Advances in Nano Research, *Vol. 16, No. 4 (2024) 375-394*
https://doi.org/10.12989/anr.2024.16.4.375

Predicting the splitting tensile strength of manufactured-sand concrete containing stone nano-powder through advanced machine learning techniques

Manish Kewalramani[1], Hanan Samadi[2], Adil Hussein Mohammed[3], Arsalan Mahmoodzadeh[2], Ibrahim Albaijan*[4], Hawkar Hashim Ibrahim[5] and Saleh Alsulamy[6]

[1]*Department of Civil Engineering, College of Engineering, Abu Dhabi University, Abu Dhabi, UAE*
[2]*IRO, Civil Engineering Department, University of Halabja, Halabja, 46018, Iraq*
[3]*Department of Communication and Computer Engineering, Faculty of Engineering, Cihan University-Erbil, Kurdistan Region, Iraq*
[4]*Mechanical Engineering Department, College of Engineering at Al-Kharj, Prince Sattam Bin Abdulaziz University, Al Kharj 16273, Saudi Arabia*
[5]*Department of Civil Engineering, College of Engineering, Salahaddin University-Erbil, 44002 Erbil, Kurdistan Region, Iraq*
[6]*Department of Architecture and Planning, College of Engineering, King Khalid University, Abha 61421, Saudi Arabia*

(*Received June 8, 2022, Revised March 6, 2024, Accepted March 7, 2024*)

Abstract. The extensive utilization of concrete has given rise to environmental concerns, specifically concerning the depletion of river sand. To address this issue, waste deposits can provide manufactured-sand (MS) as a substitute for river sand. The objective of this study is to explore the application of machine learning techniques to facilitate the production of manufactured-sand concrete (MSC) containing stone nano-powder through estimating the splitting tensile strength (STS) containing compressive strength of cement (CSC), tensile strength of cement (TSC), curing age (CA), maximum size of the crushed stone (Dmax), stone nano-powder content (SNC), fineness modulus of sand (FMS), water to cement ratio (W/C), sand ratio (SR), and slump (S). To achieve this goal, a total of 310 data points, encompassing nine influential factors affecting the mechanical properties of MSC, are collected through laboratory tests. Subsequently, the gathered dataset is divided into two subsets, one for training and the other for testing; comprising 90% (280 samples) and 10% (30 samples) of the total data, respectively. By employing the generated dataset, novel models were developed for evaluating the STS of MSC in relation to the nine input features. The analysis results revealed significant correlations between the CSC and the curing age CA with STS. Moreover, when delving into sensitivity analysis using an empirical model, it becomes apparent that parameters such as the FMS and the W/C exert minimal influence on the STS. We employed various loss functions to gauge the effectiveness and precision of our methodologies. Impressively, the outcomes of our devised models exhibited commendable accuracy and reliability, with all models displaying an R-squared value surpassing 0.75 and loss function values approaching insignificance. To further refine the estimation of STS for engineering endeavors, we also developed a user-friendly graphical interface for our machine learning models. These proposed models present a practical alternative to laborious, expensive, and complex laboratory techniques, thereby simplifying the production of mortar specimens.
Keywords: machine learning; manufactured-sand concrete; sensitivity analysis; splitting tensile strength; stone nano-powder

1. Introduction

The global building industry currently produces and utilizes a minimum of ten billion metric tons of concrete annually (Zhang *et al*. 2020). This number is anticipated to rise in response to growing demand for these construction services. Consequently, the large-scale production of concrete has resulted in excessive extraction of river sand from certain regions. The depletion of river sand is a significant environmental concern due to its nonrenewable nature. This issue encompasses various problems such as reduced riverbed levels, erosion of river banks, risks to the stability of river banks and bridges, decreased vegetation along riverbanks, disruption of aquatic biodiversity, and agricultural impacts caused by lower water tables (Zhang *et al*. 2020).

Crushing waste rock deposits generates manufactured-sand, which possesses a more angular shape and rougher surface texture in comparison to river sand (Li *et al*. 2011). While MS enhances the interlocking of particles in concrete, thereby increasing its strength, it is less workable than river sand when subjected to the same water content (Zhao *et al*. 2017). The manufacturing process of manufactured-sand concrete introduces micro-fines, which are particles that pass through a 75 µm sieve at a range of 5-20%, and these micro-fines affect various characteristics of concrete, including workability, compressive and flexural strengths, and freezing resistance (Li *et al*. 2011). Consequently, when predicting the mechanical properties of MSC, additional considerations are necessary compared to concrete made with river sand. However, conducting laboratory tests on a large number of samples to identify patterns in concrete characteristics requires significant time and resources. In response to these challenges, numerous

*Corresponding author, Assistant Professor,
E-mail: i.albaijan@psau.edu.sa

ISSN: 2287-237X (Print), 2287-2388 (Online)

governments have implemented restrictions on river sand mining, as it has become a scarce resource with high transport costs from natural deposits. Consequently, there is an urgent need to regulate sand mining and identify sustainable alternatives to river sand. Utilizing locally available and durable aggregates is crucial to achieving long-term economic and developmental objectives. This approach can help minimize construction expenses, energy consumption, and carbon dioxide emissions. Manufactured-sand (MS) has emerged as a potentially cost-effective substitute for river sand that also adheres to environmental standards. Incorporating MS as fine aggregates in concrete not only has the potential to reduce construction costs but also fosters the development of environmentally conscious building practices (Zhang *et al*. 2020, Zhao *et al*. 2017). Consequently, manufactured-sand concrete (MSC) has progressively gained essential status as an eco-friendly component within the construction industry (Zhao *et al*. 2017).

To address this challenge, the development of a predictive model for concrete characteristics prior to production becomes crucial. Currently, concrete characteristic predictions rely on linear and nonlinear regression methods (Zhao *et al*. 2017). Nevertheless, employing traditional statistical models to replicate the intricate nonlinear dynamics between concrete and the variables that impact it poses a significant challenge. This challenge arises from the inherent difficulty in precisely determining the coefficients within these models (Chou & Pham, 2013). Conducting laboratory research to study the mechanical properties of concrete also presents its own set of challenges. In order to overcome the limitations of traditional approaches, advanced methods such as machine learning (ML) and deep learning (DL) are often employed, as they eliminate the need for time-consuming and expensive experiments (Chou and Pham 2013). ML methods do not rely on explicit equations when forecasting concrete properties.

In recent years, the utilization of ML techniques, such as artificial neural networks (ANNs) and fuzzy-based algorithms has emerged as cost-effective and accurate methods for assessing various engineering variables, particularly the mechanical properties of concrete, including compressive strength (CS) and tensile strength (TS) (Zhang *et al*. 2020). A broad range of intelligent processors have been employed to investigate the correlations between each parameter and the characteristics of concrete mixtures. Yaseen *et al*. (2018) demonstrated the effectiveness of the extreme learning machine (ELM) and multivariate adaptive regression spline (MARS) model in modeling CS for lightweight foamed concrete. Notably, established models like the M5 tree and MARS exhibited improved prediction accuracy when combined with the suggested model. Ling *et al*. (2019) utilized support vector machine (SVM) coupled with the K-fold cross-validation method to estimate the CS in maritime settings, achieving superior performance compared to decision tree (DT) and ANN models. Furthermore, Tinoco *et al*. (2020) compared ANN, SVM, and multiple regression techniques for evaluating the CS of soil-cement mixtures. Key factors such as cement concentration, water/cement (W/C) ratio, curing age (CA), and organic matter content were identified as significant influencers of concrete behavior, with SVM proving to be the most effective method. Kaloop *et al*. (2020) employed gradient boosting regression (GBR) to calculate the CS of high-performance concrete, obtaining a good correlation coefficient of 0.965 and thus providing reliable estimations. Zhang *et al*. (2020) showcased the potential of three tree-based models, including GBR, random forest (RF), and decision tree regression (DTR), to forecast the mechanical characteristics of manufactured sand concrete (MSC), such as uniaxial compressive strength (UCS) and splitting tensile strength (STS).

Gap identification in previous researches and novelty of this present work - The aim of this study is to evaluate the efficacy of eleven established ML, fuzzy-based, and genetic methods in predicting the mechanical properties of MSC. These methods, including SVM, Takagi-Sugeno fuzzy model (TSF), Mamdani fuzzy inference system (MFIS), ELM, deep feed-forward neural network (DFF), ANN with varying numbers of neurons in the hidden layer (7, 13, 17, 22), gene expression programming (GEP), and radial basis function (RBF), have not been extensively explored in this research area. While previous ML-based studies have primarily focused on the strength of MSC, the STS parameter has been comparatively understudied (Zhang *et al*. 2020). The dataset used in this study consists of nine features that influence the mechanical properties of MSC, namely compressive strength of cement (CSC), tensile strength of cement (TSC), CA, maximum size of the crushed stone (D_{max}), stone nano-powder content (SNC), fineness modulus of sand (FMS), W/C, sand ratio (SR), and slump (S). The influence of these features on the STS of MSC is analyzed using sensitivity analysis based on an empirical model. By comparing the ML-estimated results with the laboratory results, the study evaluates the performance of each ML model and identifies the most suitable model for predicting STS in stone nano-powder (SNP)-reinforced MSC. The flowchart of present study is indicated in Fig. 1. Hence, this study prioritizes the use of STS as the output variable for the ML models. Past research has shown that the inclusion of a specific proportion of stone powder in MSC significantly affects both UCS and STS (Zhao *et al*. 2017, Zhang *et al*. 2020). However, there is limited research that specifically investigates the impact of stone powder on MSC strength, and even in those studies, only stone micro-powder has been utilized. This study introduces the incorporation of stone nano-powder (SN) for the first time in the production of MSC specimens. Moreover, unlike previous works, this study conducts hypothesis tests such as P-P plot of regression standardized residuals, Q-Q plot, Chi-squared automatic interaction detection (CHAID), and collinearity analysis to select the research hypothesis. Additionally, the impact of multi-collinearity on the reliability analysis of STS is investigated. The ML models are trained and tested using a dataset comprising 310 laboratory data points, with 90% allocated for training and 10% for testing.

Moving forward, this research paper will be structured as follows: Section 2 conducts statistical analysis on the

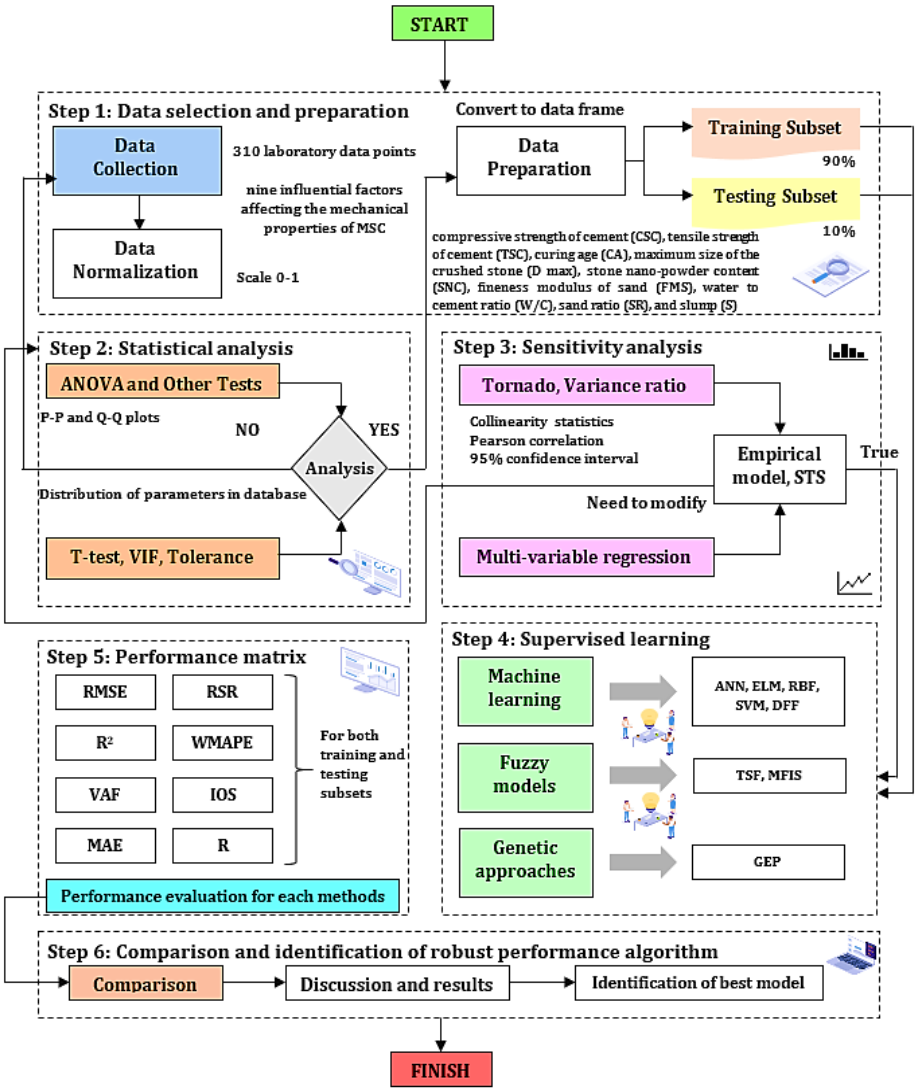

Fig. 1. The flowchart of current study to predict STS

dataset, including data preprocessing. Furthermore, hypothesis tests such as P-P plot of regression standardized residuals, Q-Q plot, and collinearity analysis are carried out to select the research hypothesis. Section 3 provides the empirical model, statistical metrics, and sensitivity analysis based on an empirical formula. Section 4 presents an overview of the eleven ML methods utilized and introduces the statistical metrics employed to assess the performance of the ML models. Also, in section 4, the research findings and their interpretations, the results of ML models are presented. Finally, section 5 concludes the paper by summarizing the key findings and contributions.

Overall, this study makes several notable contributions. Firstly, it utilizes a novel laboratory database that incorporates nine features impacting the STS of MSC. Secondly, it evaluates the performance of eleven established ML methods in estimating the STS of MSC. Thirdly, it examines the effect of each input factor on the STS of MSC through comprehensive sensitivity analysis, utilizing tornado graphs. Fourthly, it compares the behavior of ML models with laboratory tests. Lastly, it identifies the most powerful ML-based model for estimating STS in SNP-reinforced MSC. These proposed ML models provide a viable alternative to laborious, expensive, and complex laboratory methods, thereby facilitating MSC production.

2. Database development

The measurement of TS in concrete can be achieved through either the direct tensile strength (DTS) test or the STS test. Although the DTS test is more intricate due to the requirement for axial tension assurance (Zhao *et al*. 2017), it is commonly utilized in specifications and recommendations because it provides a more accurate representation of concrete's tensile characteristics. On the other hand, the STS test is more frequently employed in research and engineering as it is easier to conduct, more reliable, and exhibits less variation. Additionally, cylindrical and cubic specimens are the standard shapes used in STS testing (Zhao *et al*. 2017). Both cylindrical specimens with dimensions of 300 mm × 150 mm diameter and cubical

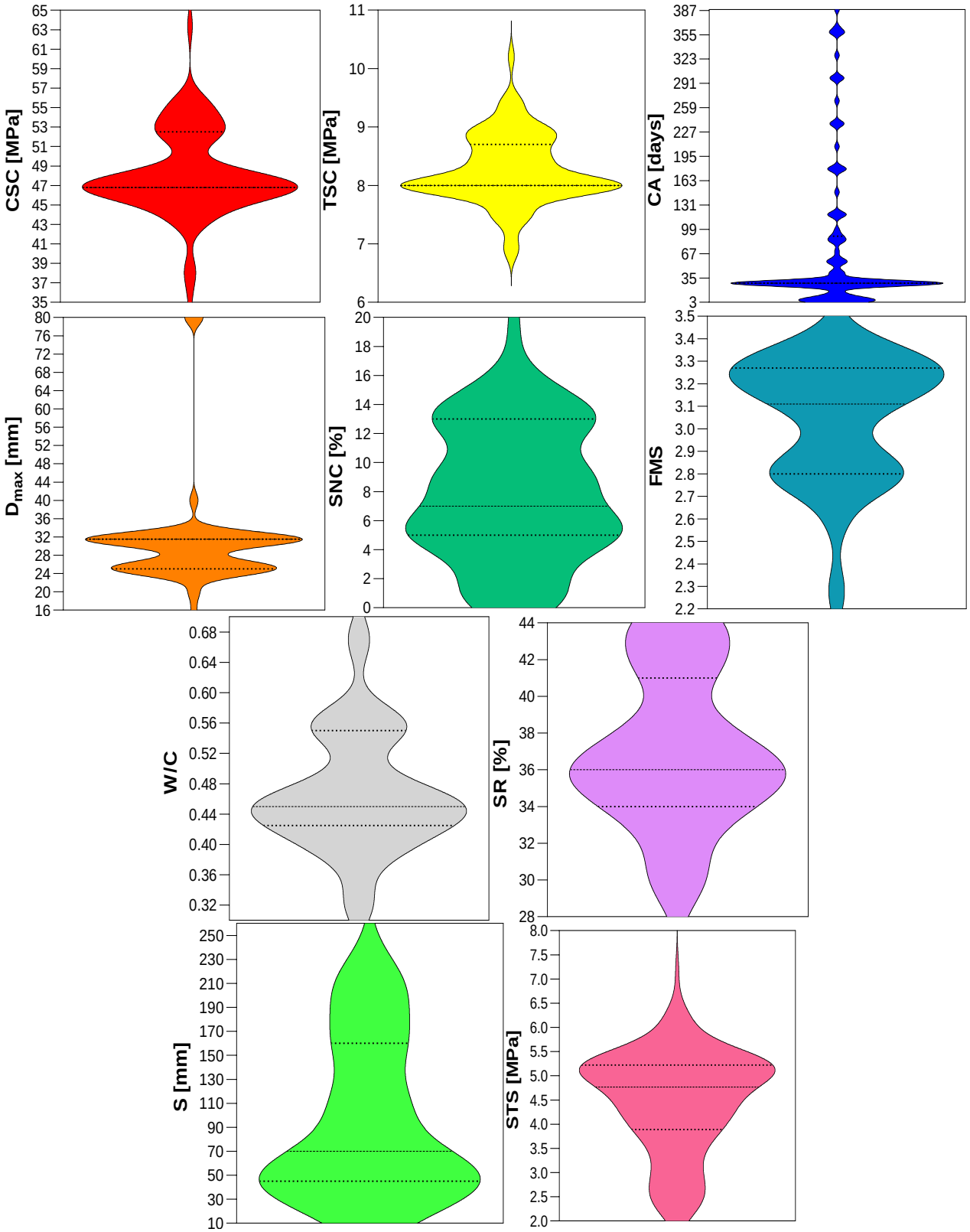

Fig. 2 Violin plots of input and output parameters

specimens with dimensions of 150 mm consistently yield similar test results (Pan *et al*. 2022).

This research paper delves into an extensive investigation on the STS of MSC under varying W/C ratios and SNC conditions, spanning a comprehensive CA period of up to 388 days. The study emphasizes the critical importance of understanding the TS characteristics of MSC, particularly in the context of reinforced concrete structures, driven by the paucity of comprehensive research in this domain.

Table 1 Summary results of statistical analysis of parameters recorded in the database

Statistics											
		CSC	TSC	CA	D	SNC	FMS	W/C	SR	S	STS
Unit		MPa	MPa	days	mm	%	-	-	%	mm	MPa
Count	Valid	310	310	310	310	310	310	310	310	310	310
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		48.1719	8.2629	77.1516	30.688	7.9581	3.0491	0.4668	37.0052	95.9484	4.4808
Std. Error of Mean		0.24218	0.0332	5.67106	0.6274	0.2652	0.0153	0.0044	0.23735	3.71494	0.0591
Median		47.1478[a]	8.1988[a]	30.8000[a]	30.338[a]	7.1613[a]	3.1186[a]	0.4506[a]	36.0338[a]	68.0000[a]	4.7700[a]
Mode		46.8000	8.0000	28.0000	31.500	5.0000	2.7700[c]	0.4500	36.0000	50.0000	5.2200[c]
Std. Deviation		4.26401	0.5856	99.8492	11.047	4.6694	0.2708	.07791	4.17890	65.4081	1.0415
Variance		18.1820	0.3430	9969.87	122.04	21.804	0.0730	0.0060	17.4630	4278.23	1.0850
Skewness		0.43500	0.4430	1.78800	3.6740	0.1930	-0.888	0.6370	0.18500	0.67800	-0.5470
Std. Error of Skewness		0.13800	0.1380	0.13800	0.1380	0.1380	0.1380	0.1380	0.13800	0.13800	0.1380
Kurtosis		1.78200	0.7730	2.03300	14.099	-0.748	0.3530	0.5300	-0.7310	-0.9000	-0.2720
Std. Error of Kurtosis		0.27600	0.2760	0.27600	0.2760	0.2760	0.2760	0.2760	0.27600	0.27600	0.2760
Range		27.9000	3.3000	385.000	64.000	20.000	1.3000	0.3800	16.0000	249.000	5.2700
Minimum		35.5000	6.9000	3.00000	16.000	0.0000	2.2000	0.3100	28.0000	11.0000	2.0000
Maximum		63.4000	10.200	388.000	80.000	20.000	3.5000	0.6900	44.0000	260.000	7.2700
Sum		14933.3	2561.5	23917.0	9513.5	2467.0	945.22	144.71	11471.6	29744.0	1389.0
Percentiles	10	44.1148[b]	7.6556[b]	6.16980[b]	22.072[b]	1.3833[b]	2.6896[b]	0.3858[b]	31.5122[b]	29.3750[b]	2.7633[b]
	20	46.3625	7.9271	14.8936	24.864	4.3651	2.7889	0.4136	33.8696	44.0370	3.5900
	25	46.4701	7.9493	17.9716	25.338	4.7587	2.8122	0.4256	34.2111	45.7407	3.8860
	30	46.5778	7.9714	21.0496	25.713	5.0706	2.8500	0.4290	34.7273	49.1875	4.0600
	40	46.7931	8.0521	27.2057	26.463	5.7273	3.0238	0.4430	35.7813	53.7705	4.3950
	50	47.1478	8.1988	30.8000	30.338	7.1613	3.1186	0.4506	36.0338	68.0000	4.7700
	60	47.5072	8.3456	34.0148	30.938	9.0820	3.1973	0.4597	37.0714	104.444	4.9925
	70	48.0966	8.4907	59.1429	31.715	10.923	3.2417	0.4776	38.4842	130.000	5.1725
	75	52.1923	8.7160	88.0000	33.382	12.381	3.2608	0.5408	40.9167	157.000	5.2225
	80	52.6154	8.8278	118.833	35.050	12.873	3.2827	0.5515	41.8781	166.400	5.2875
	90	53.7415	8.9900	255.500	38.386	14.062	3.3320	0.5623	43.3803	202.222	5.4667

a. Calculated from grouped data.
b. Percentiles are calculated from grouped data.
c. Multiple modes exist. The smallest value is shown.

The methodology employed in this research involved meticulous preparation of MSC samples, taking into account the absorption characteristics of SNC in the initial mixing water. It is noteworthy that, in comparison to concrete formulations utilizing natural sand, the sand ratio in MSC was adjusted by approximately 2% for every 2% - 3% increment in SNC content within manufactured sand, while witnessing a corresponding decrease of about 1% - 2% (Ding *et al*. 2016).

A rigorous experimental setup was established, resulting in the preparation of 310 cubic samples, each with dimensions of 15×15×15 cm, encompassing a diverse range of mixing designs. Following the determination of the specified curing age for each sample, the STS tests were meticulously carried out. Subsequent to the experimental phase, the laboratory findings were meticulously documented for each MSC sample, thus compiling a robust dataset comprising 310 data points, incorporating nine key parameters influencing the STS of the concrete specimens.

In order to effectively analyze the dataset and derive meaningful insights, it was partitioned into distinct subsets, with 90% allocated for training purposes and the remaining 10% reserved for validating the performance of ML models. This strategic division ensured the robustness and reliability of the predictive models developed in this study.

Table 1 showcases statistical metrics extracted from the dataset under scrutiny. These metrics offer concise and objective insights into different facets of the dataset, including measures of central tendency, variability, and distribution. Researchers can utilize these metrics to succinctly depict key characteristics of the dataset.

In Fig. 2, the violin plots vividly showcase the distribution of each input parameter and the output parameter. These plots provide compelling evidence that the

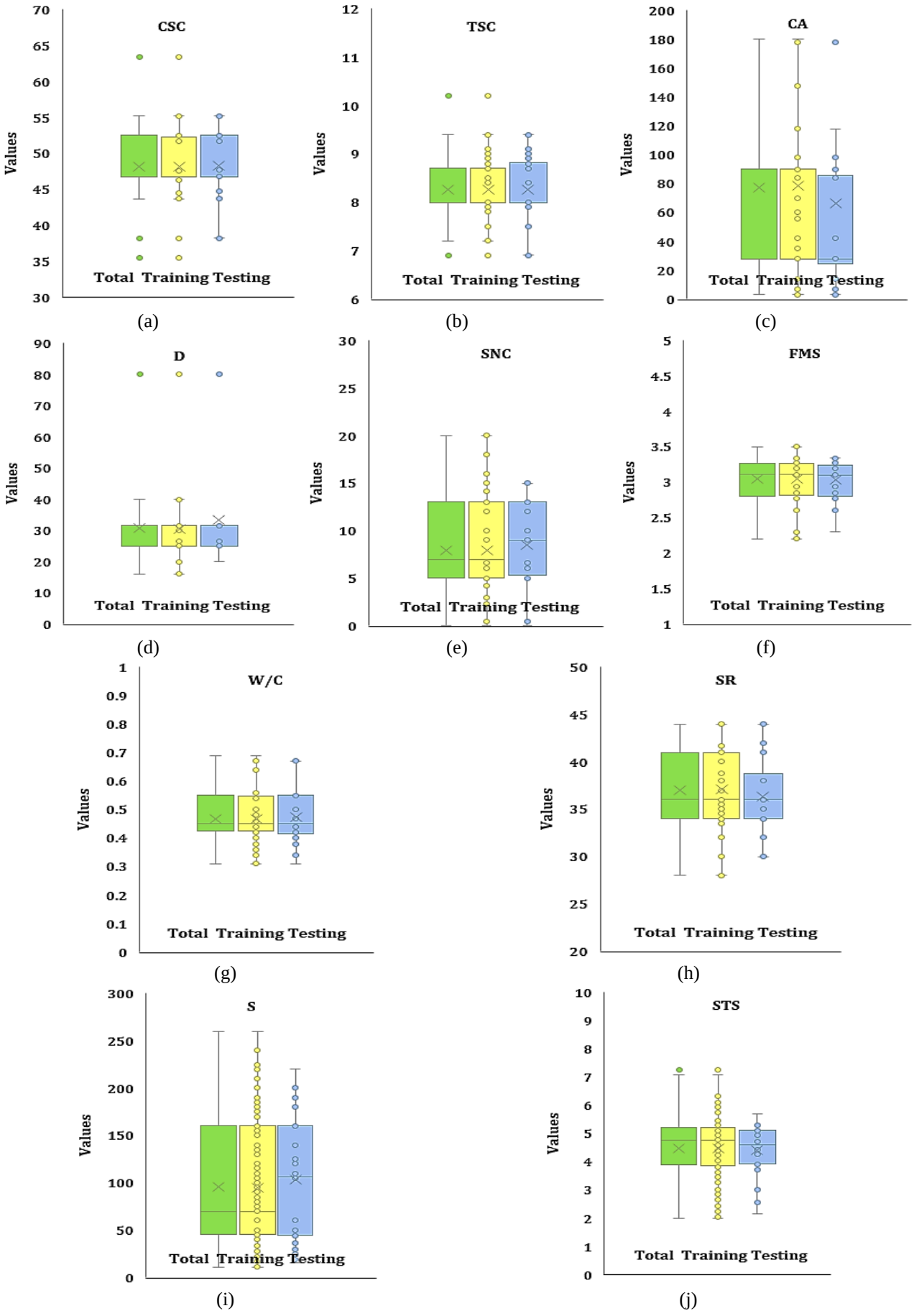

Fig. 3 Boxplots for defined parameters

Table 2 Anti-image matrices of parameters

		CSC	TSC	CA	D	SNC	FMS	WC	SR	S	STS
Anti-image Covariance	CSC	0.083	-0.079	0.027	-0.074	0.095	0.054	0.028	-0.057	-0.001	-0.089
	TSC	-0.079	0.084	-0.010	0.075	-0.086	-0.063	-0.034	0.062	-0.008	0.086
	CA	0.027	-0.010	0.646	-0.067	-0.048	-0.057	0.026	-0.108	0.024	-0.259
	D	-0.074	0.075	-0.067	0.177	-0.108	-0.022	-0.150	.0169	0.011	0.105
	SNC	0.095	-0.086	-0.048	-0.108	0.610	0.099	0.007	-0.021	0.089	-0.011
	FMS	0.054	-0.063	-0.057	-0.022	0.099	0.881	0.028	-0.010	0.062	-0.068
	WC	0.028	-.0034	0.026	-0.150	0.007	0.028	0.260	-0.199	0.019	0.004
	SR	-0.057	0.062	-0.108	0.169	-0.021	-0.010	-0.199	0.300	-0.092	0.124
	S	-0.001	-0.008	0.024	0.011	0.089	0.062	0.019	-0.092	0.824	-0.043
	STS	-0.089	0.086	-0.259	0.105	-0.011	-0.068	0.004	0.124	-0.043	0.417
Anti-image Correlation	CSC	0.330[a]	-0.949	0.117	-0.608	0.423	0.198	0.187	-0.360	-0.004	-0.476
	TSC	-0.949	0.336[a]	-0.042	0.615	-0.379	-0.232	-0.229	0.393	-0.031	0.461
	CA	0.117	-0.042	0.479[a]	-0.199	-0.076	-0.075	0.064	-0.244	0.032	-0.499
	D	-0.608	0.615	-0.199	0.285[a]	-0.330	-0.055	-0.698	0.735	0.028	0.386
	SNC	0.423	-0.379	-0.076	-0.330	0.510[a]	0.134	0.018	-0.049	0.126	-0.022
	FMS	0.198	-0.232	-0.075	-0.055	0.134	0.489[a]	0.058	-0.020	0.073	-0.111
	WC	0.187	-0.229	0.064	-0.698	0.018	0.058	0.434[a]	-0.713	0.040	0.012
	SR	-0.360	0.393	-0.244	0.735	-0.049	-0.020	-0.713	0.212[a]	-0.185	0.350
	S	-0.004	-0.031	0.032	0.028	0.126	0.073	0.040	-0.185	0.811[a]	-0.073
	STS	-0.476	0.461	-0.499	0.386	-0.022	-0.111	0.012	0.350	-0.073	0.406[a]

a. Measures of Sampling Adequacy (MSA)

Table 3 Comprehensive explanation of variance by components through PCA analysis.

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.377	23.775	23.775	2.377	23.775	23.775	2.09	20.991	20.991
2	2.138	21.381	45.156	2.138	21.381	45.156	1.98	19.860	40.851
3	1.605	16.051	61.207	1.605	16.051	61.207	1.76	17.609	58.460
4	1.109	11.091	72.298	1.109	11.091	72.298	1.38	13.838	72.298
5	0.926	9.262	81.560						
6	0.681	6.806	88.366						
7	0.642	6.416	94.782						
8	0.348	3.482	98.263						
9	0.138	1.380	99.643						
10	0.036	0.357	100.000						

Extraction Method: Principal Component Analysis.

input and output parameters possess highly favorable distributions. Their remarkable adherence to acceptable ranges not only validates their suitability but also underscores the meticulousness of the analysis conducted.

Fig. 3 exhibits boxplots categorizing the specified parameters within the database. This visual representation facilitates a comparative examination of these parameters across varying strengths, empowering researchers to pinpoint any discrepancies or outliers. The boxplots succinctly depict the statistical distribution, median, and outliers of each parameter within the total, training, and testing subsets. Such insights gleaned offer valuable understanding into the variations in STS.

2.1 Data preprocessing

In the process of preparing numerical data for analysis, it's crucial to perform data scaling. Scaling ensures that the

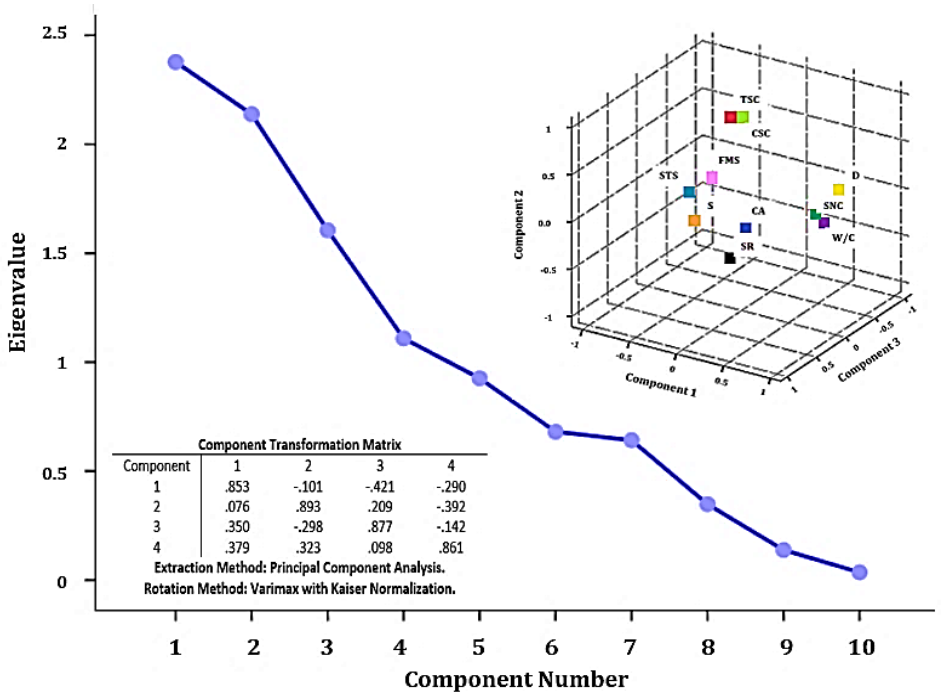

Fig. 4 The matrix of component transformation employing the Varimax technique with Kaiser normalization rotation

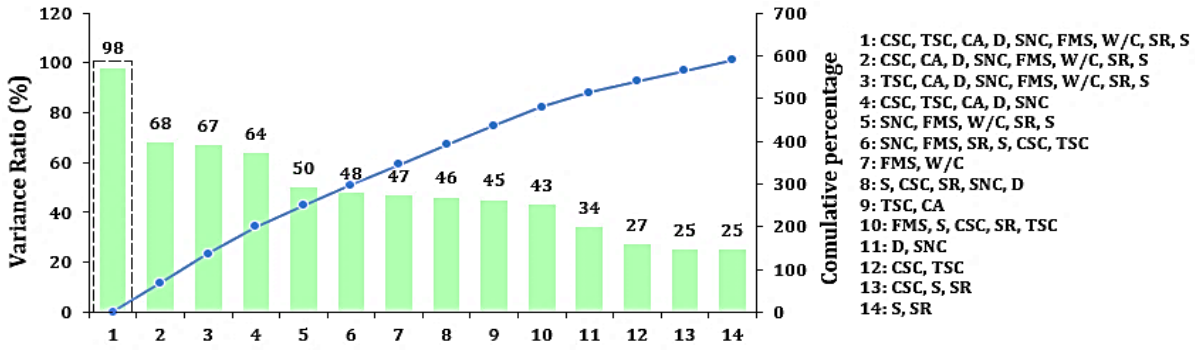

Fig. 5 Ratio of variance for each component and groups of specified features

data is suitable for various machine learning algorithms, enhancing their efficiency. To achieve this, different scaling methods are employed. In our research, we utilize the StandardScaler method to accurately represent the numerical features being studied. This method is suitable when the features in the dataset have varying ranges or are measured using different units. StandardScaler adjusts the distribution of values to align the mean with 0 and the standard deviation with 1. The formula for standard scaling is outlined below.

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where, x is a set of the observed values present in each defined factor, x_{min} is the minimum values of factor, and x_{max} is the maximum values in factor.

2.2. Statistical analysis

In this section, we conduct a principal component analysis (PCA) on the dataset. PCA is a statistical method used to identify patterns and inconsistencies in multivariate data. It also helps in reducing the dimensionality of complex datasets, such as the number of independent variables. In PCA, principal components (PCs) are combinations of the original variables that effectively capture the variability in the data. Tables 2 and 3 present the anti-image matrix and the explained total variance, respectively. Furthermore, Fig. 4 displays the component transformation matrix obtained using the Varimax method with Kaiser normalization rotation. Additionally, Fig. 5 illustrates the variance ratio of each component. It's important to note that using nine features (as listed in Table 1) explains 98% of the variance, indicating it as the most suitable model for calculating STS in terms of statistical analysis.

Tree classified mode with Chi-squared automatic interaction detection (CHAID) growing method enables decision making, identifies important variables, reveals interactions, and meaningful subgroups in the database. The results of tree classified mode with CHAID growing method is presented in Fig. 6. The applied structure to this

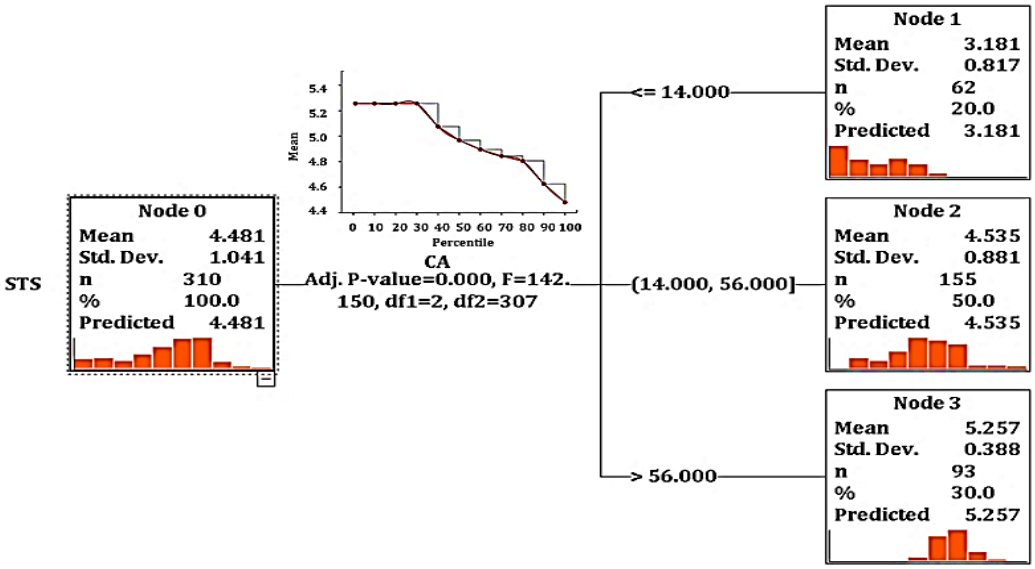

Fig. 6 The results of tree classify CHAID method for STS

Table 4 Structure of tree classify mode with CHAID growing method

Specification	Values
Growing method	CHAID
Percentile increment	10%
Maximum tree depth	3
Parent node	100
Number of terminal nodes	2
depth	1

mode is indicated in Table 4. As can be seen, this diagram and corresponding structure illustrates the hierarchy of predictor variables and their thresholds, facilitating the understanding and communication of the analysis results.

3. Empirical model

For multi-variable regression analysis, nine parameters (see Table 1) serve as independent variables, while STS acts as the dependent value. This analysis utilizes forward stepwise regression analysis (FSR) to evaluate the impact of each variable on STS. Table 5 presents the analysis of variance (ANOVA) and relevant statistical parameters derived from the formulated equation to assess STS. As indicated in Table 5, the ANOVA analysis and regression coefficients meet the criteria, validating the creation of new models for estimating STS. Considering the collinearity statistics and covariance analysis results, the identified correlations and coefficients prove accurate. The proposed empirical equation for calculating STS is outlined in Eq. 2.

$$STS\ (Mpa) = SP + MP + PP \quad (2)$$

where SP: strength parameters:

$$SP = 1.064\ CSC - 1.08\ TSC$$

MP: mechanical properties:

$$MP = 0.401\ CA - 0.592\ D + 0.077\ FMS$$

PP: physical properties:

$$PP = 0.019\ SNC - 0.016\ W/C + 0.052\ S - 0.412\ SR$$

Moreover, Fig. 7 displays four diagrams: a histogram depicting the regression standardized residual, a normal P-P plot illustrating the regression standardized residual, a histogram showcasing the predicted error rate, and a depiction of the regression standardized residual correlated with regression standardized predicted values. These diagrams pertain to the dependent parameter (STS), serving as the output factor, based on nine input parameters acting as independent variables. As shown in these diagrams, the results indicate the reliable outcomes with high accuracy in terms of statistics.

3.1 Statistical metrics

The effectiveness of the applied models utilized in this research to forecast the STS of the MSC was assessed by employing a set of nine statistical measures. These measures include the root mean square error (RMSE), mean square error (MSE), normalized root mean square error (NRMSE), root relative squared error (RRSE), relative standard error (RSE), mean absolute deviation (MAD), mean absolute percentage error (MAPE), relative root mean square error (RRMSE), and coefficient of determination (R^2) (Aretz *et al*. 2011). The formula of these loss functions are according to Eqs. (3)-(11).

R^2 assesses the proportion of variability in the dependent variable that can be explained by the independent variable(s). RMSE computes the square root of the average squared differences between predicted and actual values, giving more importance to larger errors, thus being sensitive to outliers. RRMSE expresses the variation in accuracy as a percentage, providing a relative measure of prediction accuracy by scaling RMSE with the range of actual values. MSE computes the average of the squared differences between predicted and actual values, often used in optimization and model fitting due to its simplicity in computation and interpretation. RSE is another loss

Table 5 Statistical methods employed by the developed model for estimating STS

Model	95% confidence interval for B		Correlations			Collinearity statistics	
	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
Constant	9.385	14.845					
CSC	0.210	0.286	0.019	0.461	0.363	0.111	9.648
TSC	-2.194	-1.673	-0.112	-0.452	-0.352	0.117	9.494
CA	0.004	0.004	0.501	0.514	0.358	0.903	1.546
D	-0.080	-0.033	-0.306	-0.436	-0.286	0.214	5.245
SNC	-0.021	0.031	-0.290	0.031	0.022	0.634	1.743
FMS	-0.005	0.482	0.171	0.126	0.085	0.911	1.252
W/C	-2.124	1.824	-0.452	-0.009	-0.009	0.372	3.913
SR	-0.141	-0.081	-0.217	-0.362	-0.262	0.381	2.757
S	0.001	0.003	0.083	0.082	0.055	0.882	1.356

ANOVA analysis

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	198.536	11	22.62	48.832	0.000
Residual	147.449	302	0.481		
Total	351.543	312			

function that normalizes the squared difference between predicted and actual values by the sum of squares of the actual values, facilitating model comparisons across different datasets. RRSE gauges the discrepancy between predicted and actual values by dividing RMSE by the range of observed data. MAPE quantifies the average percentage difference between predicted and actual values, particularly useful in forecasting, although it may yield undefined values when actual values include zeros. MAD calculates the average of absolute differences between predicted and actual values, offering robustness to outliers and insight into the typical error magnitude. Utilizing these diverse loss functions allows researchers and practitioners to evaluate prediction models in terms of accuracy, robustness, and applicability across various scenarios.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(P-A)^2}{n}} \tag{3}$$

$$MSE = \frac{1}{n}\sum (A-P)^2 \tag{4}$$

$$RRMSE = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(A-P)^2}{\sum_{i=1}^{n}(P)^2}} \tag{5}$$

$$NRMSE = \frac{RMSE}{average\ (P)} \tag{6}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(A-P)^2}{\sum_{i=1}^{n}(A-average\ (A))^2}} \tag{7}$$

$$RSE = \frac{\sum_{i=1}^{n}(A-P)^2}{\sum_{i=1}^{n}(A-average\ (A))^2} \tag{8}$$

$$MAD = \frac{\sum|P-average\ (P)|}{n} \tag{9}$$

$$MAPE = \frac{1}{n}\sum\left|\frac{A-P}{A}\right| \tag{10}$$

$$R^2 = 1 - \frac{sum\ squared\ regression}{sum\ of\ squares\ total} \tag{11}$$

These metrics indicate good performance when their values are low, suggesting that the model's predictions closely align with the actual values. In this case, a low value of these metrics indicates that the model is accurately capturing the variability in the data and making precise predictions. On the other hand, acceptable but suboptimal performance is indicated when these metrics have moderate values, indicating that the model's predictions are somewhat accurate but could be improved. Lastly, unacceptable performance is indicated when the values of loss functions are high, suggesting that the model's predictions are far off from the actual values, and the model is not capturing the variability in the data effectively. In this case, the model needs to be re-evaluated and potentially retrained to improve its performance (Zhou *et al*. 2021).

The results of these loss functions and statistical indexes for developed formula (Eq. 2) are presented in Table 6. As can be seen, the values of these functions show the reliable results with high accuracy and reliability. Thus, the empirical formula can be utilized to predict the STS based on defined mechanical and physical properties defined in database.

3.2 Sensitivity Analysis

In engineering, sensitivity analysis is vital for validating models and addressing uncertainties. It involves studying

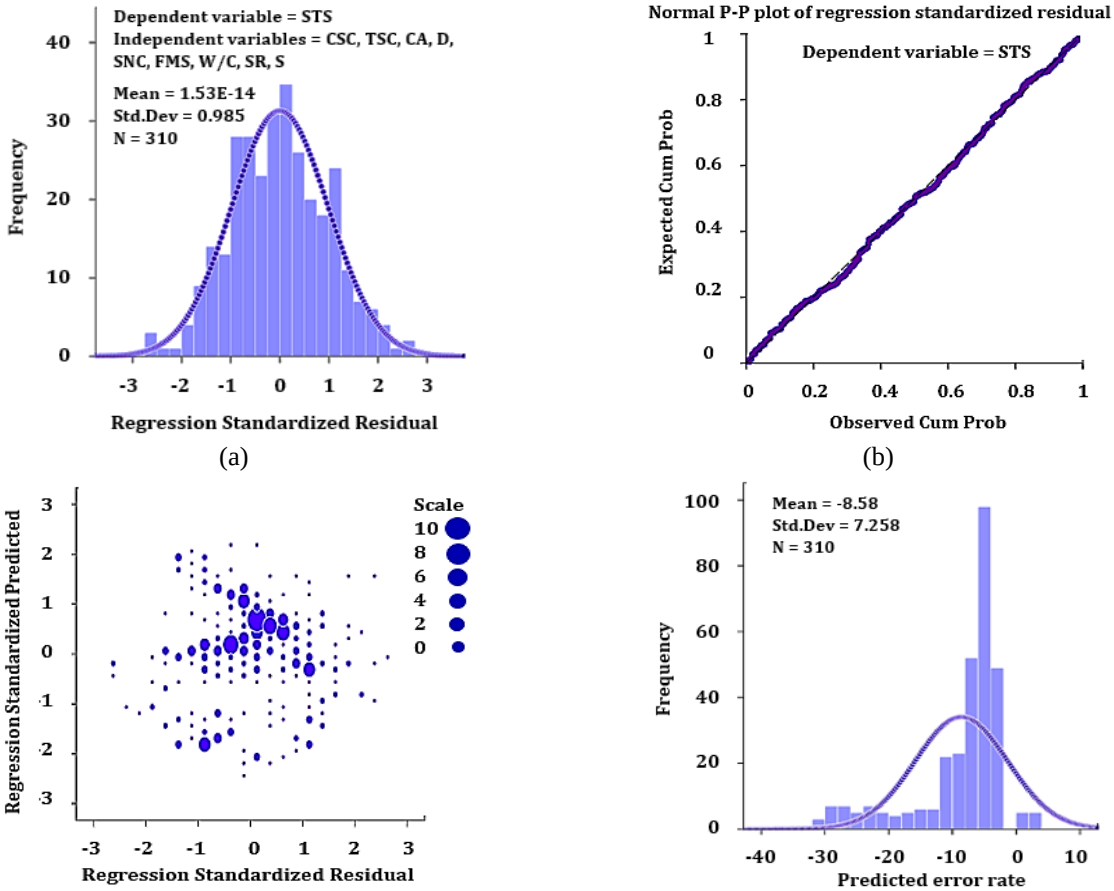

Fig. 7 Statistical diagrams for dependent parameter (STS) as output factor using forward stepwise regression analysis (FSR)

how changes in input parameters affect model outcomes, thereby grasping the influence of uncertainties on overall performance. Tornado and spider diagrams have emerged as popular tools for visualizing sensitivity analysis outcomes. Tornado diagrams succinctly depict the importance of each input parameter, ranking them based on their impact on model outputs. Conversely, spider diagrams, also called radar plots, offer a comprehensive perspective on parameter interactions. By plotting each parameter on its axis and connecting them, spider diagrams highlight areas in the parameter space affecting significant shifts in model outcomes. These visual aids enhance the clarity and understanding of sensitivity analysis, aiding informed decision-making and advancing research endeavors.

Figs. 8 and 9 depict the results of sensitivity analysis using tornado and spider diagrams for the specified input parameters. Consequently, CSC and CA significantly influence STS. Furthermore, sensitivity analysis conducted with an empirical model highlights that FMS and W/C have the least effect on STS.

Additionally, for enhanced clarity, individual diagrams depicting the influence of each input factor on STS are presented, as shown in Fig. 10. In these graphical representations, one input factor is varied while keeping the remaining input features constant, as specified in Table 7.

The ranges of parameter values necessary for conducting sensitivity analysis are illustrated in Table 8. This table presents the breadth of input variables within the database, as determined by maximum and minimum statistical measures. For instance, the minimum value for the variable CSC is recorded as zero, while the maximum value reaches 63 MPa. This range of parameters can then be categorized into sub-measures according to sequential series. To further illustrate, consider a hypothetical scenario where the input variables are being used to assess structural integrity in engineering. The range of values provides critical insights into the potential variation and limits within the data, offering valuable guidance for decision-making processes in

Table 6 Outcome of loss functions during the construction of the empirical model for STS of MSC

Loss functions	Results
RMSE	0.457
MSE	0.141
MAPE	0.012
MAD	0.109
RRMSE	0.205
NRMSE	0.010
RSE	62.461
RRSE	7.903

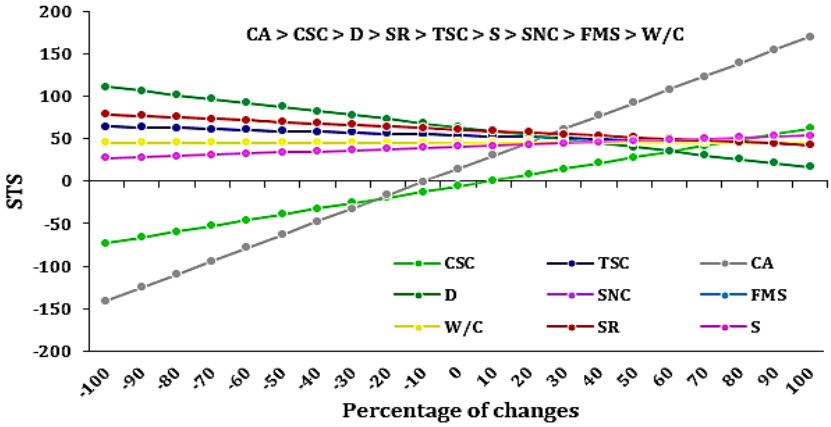

Fig 8 Spider diagram of each variables affecting STS

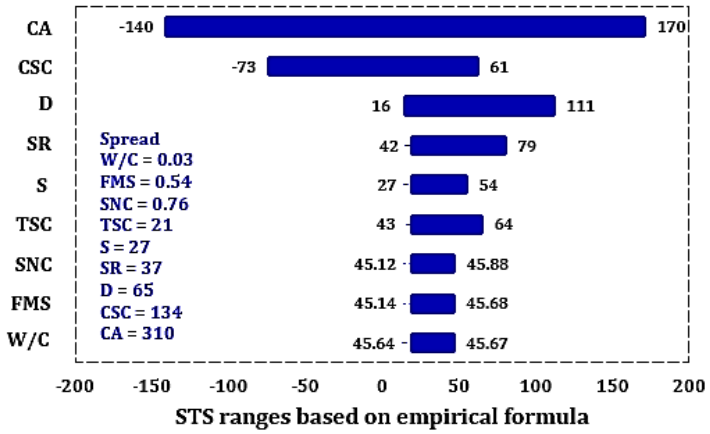

Fig. 9 Tornado graph of each variables affecting STS

Table 7 Constant values of input parameters to performing sensitivity analysis

CSC (MPa)	TSC (MPa)	CA (days)	D_{max} (mm)	SNC (%)	FMS	W/C	SR (%)	S (mm)
48	8	77	30	7.95	3	0.46	37	95

Table 8 The intervals covering the ranges of parameter values necessary for performing sensitivity analysis

Parameters	Variables
CSC (MPa)	0, 6, 12, 19, 25, 31, 38, 44, 50, 57, 63
TSC (MPa)	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
CA (day)	0, 38, 77, 116, 155, 194, 232, 271, 310, 349, 388
D_{max} (mm)	0, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80
SNC (%)	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20
FMS	0, 0.35, 0.7, 1.01, 1.4, 1.75, 2.1, 2.45, 2.8, 3.15, 3.5
W/C	0, 0.06, 0.13, 0.2, 0.27, 0.34, 0.41, 0.48, 0.55, 0.62, 0.69
SR	0, 4.4, 8.8, 13.2, 17.6, 22, 26.4, 30.8, 35.2, 39.6, 44
S	0, 26, 52, 78, 104, 130, 156, 182, 208, 234, 260

the sensitivity analysis process. Additionally, the organization of these parameters into sub-measures enables a more nuanced understanding of the data, facilitating more accurate and targeted analyses.

This method allows for assessing the impact of each factor on STS. It's important to note that changing the CSC from 0 to 60 MPa results in an increase in STS from 0 to 70 MPa. Similarly, adjusting the CA duration from 20 to 170 days leads to a variation in the STS range from 0 to 400 MPa. These results underscore the significant influence of these three parameters. However, altering the FMS range from 0-4 and W/C range from 0-0.7 does not produce

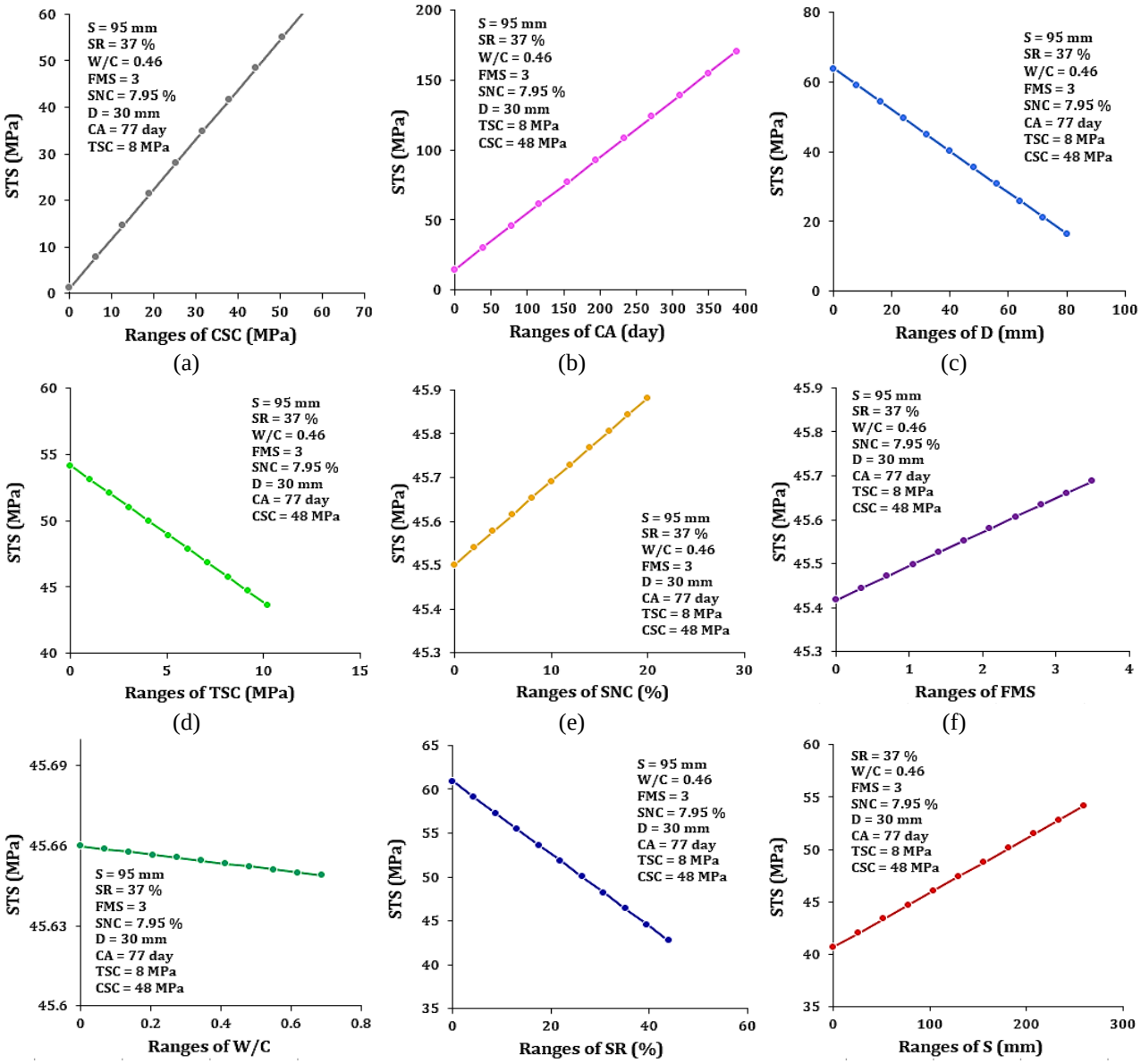

Fig. 10 Results of sensitivity analysis based on forward stepwise regression analysis (FSR) in relation to the change of input factors

substantial changes in the STS range, as it remains relatively stable.

For additional clarification and demonstration, we extended our analysis beyond single-factor sensitivity analysis to include multi-factor sensitivity analysis. In this research, we utilized the mutual information (MI) technique for sensitivity analysis. MI, a feature selection method grounded in information theory, utilizes information gain to construct decision trees. It quantifies the degree to which one variable can be deduced from observations of another. While originally intended for categorical data, MI can be adapted for numerical data. Its effectiveness can be gauged by the reduction in entropy. As shown in Eq. (12), the MI score should range from zero to infinity. A higher MI value signifies the feature's importance in model training and warrants greater attention in the final assessment. Conversely, a low MI score, such as 0, indicates minimal or no correlation between the attribute and the objective.

$$MI\,(feature, target) = Entropy\,(feature) - Entropy\,(feature \mid target) \quad (12)$$

The calculated MI scores for all inputs are depicted in Fig. 11, highlighting the link between the CA and W/C parameters and the most and least significant impacts, respectively, on the STS. The influence of other parameters on the STS properties is also noteworthy.

4. Developing predictive networks using supervised algorithms

In geotechnical projects characterized by intricate mechanical and physical conditions and a multitude of input variables, supervised learning methods like ML, deep

Table 9 The outcomes of loss functions, encompassing both the training and testing datasets

	ELM	RBF	GEP	ANN 7	ANN 13	ANN 17	ANN 22	DFF	TSF	MFIS	SVM
RMSE	0.036	0.029	0.033	0.041	0.004	0.018	0.029	0.025	0.074	0.056	0.088
MSE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003
MAPE	0.001	0.009	0.001	0.008	0.007	0.005	0.004	0.001	0.001	0.001	0.002
MAD	0.005	0.006	0.005	0.005	0.007	0.007	0.006	0.006	0.004	0.004	0.003
RRMSE	0.009	0.021	0.014	0.013	0.003	0.030	0.004	0.012	0.036	0.032	0.048
NRMSE	0.006	0.004	0.005	0.007	0.000	0.003	0.004	0.004	0.011	0.009	0.014
RSE	0.114	0.007	0.046	0.015	0.000	0.002	0.077	0.005	0.000	0.002	0.034
RRSE	0.337	0.081	0.215	0.121	0.004	0.047	0.277	0.070	0.010	0.044	0.185
R^2	0.856	0.932	0.916	0.887	0.959	0.894	0.935	0.941	0.874	0.823	0.582

Table 10 The outcomes of loss functions, encompassing the training datasets

	ELM	RBF	GEP	ANN 7	ANN 13	ANN 17	ANN 22	DFF	TSF	MFIS	SVM
RMSE	0.038	0.030	0.035	0.044	0.004	0.019	0.031	0.027	0.078	0.059	0.093
MSE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003
MAPE	0.001	0.007	0.001	0.008	0.009	0.005	0.004	0.001	0.002	0.001	0.002
MAD	0.006	0.006	0.006	0.005	0.008	0.007	0.006	0.007	0.004	0.005	0.003
RRMSE	0.010	0.022	0.015	0.013	0.003	0.032	0.004	0.012	0.038	0.033	0.051
NRMSE	0.006	0.004	0.005	0.007	0.001	0.003	0.005	0.004	0.011	0.009	0.015
RSE	0.112	0.006	0.046	0.014	0.000	0.002	0.076	0.005	0.000	0.002	0.034
RRSE	0.335	0.080	0.214	0.120	0.004	0.047	0.275	0.069	0.010	0.043	0.184
R^2	0.850	0.930	0.920	0.890	0.970	0.900	0.949	0.947	0.880	0.820	0.590

Table 11 The outcomes of loss functions, encompassing both the testing datasets

	ELM	RBF	GEP	ANN 7	ANN 13	ANN 17	ANN 22	DFF	TSF	MFIS	SVM
RMSE	0.152	0.085	0.108	0.068	0.011	0.016	0.105	0.022	0.079	0.121	0.056
MSE	0.004	0.003	0.019	0.005	0.016	0.034	0.001	0.009	0.037	0.019	0.084
MAPE	0.001	0.009	0.002	0.008	0.008	0.007	0.001	0.002	0.004	0.001	0.003
MAD	0.009	0.006	0.015	0.002	0.016	0.013	0.007	0.007	0.004	0.001	0.005
RRMSE	0.058	0.022	0.020	0.014	0.004	0.040	0.007	0.055	0.010	0.038	0.005
NRMSE	0.026	0.015	0.020	0.012	0.002	0.003	0.017	0.004	0.014	0.020	0.009
RSE	0.204	0.173	0.266	0.998	0.107	2.047	0.209	0.329	0.156	0.207	0.248
RRSE	0.452	0.416	0.515	0.999	0.328	1.431	0.457	0.574	0.395	0.455	0.498
R^2	0.840	0.860	0.859	0.830	0.853	0.800	0.780	0.880	0.833	0.770	0.410

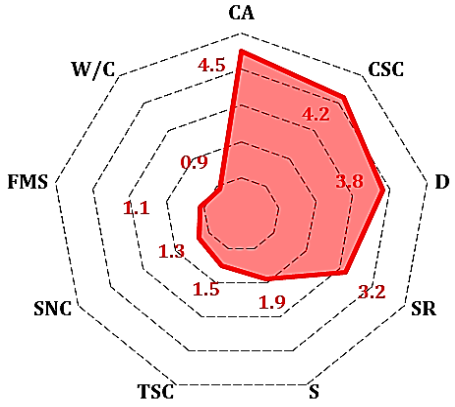

Fig. 11 Multi-factor sensitivity score of inputs

learning, genetic approaches, neural networks, and fuzzy-based models have gained significant traction. The integration of these supervised learning algorithms has demonstrated notable effectiveness in improving the precision and efficacy of models used to estimate STS of MSC in geotechnical applications.

This section presents a concise overview of several frequently utilized algorithms for STS estimation, including SVM, TSF, GEP, ELM, MFIS, ANN, DFF, and RBF.

The decision to analyze these algorithms is supported by various factors:

• They present innovative solutions to complex issues by employing novel approaches to encoding and decoding data.

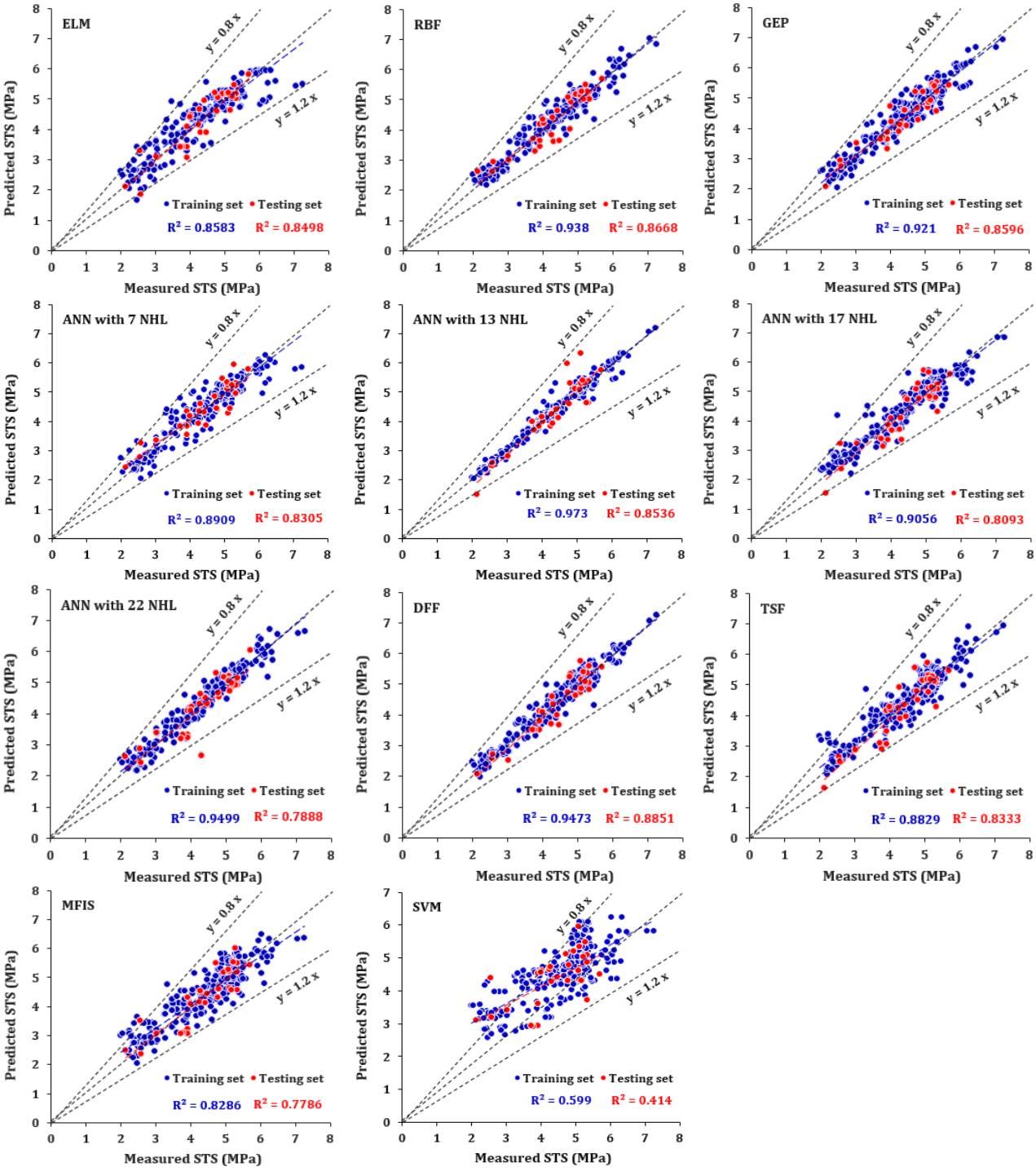

Fig. 12 A comparison between the predictions generated by supervised learning and the actual monitored data

• Their integration of a regularization parameter aids in averting overfitting, managing the complexity of the model, and preventing it from capturing noise in the dataset.

• These algorithms demonstrate proficiency in handling large datasets, utilizing kernel and non-linear techniques to transform data into higher-dimensional spaces. This adaptability facilitates the effective management of datasets that are not linearly separable.

• The use of bilingual notation in these models offers an effective strategy for addressing intricate problems. Additionally, their reputation for quick and efficient learning is particularly valuable in geotechnical applications, where timely predictions are crucial for decision-making.

• They demonstrate swift convergence, necessitating fewer iterations compared to alternative machine learning algorithms to achieve the desired level of accuracy.

• These algorithms are supported by a robust theoretical foundation that ensures convergence and optimality under specific conditions.

• They demonstrate versatility in various learning

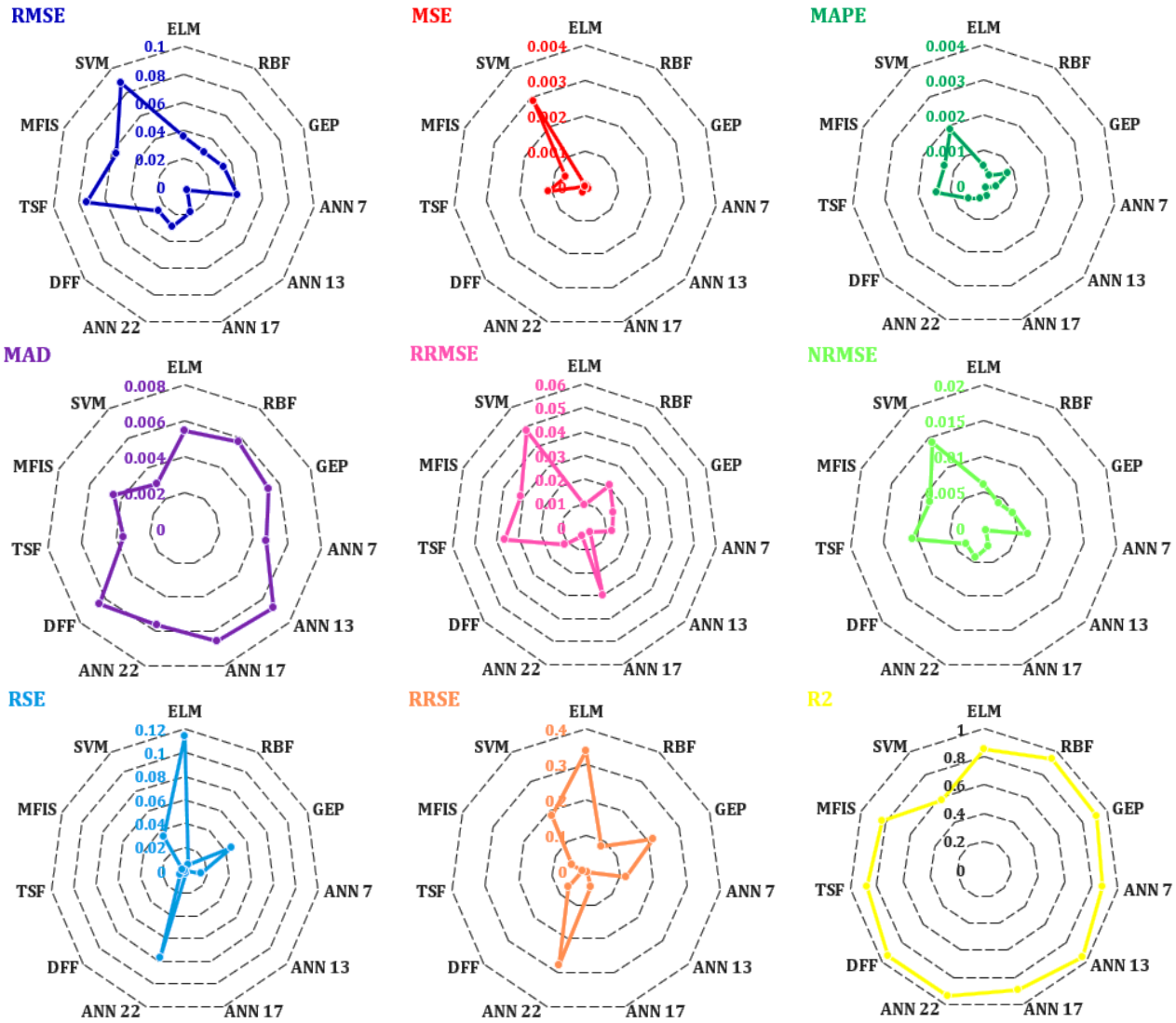

Fig. 13 Performance comparison of the supervised learning models in terms of nine types of loss functions for evaluation the results of STS

scenarios, including batch learning, sequential learning, and incremental learning, making them adaptable tools for predicting target parameters in diverse geotechnical settings and experimental setups.

• By considering the relationship between nodes over a time sequence, these models can capture temporal dependencies and make accurate predictions based on historical context.

• They serve as valuable tools in various applications, particularly in time series forecasting.

To evaluate the performance of the developed models, various loss functions and statistical indices have been proposed. These metrics serve the purpose of quantifying the disparity between the predicted and actual STS values, as well as assessing the predictive capability of the algorithms.

The current study utilized Python version 3 and MATLAB R2020a software for implementing and running ML models. The main aim was to improve the ease of managing and executing packages. Computation was performed on a computer equipped with 8GB of RAM and an 11th Generation Intel(R) Core(TM) i5-1155G7 processor running at 2.50 gigahertz. Optimization of all ML models was conducted using their respective training datasets to ensure optimal performance in predicting STS. This optimization process involved setting the hyperparameters of each model to their optimal values or states. To accomplish this, a trial-and-error method was adopted to fine-tune the hyperparameters of the models. Problem-solving often requires repeated experimentation and adjustments until a successful solution is found. Therefore, the agent in this study underwent numerous iterations before a viable solution was identified.

5. Results and discussion

The highly accurate prediction capabilities of the ML models in determining the STS of MSC have been demonstrated in the present study. The effectiveness of the

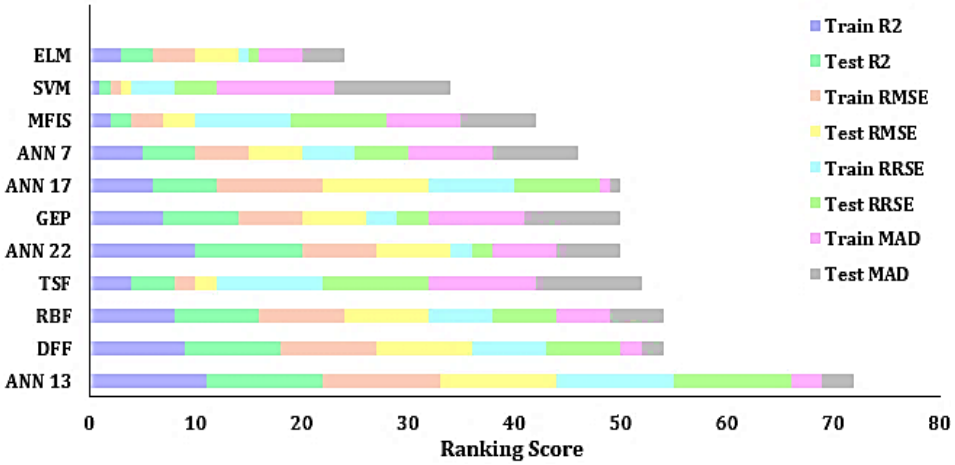

Fig. 14 Intuitive display of comprehensive ranking of eleven applied models

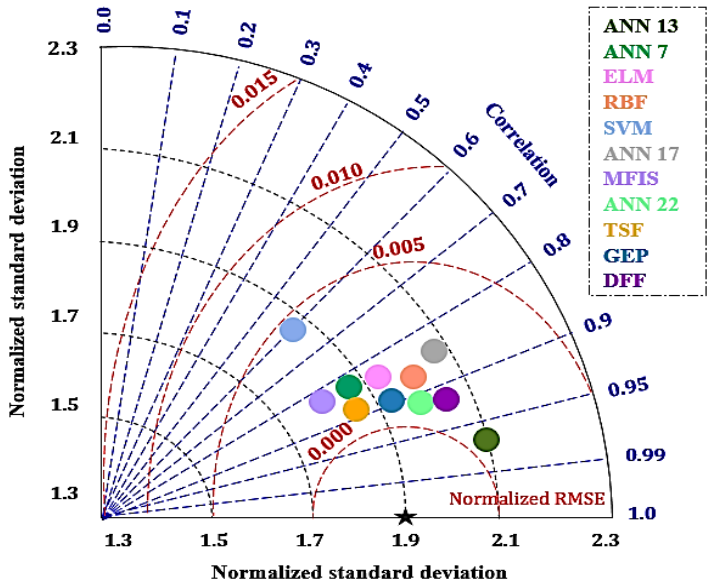

Fig. 15 Taylor chart for comparison of multiple supervised models to predict STS

ML models is evident from the strong agreement between the actual and predicted values of STS, as depicted in Fig. 12. As in Fig. 12, the ANN model with 13 NHL has provided the most accurate predictions in the training phase with a R^2 of 0.9730. In the testing phase, the DFF model has provided the most accurate predictions (R^2 = 0.8851). For the both training and testing phases, the main focus of the comparison points is on x = y line and most of them are located between x = 1.2y and x = 0.8y lines. According to these charts, it can be said that the lowest accuracy is provided by the SVM model.

To comprehensively assess the performance prediction of the ML models, several loss functions including RMSE, RRMSE, NRMSE, MSE, MAPE, and MAD were utilized. These loss functions quantify the deviation between the predicted outcomes and the actual measured values, providing a quantitative evaluation of the predictive capabilities of the ML models. The results of this evaluation can be observed in Tables 9-11.

The use of loss functions in assessing the accuracy of ML models is crucial in determining the reliability of the predictive models. By quantifying the deviation between predicted and actual values, these loss functions highlight any discrepancies and provide a metric to measure the performance of the ML models.

Tables 9-11 present the results of the evaluation, demonstrating the accuracy of the ML models in predicting the STS of MSC. The low values of the loss functions indicate a minimal deviation between the predicted and actual values, further affirming the effectiveness of the ML models.

Fig. 13 provides a visual comparison of different types of loss functions employed for predicting the STS. This comparative analysis highlights the varied performance of each loss function, with some demonstrating superior predictive abilities compared to others. The ML models, overall, exhibit superior predictive abilities, suggesting their potential to estimate the STS of MSC.

The inclusion of loss functions and their evaluation further enhances the accuracy of the ML models. By

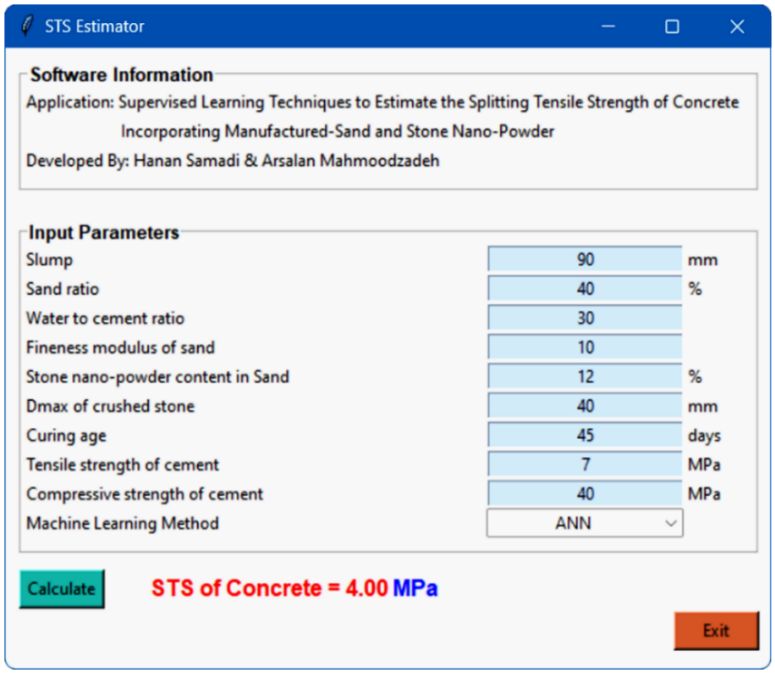

Fig. 16 GUI of the supervised learning models for estimating the STS of SNP-reinforced MSC

quantifying the deviation between predicted and actual values, loss functions help refine the predictions and improve the reliability of the models. This enhanced accuracy makes the ML models highly reliable in predicting future STS values.

The overall ranking results of the different ML models are effectively visualized in Fig. 14. Upon closer examination, it becomes evident that the ANN with 13 NHL (RMSE = 0.011, RRMSE = 0.004) and the DFF (RMSE = 0.022, MAD = 0.007) models emerge as the most robust and accurate models. This conclusion is drawn by considering both the training and testing comprehensive ranking. These models not only outperform other ML models in terms of accuracy but also exhibit a high level of reliability, further validating their predictive capabilities.

To comprehensively evaluate the prediction results of various soft computing techniques for STS output parameters, a normalized error analysis was conducted. Fig. 15 presents a Taylor diagram that provides a detailed assessment of the performance of the applied algorithms. This assessment is based on normalized standard deviation and NRMSE. The consistently excellent results showcased in the diagram reflect the superior performance of the ML models, particularly in the testing subsets. However, the ANN with 13 NHL has provided the most accuracy.

By consistently delivering excellent results in predicting the STS values based on input data derived from various input features, the ML models demonstrate their effectiveness and accuracy in predicting the behavior and performance of MSC structures.

This opens up new possibilities for the optimization of design and construction processes for concrete structures. Instead of relying solely on traditional engineering methods, engineers can now leverage ML models to gain valuable insights into how these structures will perform. This insight can help them make informed decisions, identify potential weaknesses or areas of concern, and take proactive measures to ensure the long-term durability and stability of the structures.

Furthermore, ML techniques offer the advantage of being able to analyze large amounts of data quickly and efficiently. This enables engineers to process vast datasets that would otherwise be time-consuming and challenging to analyze manually. By leveraging ML models, engineers can gain a comprehensive understanding of MSC structures and their behavior, allowing them to optimize their designs and make more accurate predictions.

6. Graphical user interphase

Estimating the compressive strength of concrete can be seamlessly accomplished through user-friendly graphical user interfaces (GUIs), serving as robust platforms housing machine learning-based prediction models. This study integrates all supervised learning models within the system. There's no necessity to retrain the supervised models for compressive strength estimation as they have already undergone training using the laboratory dataset. The GUI, illustrated in Fig. 16, allows users to input parameters, enabling the trained machine learning algorithms to generate predictions for the compressive strength output. Notably, the GUI exhibits superior accuracy in predicting the compressive strength of MSC compared to laboratory tests, achieving this with less time and reduced costs. Furthermore, the GUI can act as a valuable tool for researchers interested in exploring the science, technology, and societal implications of SNP-reinforced MSC. It also holds promise as a platform for gathering data on the compressive strength of SNP-reinforced MSC.

7. Conclusions

Findings from this investigation lead to the following conclusions:

• Supervised learning models demonstrated close approximations to experimental results. While all models showed positive outcomes, the ANN model with 13 neurons in the hidden layer and the DFF model showcased superior predictive accuracy. The ANN model achieved an R2 exceeding 0.90, an RMSE of 0.004, and an MAPE of 0.001. Similarly, the DFF model exhibited impressive metrics, with an R2 above 0.90, an RMSE of 0.025, and an MAPE of 0.003.

• Pearson correlation analysis, sensitivity analysis, and statistical analysis were employed to evaluate each parameter's contribution to the prediction challenge. Results emphasized the significant impact of each characteristic on STS, with CA identified as the most influential parameter and W/C as the least significant.

• To enhance STS estimation for engineering purposes, a GUI was developed for the supervised learning models. This GUI offers various supervised learning methods to estimate the STS of different MSC samples quickly.

• Future research utilizing the models in this study should explore the most precise algorithms and optimal parameters for predicting concrete's STS. Moreover, considering the multitude of challenges in this field, it's crucial to evaluate and explore the models' predictive capabilities in geotechnical problems. Additionally, future studies should consider the type of nanoparticles utilized in the MSC mixture.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University (KKU) for funding this work through the Research Group Program Under the Grant. code RGP2/262/44.

References

Chen, T. and Guestrin, C. (2016), "XGBoost", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

Chou, J.S. and Pham, A.D. (2013), "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength", *Constr. Build. Mater.*, **49**, 554-563. https://doi.org/10.1016/j.conbuildmat.2013.08.078

Cortes, C. and Vapnik, V. (1995), "Support-vector networks", *Mach. Learn.*, **20**(3), 273-297. https://doi.org/10.1007/BF00994018

Cover, T. and Hart, P. (1967), "Nearest neighbor pattern classification", *IEEE T. Inform. Theor.*, **13**(1), 21-27. https://doi.org/10.1109/TIT.1967.1053964

Dehestani, A., Kazemi, F., Abdi, R. and Nitka, M. (2022), "Prediction of fracture toughness in fibre-reinforced concrete, mortar, and rocks using various machine learning techniques. *Eng. Fract. Mech.*, **276**, 108914. https://doi.org/10.1016/j.engfracmech.2022.108914

Ding, X., Li, C., Xu, Y., Li, F. and Zhao, S. (2016), "Experimental study on long-term compressive strength of concrete with manufactured sand", *Constr. Build. Mater.*, **108**, 67-73. https://doi.org/10.1016/j.conbuildmat.2016.01.028

Ho, T.K. (1998), "The random subspace method for constructing decision forests", *IEEE T. Pattern Anal. Mach. Intell.*, **20**(8), 832-844. https://doi.org/10.1109/34.709601

Hou, S., Qiao, L. and Xing, L. (2022), "Artificial neural fuzzy system and monitoring the process via IoT for optimization synthesis of nano-size polymeric chains", *Adv. Nano Res.*, **14**(4), 375-386. https://doi.org/https://doi.org/10.12989/anr.2022.12.4.375

Jin, R. and Agrawal, G. (2003), "Communication and Memory Efficient Parallel Decision Tree Construction", *Proceedings of the 2003 SIAM International Conference on Data Mining*, 119-129. https://doi.org/10.1137/1.9781611972733.11

Kaloop, M. R., Kumar, D., Samui, P., Hu, J. W. and Kim, D. (2020), "Compressive strength prediction of high-performance concrete using gradient tree boosting machine", *Constr. Build. Mater.*, **264**, 120198. https://doi.org/10.1016/j.conbuildmat.2020.120198

Kazemi, F., Asgarkhani, N. and Jankowski, R. (2023), "Machine learning-based seismic response and performance assessment of reinforced concrete buildings", *Arch. Civil Mech. Eng.*, **23**(2), 94. https://doi.org/10.1007/s43452-023-00631-9

Aretz, K., Bartram, S. M. and Pope, P. F. (2011), "Asymmetric loss functions and the rationality of expected stock returns", *Int. J. Forecast.*, **27**(2), 413-437.

Li, B., Ke, G. and Zhou, M. (2011), "Influence of manufactured sand characteristics on strength and abrasion resistance of pavement cement concrete", *Constr. Build. Mater.*, **25**(10), 3849-3853. https://doi.org/10.1016/j.conbuildmat.2011.04.004

Ling, H., Qian, C., Kang, W., Liang, C. and Chen, H. (2019), "Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment", *Constr. Build. Mater.*, **206**, 355-363. https://doi.org/10.1016/j.conbuildmat.2019.02.071

Liu, Q., Peng, K., Zeng, J., Marzouki, R., Majdi, A., Jan, A., Salameh, A. A. and Assilzadeh, H. (2022), "Effects of mining activities on Nano-soil management using artificial intelligence models of ANN and ELM", *Adv. Nano Res.*, **12**(6), 549-566. https://doi.org/https://doi.org/10.12989/anr.2022.12.6.549

Mamdani, M. (1974), "The ideology of population control. *Concern. Demograph.*, **4**(2), 13-22. PMID: 12307029.

Moayedi, Kalantar, Foong, Tien Bui and Motevalli. (2019), "Application of three metaheuristic techniques in simulation of concrete slump", *Appl. Sci.*, **9**(20), 4340. https://doi.org/10.3390/app9204340

Pan, J., Zhong, W., Wang, J. and Zhang, C. (2022), "Size effect on dynamic splitting tensile strength of concrete: Mesoscale modeling", *Cement Concr. Compos.*, **128**, 104435. https://doi.org/10.1016/j.cemconcomp.2022.104435

Quinlan, J.R. (1986), "Induction of decision trees", *Mach. Learn.*, **1**(1), 81-106. https://doi.org/10.1007/BF00116251

Rasmussen, C.E. (2004), "Gaussian processes in machine learning", *Adv. Lecture. Mach. Learn.*, 63-71. https://doi.org/10.1007/978-3-540-28650-9_4

Ren, W., Wu, X. and Cai, R. (2022), "A hybrid artificial intelligence and IOT for investigation dynamic modeling of nano-system", *Adv. Nano Res.*, **13**(2), 165-174. https://doi.org/https://doi.org/10.12989/anr.2022.13.2.165

Tinoco, J., Alberto, A., da Venda, P., Gomes Correia, A. and Lemos, L. (2020), "A novel approach based on soft computing techniques for unconfined compression strength prediction of soil cement mixtures", *Neural Comput. Appl.*, **32**(13), 8985-8991. https://doi.org/10.1007/s00521-019-04399-z

Walczak, S. and Cerpa, N. (2003), "Artificial neural networks",

Encyclopedia Phys. Sci. Technol., **65**(6), 631-645. https://doi.org/10.1016/B0-12-227410-5/00837-1

Yaseen, Z.M., Deo, R C., Hilal, A., Abd, A.M., Bueno, L.C., Salcedo-Sanz, S. and Nehdi, M.L. (2018), "Predicting compressive strength of lightweight foamed concrete using extreme learning machine model", *Adv. Eng. Softw.*, **115**, 112-125. https://doi.org/10.1016/j.advengsoft.2017.09.004

Zhang, J., Li, D. and Wang, Y. (2020), "Predicting mechanical properties of manufactured-sand concrete using tree-based machine learning models", *J. Clean. Prod.*, **258**, 120665. https://doi.org/10.1016/j.jclepro.2020.120665

Zhao, S., Ding, X., Zhao, M., Li, C. and Pei, S. (2017), "Experimental study on tensile strength development of concrete with manufactured sand", *Constr. Build. Mater.*, **138**, 247-253. https://doi.org/10.1016/j.conbuildmat.2017.01.093

Zhou, X., Wang, P., Al-Dhaifallah, M., Rawa, M. and Khadimallah, M. A. (2022), "A machine learning-based model for the estimation of the critical thermo-electrical responses of the sandwich structure with magneto-electro-elastic face sheet", *Adv. Nano Res.*, **12**(1), 81-99. https://doi.org/https://doi.org/10.12989/anr.2022.12.1.081

Zhou, X., Liu, X., Jiang, J., Gao, X. and Ji, X. (2021, July), "Asymmetric loss functions for learning with noisy labels", *Proceedings of the International Conference on Machine Learning*, 12846-12856. PMLR. https://doi.org/10.1016/j.ijforecast.2009.10.008

CC