \times Age^2
\end{aligned} \quad (4)$$

Table 3 presents the analysis of variance (ANOVA) results for Eq. 3 in predicting STS. This table provides detailed statistical information, including the F-value, P-value, and the contribution of each factor to the total variance. Analysis of these data is essential for evaluating the validity of the model and identifying influential variables.

The results in Table 3 indicate that the quadratic model used to predict STS of concrete is statistically significant, identifying the general trend and the relationship between the input and response variables. The model F value of 13.21 and a p-value of less than 0.0001 confirm significant relationships between the main parameters and the response with high confidence. However, the Lack of Fit test ($p=0.0014$) reveals that the model does not perfectly reproduce the experimental data. This means that although the model is statistically valid and can describe the general patterns between the variables, it cannot fully explain all the variability observed in the experimental data. Such discrepancies may be attributed to the complex behavior of concrete, insufficient data at certain critical points, or the omission of influential variables.

Next, the quadratic Equation predicting the elastic modulus and its ANOVA table are examined using the RSM method. Table 4 shows the ANOVA analysis of Eq. 4 for predicting E.

According to Table 4, the quadratic model developed for predicting E is statistically significant, with an F-value of 95.74 and a p-value < 0.0001. This demonstrates that the model effectively explains the variance in the experimental data. The primary factors, such as W/C, CA/C, FA/TA, and Age, are statistically significant, along with their interactions and second-order terms, indicating a substantial influence on the response. In contrast, the variable WFS/FA with p-value = 0.755 is insignificant and has no notable impact on the model. The residual value and the mean square error (2.35) indicate that the unexplained variation by the model is relatively low and that the model has acceptable accuracy. Moreover, Lack of Fit had an F-value of 1.41 and a p-value of 0.26 with pure error of 1.72, indicating that Lack of Fit is insignificant. The insignificance of Lack of Fit further confirms that the provides an acceptable level of accuracy in predicting E.

Source	Sum of squares	df	Mean square	F value	p-value Prob > F	
Model	63.55	13	4.89	13.21	<0.0001	significant
A-WFS/C	2.86	1	2.86	7.74	0.0059	
B-W/C	1.07	1	1.07	2.88	0.0911	
C-CA/C	19.32	1	19.32	52.20	<0.0001	
E-WFS/FA	4.08	1	4.08	11.02	0.0010	
F-1000SP/C	14.32	1	14.32	38.69	<0.0001	
G-Age	4.25	1	4.25	11.48	0.0008	
AG	3.35	1	3.35	9.06	0.0029	
BF	1.68	1	1.68	4.54	0.0342	
BG	4.38	1	4.38	11.83	0.0007	
CF	9.59	1	9.59	25.92	<0.0001	
CG	4.38	1	4.38	11.84	0.0007	
EG	3.67	1	3.67	9.92	0.0019	
G ²	3.87	1	3.87	10.47	0.0014	
Residual	84.39	228	0.37			
Lack of fit	84.14	221	0.38	10.66	0.0014	Significant
Pure error	0.25	7	0.036			
Cor total	147.93	241				

Table 3. ANOVA evaluation of the quadratic model for STS prediction.

Source	Sum of squares	df	Mean square	F value	p-value Prob > F	
Model	6303.86	28	225.14	95.74	<0.0001	Significant
A-WFS/C	28.15	1	28.15	11.97	0.0008	
B-W/C	89.26	1	89.26	37.96	<0.0001	
C-CA/C	54.23	1	54.23	23.06	<0.0001	
D-FA/TA	38.64	1	38.64	16.43	<0.0001	
E-WFS/FA	0.23	1	0.23	0.097	0.7555	
F-1000SP/C	12.74	1	12.74	5.42	0.0216	
G-Age	319.52	1	319.52	135.87	<0.0001	
AB	37.33	1	37.33	15.88	0.0001	
AC	63.63	1	63.63	27.06	<0.0001	
AD	10.51	1	10.51	4.47	0.0366	
AE	67.93	1	67.93	28.89	<0.0001	
AG	344.12	1	344.12	146.34	<0.0001	
BC	76.48	1	76.48	32.52	<0.0001	
BE	49.19	1	49.19	20.92	<0.0001	
BF	60.25	1	60.25	25.62	<0.0001	
BG	95.18	1	95.18	40.48	<0.0001	
CD	155.23	1	155.23	66.01	<0.0001	
CE	10.14	1	10.14	4.31	0.0400	
CF	20.24	1	20.24	8.61	0.0040	
CG	111.63	1	111.63	47.47	<0.0001	
DE	46.05	1	46.05	19.58	<0.0001	
DG	327.49	1	327.49	139.27	<0.0001	
A ²	73.85	1	73.85	31.40	<0.0001	
B ²	147.49	1	147.49	62.72	<0.0001	
C ²	181.26	1	181.26	77.08	<0.0001	
D ²	20.05	1	20.05	8.53	0.0042	
E ²	18.60	1	18.60	7.91	0.0058	
G ²	83.32	1	83.32	35.43	<0.0001	
Residual	275.13	117	2.35			
Lack of fit	254.52	105	2.42	1.41	0.2599	Not significant
Pure error	20.61	12	1.72			
Cor total	6578.99	145				

Table 4. ANOVA evaluation of the quadratic model for E prediction.

GEP

Next, the GEP method was employed to predict STS and E. This approach was selected because of its high ability to discover complex nonlinear relationships among variables. It is also capable of producing accurate and interpretable models. For model development, the available dataset was divided into two categories:

- 80% of the data for the model training phase to learn the relationships between inputs and outputs.
- 20% of the data for testing the model and evaluating its prediction accuracy.

Table 5 summarizes the parameters used in the GEP models. These parameters include model inputs and related settings used to optimize and increase the accuracy of STS and E predictions.

The equations obtained for predicting STS and E by the GEP method are as follows:

$$\begin{aligned}
 STS = & [\cos((CA/C) + ((0.58 - FA/TA) \times W/C) + \cos(9.11))] \\
 & + [\cos(\cos(W/C - Age))] \\
 & + [\cos(\cos(CA/C + ((0.58 - FA/TA) \times W/C) + \cos(9.33))] \\
 & + [\cos(\cos(((FA/TA - 1000SP/C) \times \cos(CA/C))) + CA/C)]
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 E = & [CA/C / (CA/C - WFS/FA)] \\
 & + \dots [WFS/FA^3] + [WFS/FA] \\
 & + \dots [(WC - FA/TA) - WFS/C^3] + ((WFS/C \times FA/TA) + FA/TA) \\
 & + \dots [((-3.912 - 3.64) - (FA/TA - CA/C) \times ((-1.93/W/C) - 2.201))]
 \end{aligned} \tag{6}$$

Parameter	Definition	Model-STS	Model-E
P1	Fitness Function	RMSE	RSE
P2	Function set	+, -, ×, /, Exp, Sin, Cos, Asin	+, -, ×, /, Inv, X ² , X ³
P3	Chromosomes	30	30
P4	Head size	8	6
P5	Number of genes	4	5
P6	Linking function	Addition	Addition
P7	Mutation rate	0.0667	0.0769
P8	Inversion rate	0.1	0.1
P9	One-point recombination rate	0.2	0.1
P10	Two-point recombination rate	0.2	0.2
P11	Gene recombination rate	0.3	0.4
P12	Gene transposition rate	0.1	0.1
P13	Constant per gene	10	10
P14	Training-R ²	0.331	0.572
P15	Testing-R ²	0.393	0.522

Table 5. Summary of parameters used in GEP models.

Figures 4 and 5 illustrate the expression trees corresponding to Eqs. (5) and (6), which describe the relationships between the input variables and the output parameters, STS and E. In these models, variables d_0 to d_6 represent WFS/C, WFS/FA, FA/TA, W/C, CA/C, 1000SP/C, and Age, respectively, while c_0 to c_6 denote the constants automatically generated during the model development process.

Performance of methods in predicting STS and E and sensitivity analysis of input parameters

Several statistical indices, R, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Standard Deviation (StD), and Scatter Index (SI), were used to evaluate the performance of different methods in predicting STS and E. The equations for calculating these statistical parameters are presented below.

$$R = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2} \sqrt{\sum (\hat{y}_i - \bar{\hat{y}})^2}} \quad (7)$$

$$MAE = \frac{\sum_{i=1}^n |Y_{i(Act)} - Y_{i(Pre)}|}{N} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{i(Act)} - Y_{i(Pre)})^2}{n}} \quad (9)$$

$$StD = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad (10)$$

$$SI = \frac{RMSE}{(1/n) \sum_{i=1}^n y_{i(Actual)}} \quad (11)$$

In Eqs. (7 to 11), \bar{y} is the mean of the data, n is the total number of data, $Y_{i(Act)}$ is the actual value, and $Y_{i(Pre)}$ is the predicted value of the i th sample of the data set.

Figures 6 and 7 illustrate the relationship between the predicted and actual values of STS and E for the three GMDH, RSM, and GEP models. In both cases, the closer the data points are to the 45-degree line, the higher the model accuracy. In the STS prediction, the GMDH model, with a correlation coefficient of $R=0.731$, provides the best fit to the actual data. The data are mainly within the $\pm 20\%$ error range, indicating satisfactory accuracy. The RSM model has an acceptable but slightly weaker performance, and the GEP model exhibited the widest error dispersion.

In the prediction of E, the RSM model with $R=0.978$ provided the highest correlation between the actual and predicted values and the closest fit to the ideal line. The GMDH and GEP models, with $R=0.855$ and $R=0.744$, demonstrated weaker predictive capability and greater deviation from the measured values. Therefore, the RSM model in estimating E and the GMDH model in predicting STS were identified as the most accurate models.

In selecting the most appropriate model, in addition to the correlation coefficient, other statistical indices such as MSE, MAE, and RMSE should also be considered to achieve a more accurate evaluation of the performance of

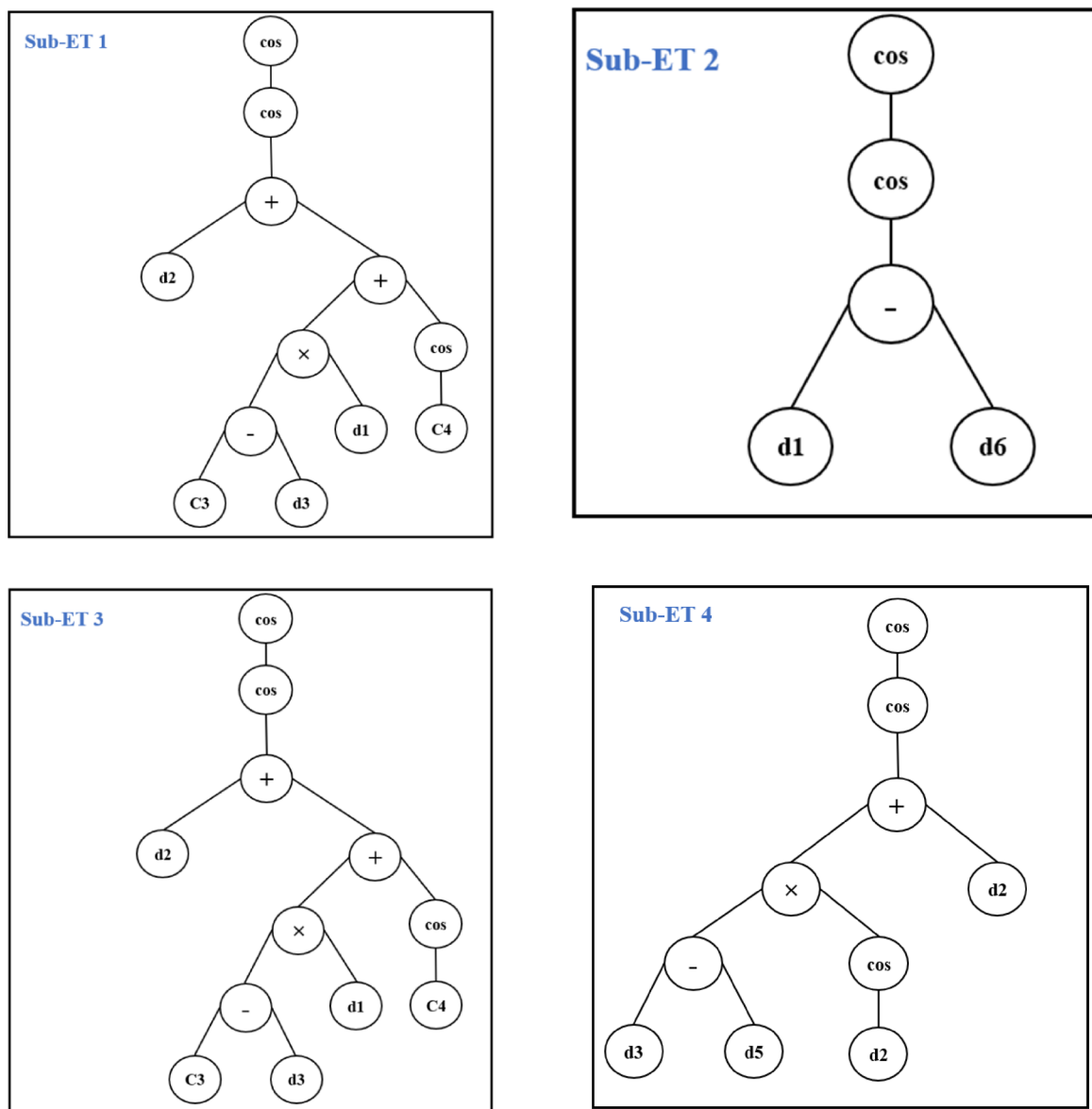


Fig. 4. Expression trees (Sub-ET1 to Sub-ET4) of the GEP model for predicting STS.

each model. Figure 8 compares different statistical indices to assess the performance of GMDH, GEP, and RSM models in predicting STS. These indices provide a basis for evaluating the accuracy and predictive capability of each model.

Figure 8 shows that the GMDH model achieved the highest R-value, indicating a strong correlation between the actual and predicted values. It also has the lowest RMSE=0.533 and MAE=0.434, indicating its superior predictive accuracy compared with the other methods. In contrast, the GEP model has the lowest R-value and the highest MSE (0.648) and MAE (0.5406), reflecting the weakest predictive performance and the largest error among the models. The RSM model performed between the two models, with an R-value of 0.655, demonstrating a moderate correlation between the actual and predicted data.

To assess the statistical significance of the difference in model performance, the Wilcoxon signed-rank test was performed on the absolute error values between the actual and predicted data. The results revealed that the difference in performance between the RSM and GMDH models ($p=0.0066$) and between RSM and GEP ($p=0.0125$) was statistically significant. In addition, the difference between the GMDH and GEP models was also significant ($p=4.74 \times 10^{-5}$), indicating a significant difference in the prediction accuracy of the two models. Figure 9 also presents the residual plots for the compared models. As shown, the GMDH model has a relatively symmetrical distribution of residuals without systematic patterns and fewer outliers than the other models. In contrast, the GEP model has a wider dispersion and several significant errors in areas with high STS values. The RSM model, although generally consistent, displays more fluctuations in residuals for boundary data. Accordingly, the GMDH model provides the most accurate and stable performance in predicting the STS of concrete and was selected for further analyses, including sensitivity analysis.

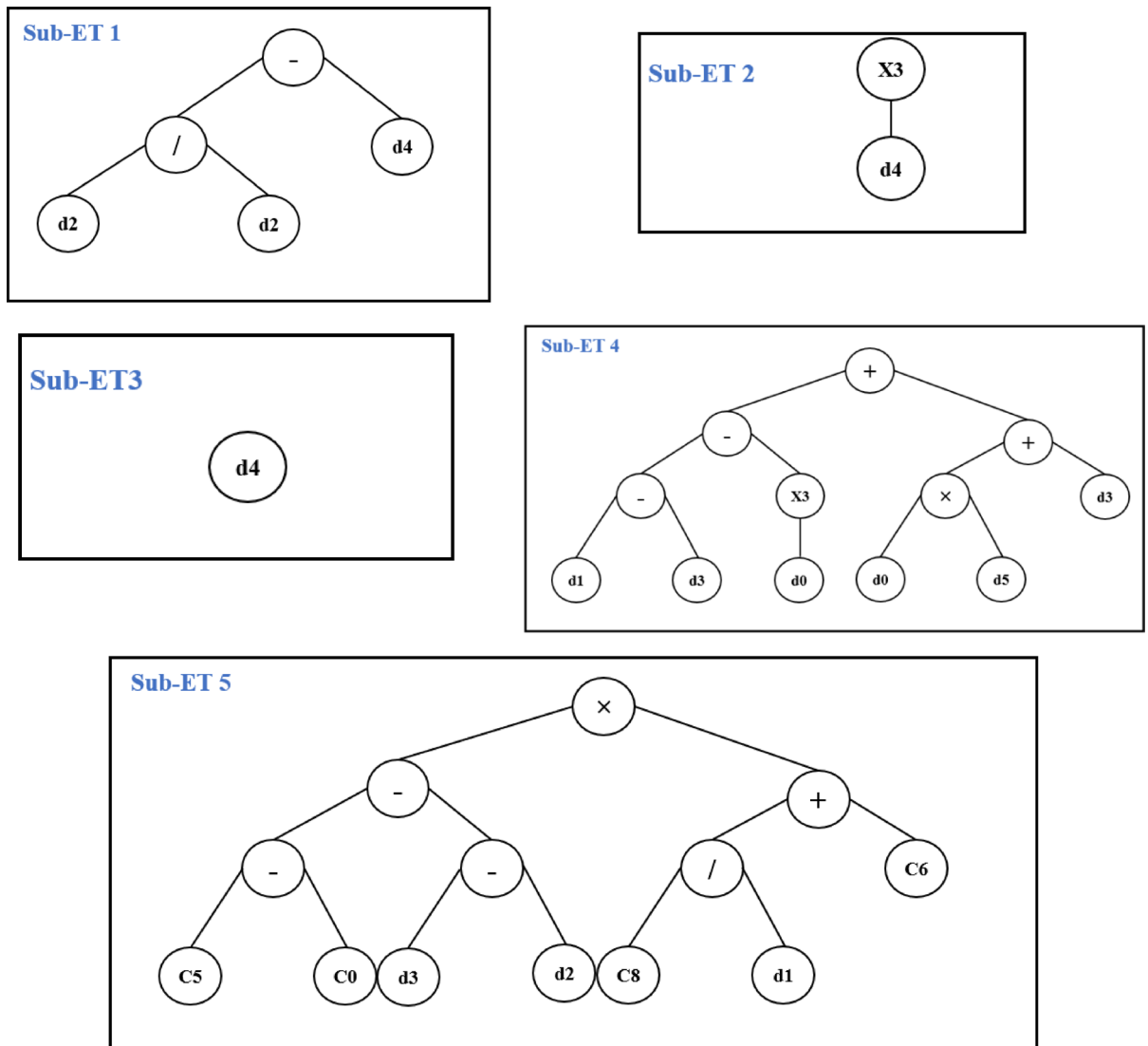


Fig. 5. Expression trees (Sub-ET1 to Sub-ET5) of the GEP model for predicting E.

Various statistical indices are shown in Fig. 10 to evaluate the performance of different models in predicting E.

Figure 10 demonstrates that, for predicting E, the RSM method had the best performance, with the highest correlation coefficient ($R=0.978$) and the lowest errors (RMSE = 1.372 and MAE = 1.088). The GMDH method also performed well but with lower performance than RSM, with $R=0.855$ and moderate errors (RMSE = 3.478 and MAE = 2.702). In contrast, the GEP method showed the weakest predictive capability, with the lowest R-value ($R=0.744$) and the highest errors (RMSE = 4.48 and MAE = 3.346).

Furthermore, the results of the Wilcoxon signed-rank test based on the absolute error values revealed that the difference between the RSM model and the other models is statistically significant (RSM vs. GMDH: $p=2.62 \times 10^{-16}$, RSM vs. GEP: $p=9.05 \times 10^{-15}$). In contrast, the difference between GMDH and GEP is not statistically significant ($p=0.8276$). These findings indicate that the RSM model is the most statistically accurate method for predicting the E, whereas the GMDH and GEP models have comparable performance. Additionally, the residual plots shown in Fig. 11 confirm this conclusion. The RSM model demonstrates that the distribution of errors is more symmetrical around the zero axis without any systematic pattern, while the GEP model shows wider dispersion and a greater number of outliers.

The selected GMDH and RSM models were used to perform sensitivity analysis for STS and E, respectively. Sensitivity analysis evaluates the effect of each input variable on the model output and plays a key role in understanding the relationships among variables. This process involves controlled variations in the values of input parameters such as WFS/C, WFS/FA, FA/TA, W/C, CA/C, 1000SP/C, and Age while keeping other factors constant. In this analysis, the values of input variables are changed relatively at different levels to determine the response of the model to these variations. Then, the effect of each parameter on the model output is discussed, and their importance in predicting STS and E is determined. For this purpose, the method presented by Liang et al.⁴⁵ has been used to assess the relative sensitivity of each variable.

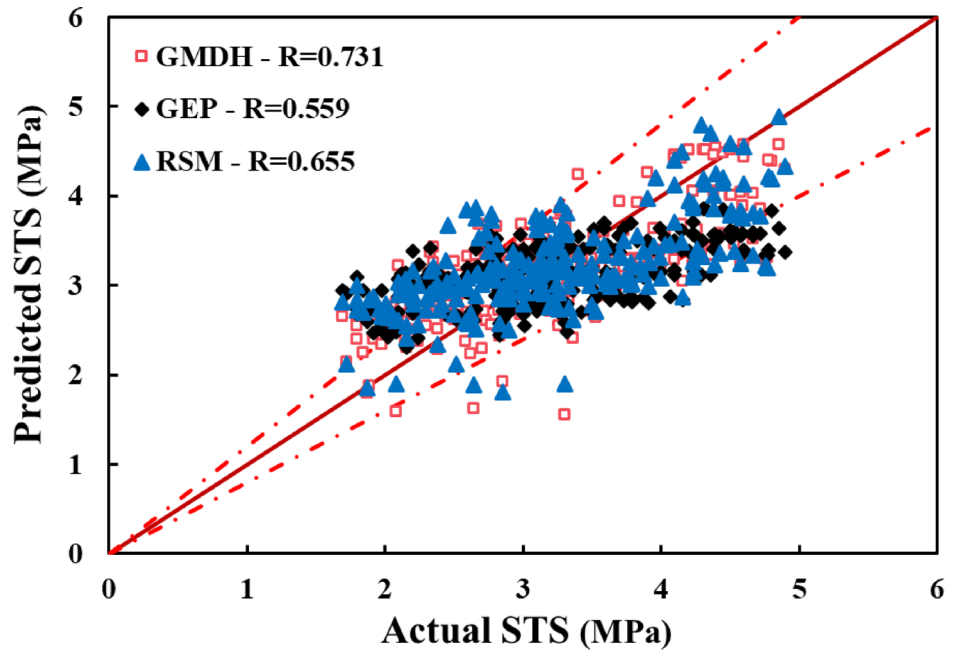


Fig. 6. Correlation between the predicted and actual STS values using the GEP, RSM, and GMDH models.

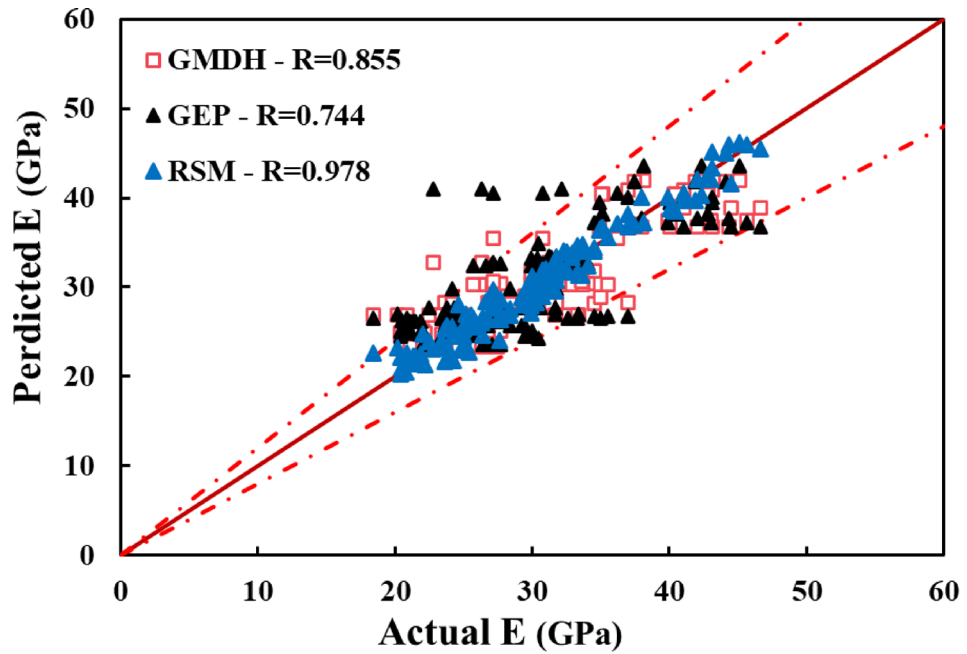


Fig. 7. Correlation between the predicted and actual E values using the GEP, RSM, and GMDH models.

$$SL \text{ of } X_i (\%) = \frac{1}{M} \sum_{j=1}^M \left(\frac{\% \text{ change in output}}{\text{change in input}} \right)_j \times 100 \quad (12)$$

Figures 12 and 13 show the sensitivity analysis of the GMDH model for STS prediction and the RSM model for E prediction, respectively.

The analysis results indicate that the CA/C ratio has the most significant influence on STS, with the model exhibiting high sensitivity to changes in this parameter across all ranges. This finding highlights the strong dependence of STS on the aggregate-to-cement ratio. The WFS/C ratio was identified as the second most influential variable, where higher values lead to a greater model. Also, Ageshowed the lowest sensitivity, which

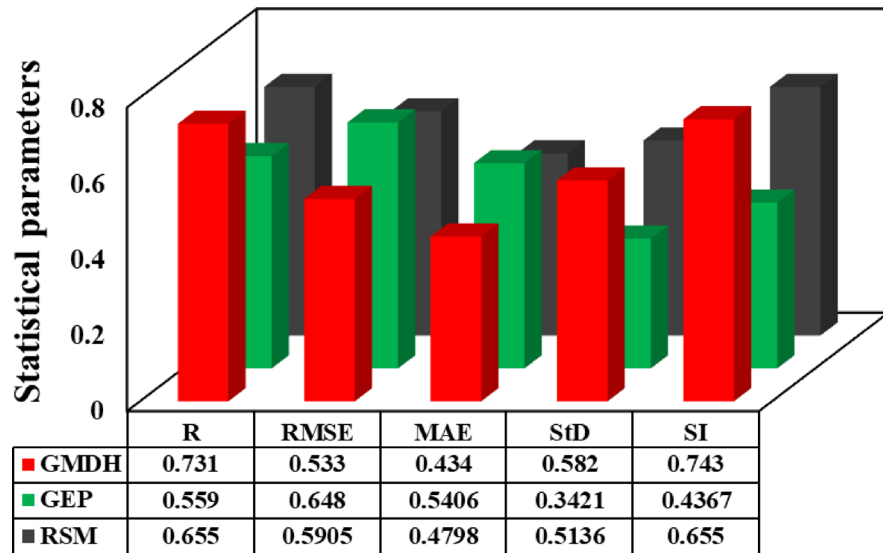


Fig. 8. Comparison of statistical indices of GMDH, GEP and RSM models for STS prediction.

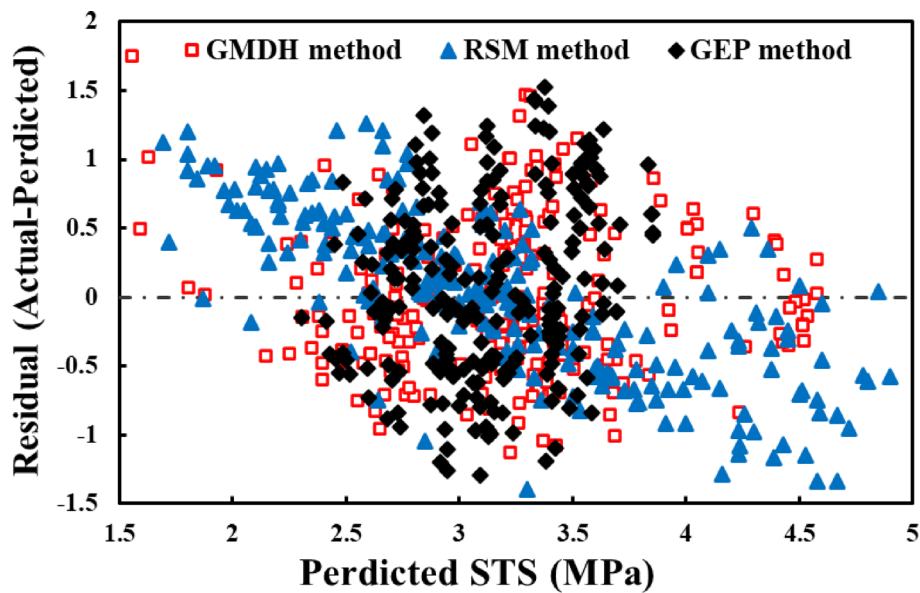


Fig. 9. Residual plots of GMDH, RSM, and GEP models for predicting the STS of concrete.

was considered the least significant factor. These results suggest that optimizing the aggregate-to-cement and recycled sand-to-cement ratios can significantly enhance the prediction and control of splitting tensile strength. In contrast, the effects of other parameters are relatively limited. Overall, this analysis provides valuable insight for the optimal design of concrete mixtures and the enhancement of their mechanical performance.

The sensitivity analysis of the RSM model for predicting the E indicates that the W/C ratio significantly influences the model output. In contrast, other variables, such as WFS/C, CA/C, and FA/TA, exhibit a comparatively lower impact. Additionally, minor variations in the value of 1000SP/C and Age have a negligible effect on the model. Overall, this analysis confirms the importance of accurately controlling the water-cement ratio in designing concrete mixtures to achieve optimal E values.

Conclusion

This study used interpretable and semi-interpretable soft computing methods, including GMDH, GEP, and RSM, were employed to develop predictive relationships for estimating E and STS of concrete containing WFS. In addition, a sensitivity analysis was conducted to evaluate the influence of input parameters on these mechanical properties. The key findings of this research are summarized as follows:

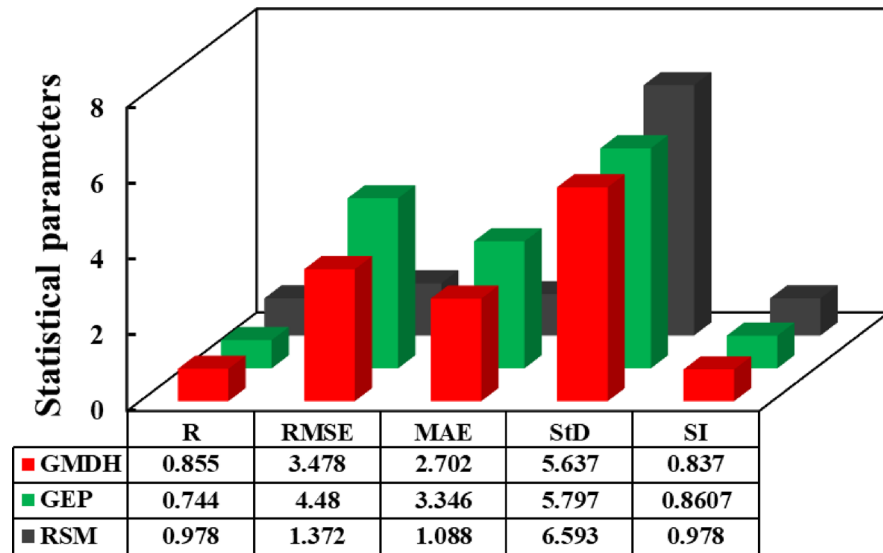


Fig. 10. Comparison of statistical indices of GMDH, GEP and RSM models for E prediction.

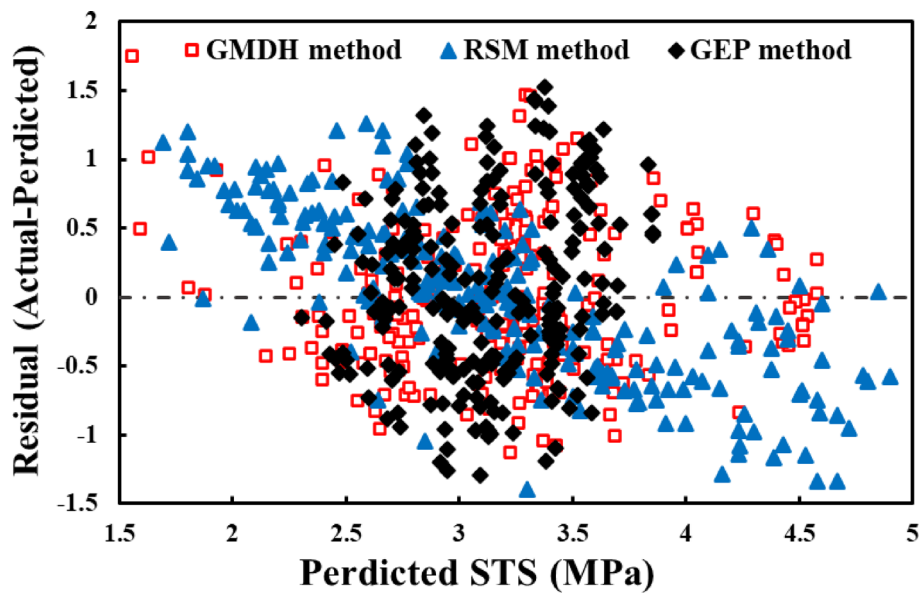


Fig. 11. Residual plots of GMDH, RSM, and GEP models for predicting the E of concrete.

- The GMDH model achieved the highest R-value in predicting STS. This model had the lowest error values, with $RMSE=0.533$ and $MAE=0.434$. Based on all statistical indices, it demonstrated the best overall performance among the evaluated methods.
- Among the three methods used to predict E, the RSM method performed the best with the highest correlation coefficient $R=0.978$ and the lowest errors ($RMSE=1.372$ and $MAE=1.088$). The GMDH method had an acceptable performance but with lower accuracy than RSM, with $R=0.855$ and moderate errors ($RMSE=3.478$ and $MAE=2.702$).
- The sensitivity analysis revealed that the CA/C ratio had the most significant influence on STS. The WFS/C ratio was identified as the second most influential parameter, with model sensitivity increasing as its value increased. Other parameters, such as WFS/FA and 1000SP/C, showed comparatively minor effects on STS.
- Sensitivity analysis of the RSM model for E indicated that the W/C ratio had the most significant influence on E. Other variables, including WFS/C, CA/C, and FA/TA, showed comparatively lower effects.

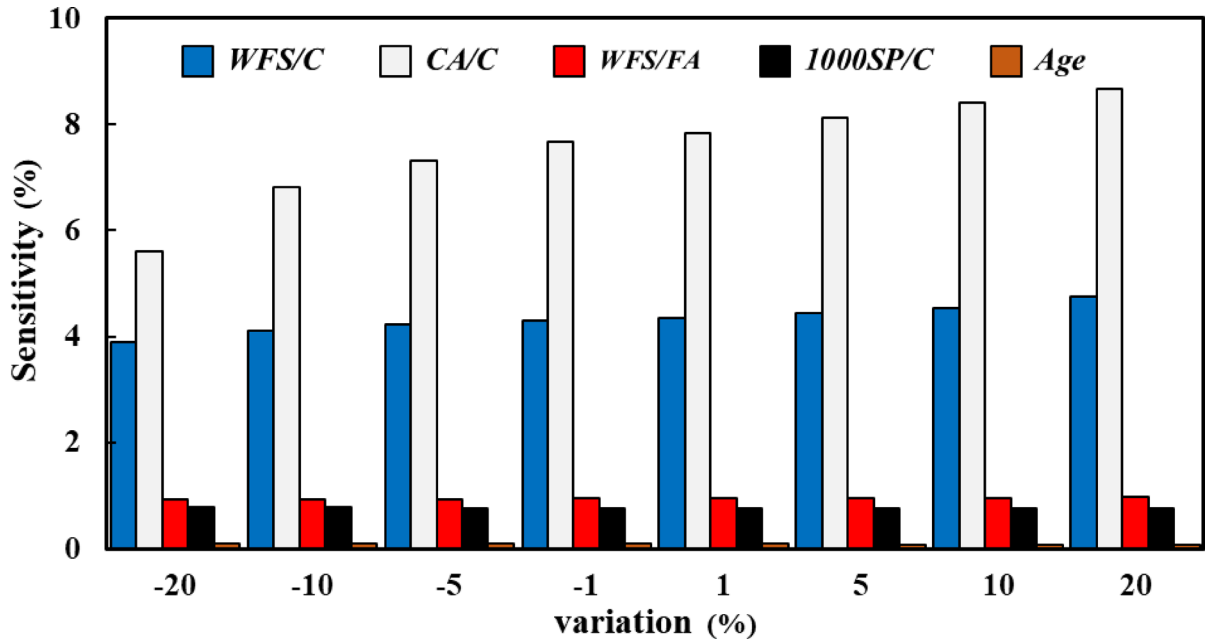


Fig. 12. Sensitivity analysis of input parameters based on the GMDH model for STS prediction.

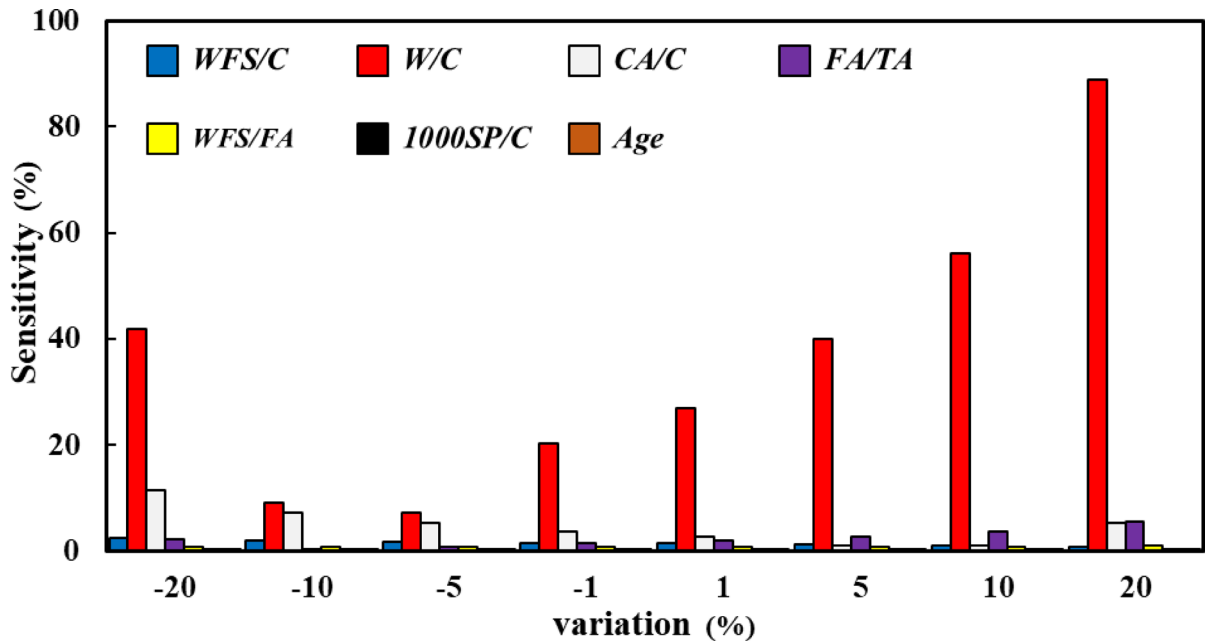


Fig. 13. Sensitivity analysis of input parameters based on the RSM model for E prediction.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to restricted access, but are available from the corresponding author on reasonable request.

Received: 23 September 2025; Accepted: 30 October 2025

Published online: 28 November 2025

References

1. Seyedkazemi, A., Kutanaei, S. S., Joybari, H. R., Rezaei, S. & Davoudi-Kia, A. Investigating the static and dynamic behavior of sheet piles with recycled lightweight materials. *Iran. J. Sci. Technol. Trans. Civil Eng.* **49**(5), 3967–3975 (2025).

2. Matajinimvar, S., Choobbasti, A. J. & Kutanaei, S. S. The effect of construction moisture content on the mechanical, shear and environmental characteristics of clay stabilized with cement and CS: A micro and macro study. *Case Stud. Constr. Mater.* **22**, e04818 (2025).
3. Ramezani, M., Soleimani Kutanaei, S., Seyedkazemi, A. & Esfandiari Fard, A. Enhancing the durability of cement-stabilized clayey sand with zeolite and PVA fibers under wet–dry and freeze–thaw cycles. *Transp. Infrastructure Geotechnology.* **12**(5), 162 (2025).
4. Ramezani, M., Kutanaei, S. S., Seyedkazemi, A. & Fard, A. E. Sustainable stabilization of fiber-reinforced clayey sand using zeolite as partial cement replacement. *Case Stud. Constr. Mater.* e04868 (2025).
5. Fakhrabadi, A., Ghadakpour, M., Choobbasti, A. J. & Kutanaei, S. S. Influence of the non-woven geotextile (NWG) on the engineering properties of clayey-sand treated with copper slag-based geopolymer. *Constr. Build. Mater.* **306**, 124830 (2021).
6. Fakhrabadi, A., Ghadakpour, M., Choobbasti, A. J. & Kutanaei, S. S. Evaluating the durability, microstructure and mechanical properties of a clayey-sandy soil stabilized with copper slag-based geopolymer against wetting-drying cycles. *Bull. Eng. Geol. Environ.* **80**(6), 5031–5051 (2021).
7. Kutanaei, S. S., Afrakoti, M. T. P. & Choobbasti, A. J. Effect of coal waste on grain failure of cement-stabilized sand due to compaction. *Arab. J. Geosci.* **14**(12), 1105 (2021).
8. Janalizadeh Choobbasti, A., Farrokhzad, F. & Nadimi, A. & ... Effects of copper sludge on cemented clay using ultrasonic pulse velocity. *J. Adhes. Sci. Technol.* **33**(4), 433–444.
9. Alonso-Santurde, R., Coz, A., Viguri, J. R. & Andrés, A. Recycling of foundry by-products in the ceramic industry: green and core sand in clay bricks. *Constr. Build. Mater.* **27**(1), 97–106 (2012).
10. Navarro-Blasco, I., Fernández, J. M., Duran, A., Sirera, R. & Álvarez, J. I. A novel use of calcium aluminate cements for recycling waste foundry sand (WFS). *Constr. Build. Mater.* **48**, 218–228 (2013).
11. Behnood, A. & Golafshani, E. M. Machine learning study of the mechanical properties of concretes containing waste foundry sand. *Constr. Build. Mater.* **243**, 118152 (2020).
12. Tavana Amlashi, A. et al. AI-based formulation for mechanical and workability properties of eco-friendly concrete made by waste foundry sand. *J. Mater. Civ. Eng.* **33**(4), 04021038 (2021).
13. Tavakoli, H. R., Omran, O. L. & Kutanaei, S. S. Prediction of energy absorption capability in fiber reinforced self-compacting concrete containing nano-silica particles using artificial neural network. *Latin Am. J. Solids Struct.* **11**, 966–979 (2014).
14. Tavakoli, H., Lotfi-Omran, O., Shiade, M. F. & Kutanaei, S. S. Prediction of combined effects of fibers and nanosilica on the mechanical properties of self-compacting concrete using artificial neural network. *Latin Am. J. Solids Struct.* **11**, 1906–1923 (2014).
15. Mostafaei, Y. & Soleimani Kutanaei, S. Modeling and prediction of embankment dam displacement under earthquake loading using artificial neural networks and swarm optimization algorithm. *Int. J. Geo-Engineering.* **16**(1), 8 (2025).
16. Fahimi, R., Seyedkazemi, A. & Kutanaei, S. S. Prediction and sensitivity analysis of embankment dam settlement under earthquake loading using gene expression programming. *Geomech. Geoeng.* **20**(1), 115–137 (2025).
17. Vafaei, A., Davoudi-Kia, A., Kutanaei, S. S. & Taslimi, M. Data-driven investigation of compressive strength of FRP-confined concrete columns using a unified model based on RSM considering interactions between parameters. *Struct. Concrete.* **25**(3), 2183–2205 (2024).
18. Kutanaei, S. S. & Choobbasti, A. J. Prediction of combined effects of fibers and cement on the mechanical properties of sand using particle swarm optimization algorithm. *J. Adhes. Sci. Technol.* **29**(6), 487–501 (2015).
19. Ghazavi, M. & Afrakoti, M. T. P. Unconfined compressive strength prediction of soils improved with biopolymers: machine learning approach. *Transp. Infrastructure Geotechnology.* **12**(1), 14 (2025).
20. Amin, M. N. et al. Application of soft-computing methods to evaluate the compressive strength of self-compacting concrete. *Materials* **15**(21), 7800 (2022).
21. Faraj, R. H., Mohammed, A. A., Omer, K. M. & Ahmed, H. U. Soft computing techniques to predict the compressive strength of green self-compacting concrete incorporating recycled plastic aggregates and industrial waste ashes. *Clean Technol. Environ. Policy.* **24**(7), 2253–2281 (2022).
22. Rezaei, S., Choobbasti, A. J. & Kutanaei, S. S. Site effect assessment using microtremor measurement, equivalent linear method, and artificial neural network (case study: Babol, Iran). *Arab. J. Geosci.* **8**(3), 1453–1466 (2015).
23. Kutanaei, S. S. & Choobbasti, A. J. Mesh-free modeling of liquefaction around a pipeline under the influence of trench layer. *Acta Geotech.* **10**(3), 343–355 (2015).
24. Mashhadban, H., Beitollahi, A. & Kutanaei, S. S. Identification of soil properties based on accelerometer records and comparison with other methods. *Arab. J. Geosci.* **9**(6), 427 (2016).
25. Choobbasti, A. J., Tavakoli, H. & Kutanaei, S. S. Modeling and optimization of a trench layer location around a pipeline using artificial neural networks and particle swarm optimization algorithm. *Tunn. Undergr. Space Technol.* **40**, 192–202 (2014).
26. Kutanaei, S. S. & Choobbasti, A. J. Prediction of liquefaction potential of sandy soil around a submarine pipeline under earthquake loading. *J. Pipeline Syst. Eng. Pract.* **10**(2), 04019002 (2019).
27. Fereidouni, P., Seyedkazemi, A., Kutanaei, S. S. & Davoudi-Kia, A. Modeling and sensitivity analysis of the compressive strength of recycled brick aggregates concrete using GMDH, GEP and RSM methods. *Results Eng.* **26**, 104891 (2025).
28. Hoseini, S., Seyedkazemi, A., Davoudi-Kia, A. & Kutanaei, S. S. A data mining approach for proposing a relationship to predict self-compaction concrete crack width after the self-healing period. *Results Eng.* 104980 (2025).
29. Tavakoli, H. & Soleimani Kutanaei, S. Evaluation of effect of soil characteristics on the seismic amplification factor using the neural network and reliability concept. *Arab. J. Geosci.* **8**(6), 3881–3891 (2015).
30. Vafaei, A., Choobbasti, A. J., Vafaei, A., Afrakoti, M. P. & Kutanaei, S. S. Prediction of the peak shear strength parameters of Babolsar sand using particles swarm optimization and artificial neural network methods. *Transp. Infrastructure Geotechnology.* **11**(2), 470–500 (2024).
31. Kutanaei, S. S. et al. Application of LRBFDQ and CVBFEM methods for evaluating saturated sand liquefaction around buried pipeline. *J. Pipeline Syst. Eng. Pract.* **13**(1), 04021077 (2022).
32. Arjomand, M. A., Mostafaei, Y. & Kutanaei, S. S. Modeling and sensitivity analysis of bearing capacity in driven piles using hybrid ANN–PSO algorithm. *Arab. J. Geosci.* **15**(3), 309 (2022).
33. Kia, M., Bayat, M., Emadi, A., Kutanaei, S. S. & Ahmadi, H. R. Reliability based seismic fragility analysis of Bridge. *Computers Concrete.* **29**(1), 59 (2022).
34. Seyyedi, S. M., Ghadakpour, M., Bayat, M., Pilehvar, M. & Kutanaei, S. S. & ... CVFEM modeling of fluid flow induced by convective heat transfer from a hot pipe buried in soil. *J. Therm. Anal. Calorim.* **146**(1).
35. Gholampour, S., Taghipour, R., Felourdi, H. K. & Kutanaei, S. S. Investigating the effect of rotational components on the progressive collapse of steel structures. *Eng. Fail. Anal.* **121**, 105094 (2021).
36. Javed, M. F. et al. Comparative analysis of various machine learning algorithms to predict strength properties of sustainable green concrete containing waste foundry sand. *Sci. Rep.* **14**(1), 14617 (2024).
37. Iqbal, M. F. et al. Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. *J. Hazard. Mater.* **384**, 121322 (2020).
38. Jakubowski, J. & Tomczak, K. Deep learning metasensor for crack-width assessment and self-healing evaluation in concrete. *Constr. Build. Mater.* **422**, 135768 (2024).
39. Kumar, A. et al. Compressive strength prediction of lightweight concrete: machine learning models. *Sustainability* **14**(4), 2404 (2022).

40. Shahrokhishahraki, M., Malekpour, M., Mirvalad, S. & Faraone, G. Machine learning predictions for optimal cement content in sustainable concrete constructions. *J. Building Eng.* **82**, 108160 (2024).
41. Singh, G. & Siddique, R. Effect of waste foundry sand (WFS) as partial replacement of sand on the strength, ultrasonic pulse velocity and permeability of concrete. *Constr. Build. Mater.* **26**(1), 416–422 (2012).
42. Manoharan, T., Lakshmanan, D., Mylsamy, K., Sivakumar, P. & Sircar, A. Engineering properties of concrete with partial utilization of used foundry sand. *Waste Manage.* **71**, 454–460 (2018).
43. Gurumoorthy, N. & Arunachalam, K. Micro and mechanical behaviour of treated used foundry sand concrete. *Constr. Build. Mater.* **123**, 184–190 (2016).
44. Torres, A., Bartlett, L. & Pilgrim, C. Effect of foundry waste on the mechanical properties of Portland cement concrete. *Constr. Build. Mater.* **135**, 674–681 (2017).
45. Liong, S., Lim, W. & Paudyal, G. River stage forecasting in Bangladesh: neural network approach. *J. Comput. Civil Eng.* **14**(1), 1–8 (2000).

Author contributions

Elahe Moshaver, Ali Seyedkazemi, Saba Jahangir contributed to the preparation of all parts of the article.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025