https://doi.org/10.31436/iiumej.v26i3.3730

MACHINE LEARNING AND RSM FOR STRENGTH FORECASTING IN SUSTAINABLE SCGC

SAMEH FUQAHA[1], AHMAD ZAKI[2*], GUNTUR NUGROHO[2]

[1]*Dept. of Civil Engineering, Postgraduate Studies, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia*
[2]*Dept. of Civil Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia*

**Corresponding author: ahmad.zaki@umy.ac.id*

(Received: 17 May 2025; Accepted: 3 July 2025; Published online: 9 September 2025)

ABSTRACT: This research focuses on the predictive modeling of flexural (Ff) and splitting tensile (Ft) strengths in Self-Compacting Geopolymer Concrete (SCGC) to support sustainable mix design optimization. A curated dataset comprising 544 experimental records was utilized to train and evaluate eight supervised machine learning (ML) algorithms. These included Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests, Gradient Boosting, CN2 Rule Induction, Naïve Bayes, Decision Trees, and Stochastic Gradient Descent. The predictive performance of each model was assessed using multiple statistical metrics, such as RMSE, R^2, and accuracy percentage. Among the models, SVM and KNN achieved the highest precision, with R^2 values of 0.99 and RMSE as low as 0.10 MPa. Additionally, statistical techniques were applied to identify influential input variables, confirming the dominant role of binder constituents in determining tensile-related strength. The models demonstrated strong generalization on unseen data and minimal sensitivity to activator dosage or curing age. These results validate the effectiveness of ML-driven tools for SCGC prediction and offer a scalable framework for integrating data analytics into sustainable concrete design and performance optimization.

ABSTRAK: Kajian ini memfokuskan kepada pemodelan ramalan bagi kekuatan lenturan (Ff) dan tegangan belahan (Ft) dalam Konkrit Geopolimer Pemadat Kendiri (SCGC) bagi menyokong pengoptimuman reka bentuk campuran mampan. Satu set data terpilih yang merangkumi 544 rekod eksperimen telah digunakan bagi melatih dan menilai lapan algoritma pembelajaran mesin (ML) terselia. Algoritma tersebut termasuk Mesin Sokongan Vektor (SVM), K-Nearest Neighbors (KNN), Rawak Forests, Gradient Boosting, CN2 Rule Induction, Naïve Bayes, Pokok Keputusan, dan Stochastic Gradient Descent. Prestasi ramalan setiap model dinilai menggunakan pelbagai metrik statistik seperti RMSE, R^2, dan peratusan ketepatan. Antara model tersebut, SVM dan KNN mencapai ketepatan tertinggi dengan nilai R^2 sebanyak 0.99 dan RMSE serendah 0.10 MPa. Tambahan, teknik statistik turut digunakan bagi mengenal pasti pemboleh ubah input berpengaruh, sekali gus mengesahkan peranan dominan konstituen pengikat dalam menentukan kekuatan berkaitan tegangan. Model yang dibangunkan menunjukkan keupayaan generalisasi yang kukuh terhadap data baharu serta kepekaan minimum terhadap dos pengaktif atau umur pengerasan. Dapatan ini mengesahkan keberkesanan alat berasaskan ML bagi meramal SCGC dan menawarkan kerangka boleh skala bagi mengintegrasikan analitik data ke dalam reka bentuk konkrit mampan serta pengoptimuman prestasi.

KEY WORDS: *Self-Compacting Geopolymer Concrete (SCGC); Flexural Strength Prediction; Splitting Tensile Strength; Machine Learning Models; Response Surface Methodology (RSM)*

https://doi.org/10.31436/iiumej.v26i3.3730

1. INTRODUCTION

Self-Compacting Geopolymer Concrete (SCGC) is an innovative material that merges the high flowability and workability of self-compacting concrete (SCC) with the eco-friendly properties of geopolymer technology. This integration provides multiple benefits, especially by enhancing eco-friendly characteristics and improving structural performance [1]. Unlike conventional Portland cement systems, using alkaline activators, geopolymer binders are formulated by activating aluminosilicate-rich industrial residues, such as fly ash, ground granulated blast furnace slag (GGBS), and metakaolin. This approach significantly curtails the greenhouse gas emissions typically linked to traditional cement manufacturing [2]. The substitution of Portland cement with pozzolanic materials has been shown to lower carbon dioxide emissions and energy consumption throughout the concrete's life cycle [3].

SCGC possesses the unique ability to flow and fill formworks solely under the influence of gravity, removing the necessity for external vibration. This property proves highly advantageous in casting intricate geometries and heavily congested reinforcement areas. Beyond improving ease of placement and construction efficiency, SCGC exhibits remarkable performance in terms of strength, durability, and resistance to elevated temperatures, often matching or exceeding the qualities observed in traditional self-compacting concrete [4]. Furthermore, its excellent chemical stability and minimal shrinkage render it highly appropriate for demanding applications, including coastal infrastructure, industrial facilities, and thermally stressed environments [5].

Despite the promising performance of SCGC, optimizing its mix design to achieve desired fresh and hardened properties remains complex due to the nonlinear interaction among multiple influencing variables [6]. Therefore, modern machine learning (ML) techniques have been increasingly utilized to develop predictive models capable of estimating key mechanical properties. Previous studies have successfully implemented ML models such as Random Forest, XGBoost, Multi-Expression Programming, Genetic Programming, and Artificial Neural Networks to forecast the strength of SCC and geopolymer concretes [7–10].

In recent advancements related to alkali-activated concrete systems, numerous studies have explored the integration of ML and response surface methodology (RSM) for predictive modeling and process optimization. Dong et al. [11] introduced a novel framework utilizing extended support vector regression (X-SVR) to estimate chloride-induced corrosion in alkali-activated slag (AAS) concrete. By combining multi-ionic transport simulations with hydrate-based models of chloride binding, the study offered a robust tool for evaluating corrosion risk with enhanced computational efficiency. Complementing this, Iqbal et al. [12] investigated the rheological behavior of alkali-activated materials (AAMs) incorporating hydrated lime. Various ML algorithms were deployed to predict key properties such as plastic viscosity and yield stress, including stochastic gradient descent, K-nearest neighbors (KNN), and support vector machines (SVM). The KNN model demonstrated the highest predictive accuracy, and RSM was further applied to analyze the influence of mix constituents, revealing that hydrated lime increased viscosity. At the same time, fly ash contributed to reducing the yield stress.

Building on these methodologies, Parhi et al. [13] developed a hybrid ML model that merged three base algorithms with a meta-learning layer, surpassing conventional approaches in terms of performance. In this case, RSM served as an optimization tool to fine-tune variables such as curing temperature, alkaline activator concentration (Na_2SiO_3/NaOH ratio), and superplasticizer dosage, achieving compressive strengths within the range of 39–43 MPa. While these approaches have proven effective, much of the existing research focuses predominantly on compressive strength, often neglecting other essential mechanical properties

[14]. Moreover, a limited comparative evaluation of different ML techniques for diverse performance metrics is crucial for selecting the most suitable model for a given application. Jiang et al [15] applied ensemble learning to predict the mechanical behavior of SCGC containing calcium carbide waste and rice husk ash, highlighting the improved generalization ability of ensemble models over standalone algorithms.

Addressing existing research gaps, this work develops and analyzes combined machine learning approaches to forecast the flexural and splitting tensile strengths of SCGC. A diverse range of ML algorithms is investigated, including ensemble techniques such as boosting and bagging, SVM, KNN, and rule-based classifiers, to identify models that most effectively capture the complex tensile behavior of SCGC. The models are trained on an experimentally validated dataset incorporating mix designs with sustainable binders such as fly ash and ground granulated blast furnace slag (GGBS), ensuring relevance to real-world construction practices. Beyond benchmarking predictive performance, the study emphasizes model interpretability by integrating RSM and sensitivity analysis using the Hoffman & Gardner approach. These methods provide both high-fidelity predictions and clear insights into the influence of key input parameters. Ultimately, by developing interpretable and reliable ML tools, this research supports the practical integration of artificial intelligence in concrete technology and contributes to scalable, environmentally responsible engineering solutions.

2. METHODOLOGY

2.1. Data Compilation and Statistical Summary

In this study, a comprehensive dataset was constructed to facilitate the predictive modeling of SCGC's mechanical properties. A total of 544 experimental records were compiled from peer-reviewed journal articles and technical sources, as in Table 1. These entries represent various mixed designs and curing ages, covering diverse proportions of geopolymer constituents and mechanical performance results.

Table 1. Compiled Experimental Records from Prior Literature Sources

References	Data	References	Data
Mazumder et al. [16]	132	Aliabdo et al. [17]	39
Ngernkham et al. [18]	42	Arun et al. [19]	12
Lavanya et al. [20]	51	Deilami et al. [21]	36
Rangan et al. [22]	52	Nishanth et al. [23]	6
Vijaya et al. [24]	32	Alam et al. [25]	108
Tho-In et al. [26]	9	Bheel et al. [27]	85

Each record in the dataset encapsulates a comprehensive set of input and output variables crucial for ML analysis in the context of geopolymer concrete performance prediction. The input variables include the dosage of GGBS, measured in kilograms per cubic meter (kg/m³), which serves as a primary binder component, and FA, also in kg/m³, commonly used to enhance workability and long-term strength. Additionally, the alkaline activators, represented by NaOH and Na_2SiO_3, both quantified in kg/m³, play a key role in initiating the geopolymerization process. The curing age (Age), expressed in days, reflects the time at which strength tests are conducted, capturing the time-dependent evolution of mechanical properties.

The output variables focus on mechanical performance, specifically the Ff and Ft, expressed in megapascals (MPa). Ff provides insights into the material's resistance to bending under load, while Ft indirectly evaluates its capacity to withstand tensile stresses. These output

parameters are vital for understanding structural behavior and for training ML models to predict the mechanical performance of geopolymer concrete based on its compositional and processing variables.

These variables were selected based on their relevance to geopolymerization mechanisms and their influence on the strength development of SCGC [28]. The compiled data represents various curing conditions, mix ratios, and testing procedures, ensuring diversity and representativeness across the parameter space.

2.1.1. Dataset Preparation and Model Structuring

The dataset was randomly divided into two primary subsets: a training set and a validation set, to promote the development of reliable and unbiased ML models. The training portion, comprising 408 data points (representing 75% of the total dataset), was used to fit the models by enabling them to learn the complex relationships between input features and target outputs. The remaining 25% validation set, comprising 136 samples, evaluated how well the trained models could predict outcomes on previously unseen data, testing their generalizability. This random allocation helps reduce the likelihood of overfitting, ensuring the models remain effective in real-world applications. To further reinforce the model's stability and reliability, 10-fold cross-validation was implemented during the training phase. This method offers additional protection against overfitting, particularly valuable when dealing with multiple output variables and inherent non-linear interactions among the predictors [29].

2.1.2. Statistical Analysis and Correlation Mapping

Initial data exploration involved computing key statistical descriptors for all input features and target outputs, including mean, standard deviation, minimum, maximum, and variance. These metrics were evaluated separately for the training set (408 samples) and the testing set (136 samples) to ensure consistent data distribution across both subsets. The summarized results are presented in Tables 2 and 3. This analysis was crucial for assessing variability, detecting potential outliers, and evaluating the overall range of values that could impact model performance. The training set demonstrated sufficient variability across all features, with GGBS and FA showing high standard deviations (156.8 and 156.6 kg/m³, respectively). At the same time, Ff and Ft had standard deviations of 1.6 MPa and 1.1 MPa, respectively. Similar distributions were observed in the testing set, confirming that the data partitioning preserved the underlying patterns and feature dispersion necessary for reliable model training and validation.

Table 2. Statistical Summary of Training Set (n = 408)

Statistic	GGBS (kg/m³)	FA (kg/m³)	NaOH (kg/m³)	Na_2SiO_3 (kg/m³)	Age (days)	Ff (MPa)	Ft (MPa)
Max	525.6	515.5	95.2	133.6	28.0	7.1	5.2
Min	0.0	0.0	52.7	75.0	7.0	0.2	0.1
Avg	247.3	237.9	68.1	101.8	16.0	3.2	2.2
SD	156.8	156.6	10.6	10.7	8.6	1.6	1.1
Var	0.6	0.6	0.2	0.1	0.5	0.5	0.5

https://doi.org/10.31436/iiumej.v26i3.3730

Table 3. Statistical Summary of Testing Set (n = 136)

Statistic	GGBS (kg/m³)	FA (kg/m³)	NaOH (kg/m³)	Na_2SiO_3 (kg/m³)	Age (days)	Ff (MPa)	Ft (MPa)
Max	512.5	510.2	93.6	132.0	28.0	6.9	5.1
Min	0.0	0.0	53.0	75.1	7.0	0.3	0.2
Avg	245.9	238.8	68.6	101.5	15.8	3.0	2.1
SD	157.3	157.7	11.1	10.9	8.9	1.8	1.1
Var	0.6	0.7	0.2	0.1	0.5	0.5	0.5

Furthermore, a Pearson correlation analysis was performed on the entire dataset comprising 544 samples to assess the linear relationships between the input variables (GGBS, FA, NaOH, Na_2SiO_3, and Age) and the target mechanical properties Ff and Ft. As illustrated in Figure 1, GGBS demonstrates a strong positive correlation with both Ff (0.77) and Ft (0.75), Indicating that higher levels of GGBS content are strongly associated with improvements in these strength characteristics. In contrast, FA, NaOH, and Na_2SiO_3 exhibit very weak correlations (near zero) with the output variables, indicating minimal direct linear influence on strength. Notably, the curing Age also shows moderate positive correlations with Ff (0.32) and Ft (0.30), reaffirming the critical role of hydration time in strength development. Most significantly, the interdependence between Ff and Ft is extremely high, which reflects their strong linear relationship as both are mechanical indicators of concrete performance.

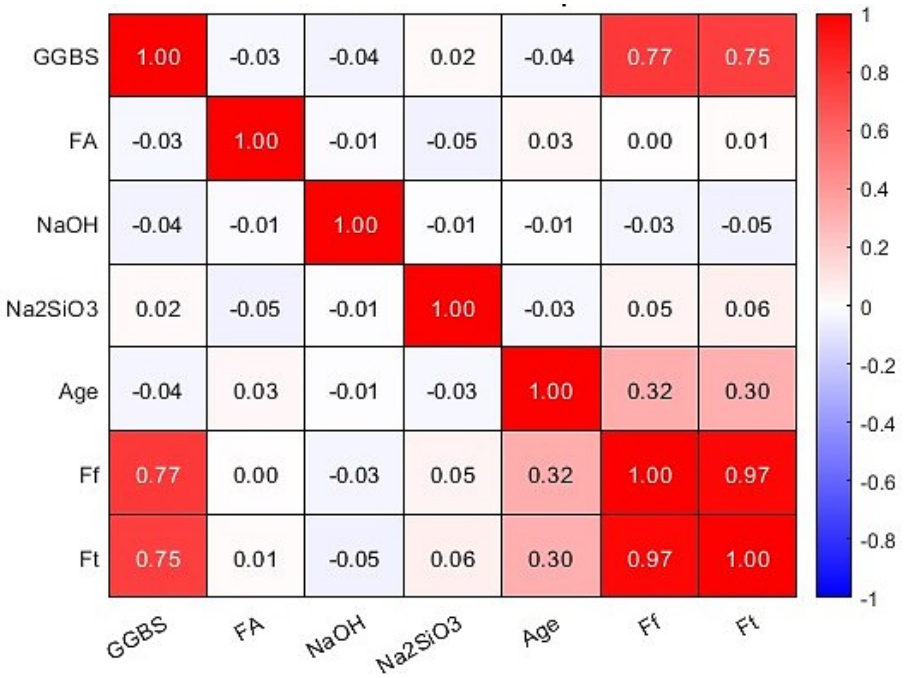

Figure 1. Correlation heatmap illustrates the Pearson correlation coefficients among input variables and output properties Ff, and Ft

2.1.3. Suitability of Dataset Size and Modeling Considerations

The dataset used in this study comprises 544 records, which is considered moderately large for machine learning applications in the concrete domain. Prior studies have successfully trained reliable models on smaller datasets, often fewer than 500 samples, demonstrating that this size provides adequate variability and learning depth for predictive modeling [30,31]. For algorithms, this dataset size offers a good balance between training volume and prediction

accuracy. However, further data enrichment may be required for more complex deep learning architectures.

The study incorporates cross-validation and bootstrapping techniques to enhance model generalization and stability. Future improvements could include the application of data augmentation or synthetic data generation methods, especially if domain-specific edge cases or underrepresented mix designs are identified [32].

Figure 2 schematically presents the methodology flow, highlighting the sequence from data acquisition through preprocessing, training, evaluation, and validation. The integration of statistical checks, data balancing, and robust model validation techniques ensures the proposed framework's reliability and replicability.

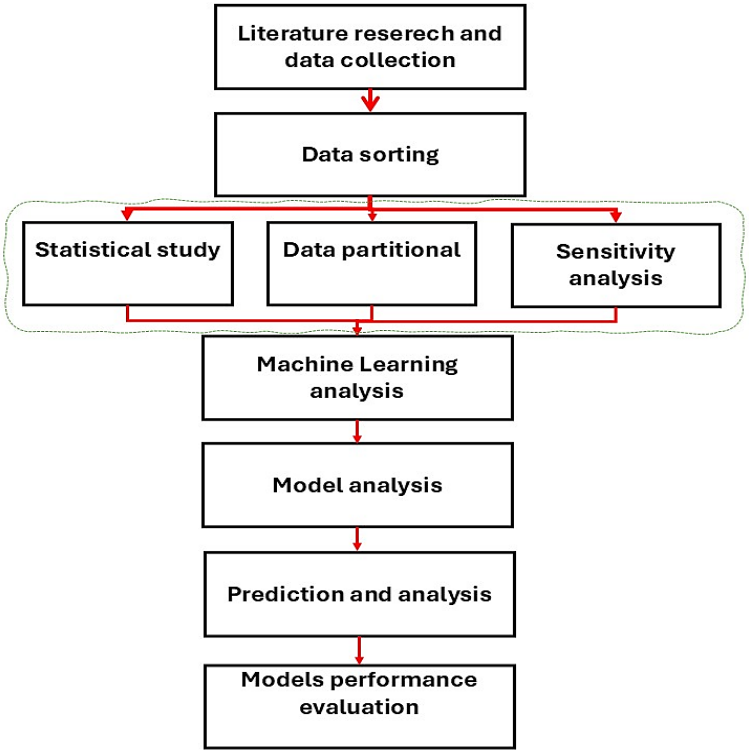

Figure 2. Workflow diagram

2.2. Machine Learning Framework

The selection of ML models plays a critical role in developing reliable predictive systems for estimating SCGC's mechanical behavior. This study employed eight supervised learning algorithms, including ensemble, rule-based, probabilistic, and kernel-based methods, to achieve both accuracy and interpretability.

2.2.1. Gradient Boosting (GB)

GB is an ensemble technique that constructs a series of weak models, commonly decision trees, where each model is trained to reduce the residual errors of the previous one. This step-by-step optimization process minimizes a specific loss function through gradient-based updates. The final predictive model is represented in Eq. (1).

$$y^{-}(x) = \sum_{m=1}^{M} \alpha_m h_m(x) \quad (1)$$

In the context of the GB model, the predicted output $y^{-}(x)$ is computed as the weighted sum of multiple weak learners, representing the final predicted value for a given input x. The

term α_m enotes the weight assigned to the m[th] weak learner, reflecting its contribution to the overall prediction. The function $h_m(x)$ refers to the individual weak learner, typically a decision tree, trained during the m^{th} iteration. The model constructs a total of M such learners, each aimed at minimizing the residual error from the previous iteration, thus refining the prediction iteratively. This ensemble approach enables Gradient Boosting to capture complex, nonlinear relationships and enhance predictive accuracy.

Its strength lies in achieving high predictive performance for complex regression tasks while effectively handling linear and non-linear relationships. Figure 3 represents the Gradient Boosting framework applied to regression tasks. The process begins with the initial dataset (Database X), where the first decision tree $f_1(x, \theta_1)$ is trained. Subsequent trees (f_2, f_3,..., f_k) are iteratively constructed to predict the residuals from the previous trees. Each new tree corrects the errors of the ensemble up to that point. The final model output is the sum of predictions from all trees $\sum f_k(x, \theta_k)$, where an objective function guides node splitting to minimize prediction error.

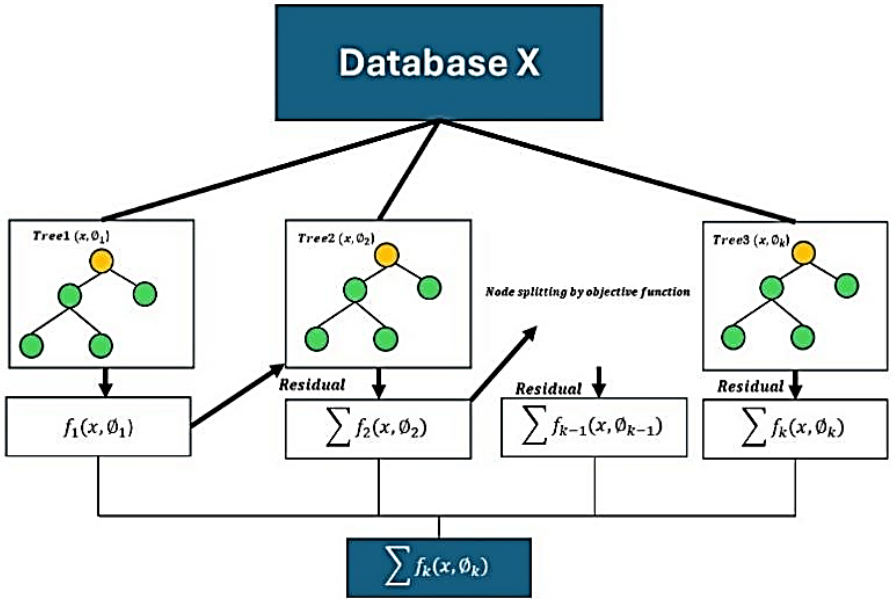

Figure 3. Schematic representation of the Gradient Boosting framework applied to regression tasks

2.2.2. CN2 Rule Induction

The CN2 algorithm is a rule-based learning method designed to produce understandable decision rules, often used for classification problems. It creates clear if-then statements by maximizing a quality measure that evaluates how effectively each rule represents the data [33]. A commonly used heuristic is in Eq. (2).

$$Score = \frac{Correct\ Predictions}{Total\ Predictions} \times \log\left(\frac{Correct\ Predictions}{Total\ Observations}\right) \quad (2)$$

CN2 supports categorical and numerical features and is appreciated for its transparency and comprehensibility [29]. Figure 4 shows the CN2 framework. The left side shows the construction of input behavior indicator sequences (S1 to Sn) from raw input entries, while the right side depicts the construction of output technical indicators (P1 to Pn). The final output is a set of predictions or classifications (C1 to CN), reflecting the model's learned patterns from both behavioral and technical indicators.

https://doi.org/10.31436/iiumej.v26i3.3730

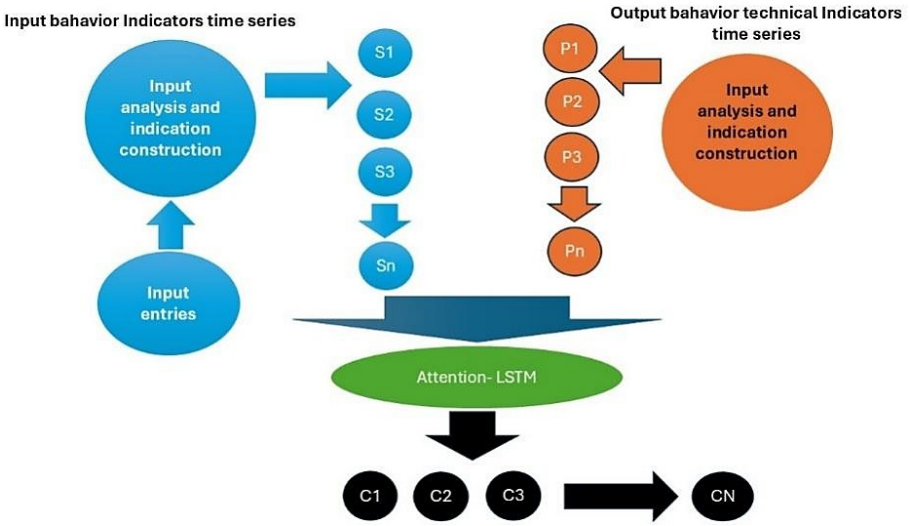

Figure 4. Diagram illustrating the Attention-LSTM framework for time-series modeling of input and output behavior indicators

2.2.3. Naive Bayes Classifier

Naive Bayes is a probabilistic model rooted in Bayes' Theorem, assuming conditional independence among features given the class label [34]. Despite its simplicity, it is highly effective for high-dimensional problems. The model estimates the posterior class probability as Eq. (3).

$$P(C|X) = \frac{P(X|C)\cdot P(C)}{P(X)} \tag{3}$$

where $P(C \mid X)$ is the probability of class C given feature vector X, $P(X \mid C)$ is the likelihood of X under class C, $P(C)$ is the prior class probability, $P(X)$ is the marginal probability of X. Figure 5 illustrates the Naïve Bayes classification process. The left side shows a dataset of mixed features (represented by different shapes and colors). The core of the process applies Bayes' Theorem, where the probability of class C given input X is computed based on prior and likelihood probabilities. The output on the right shows the data being classified into different categories (triangles, rectangles, circles), demonstrating the probabilistic sorting capability of the Naïve Bayes model.

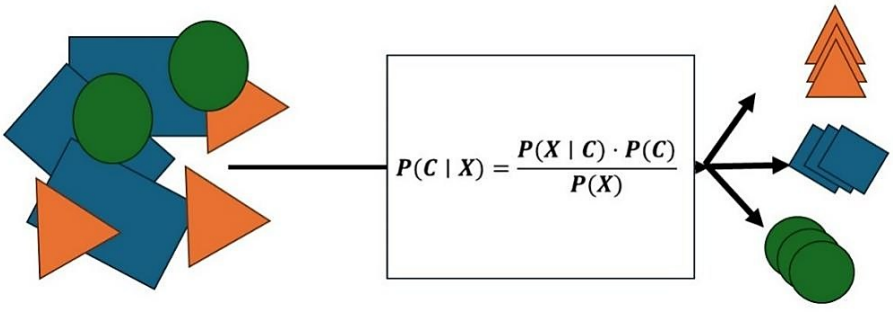

Figure 5. Illustration of the Naïve Bayes classification process

2.2.4. Support Vector Machines (SVM)

Support Vector Machines seek to construct a hyperplane that maximizes the margin between classes in a high-dimensional space [35]. For a linear SVM, the optimization problem is defined as Eq. (4).

https://doi.org/10.31436/iiumej.v26i3.3730

$$min\ \frac{1}{2}\ ||\ w\ ||^2\ \text{ subject to } y_i(w^t x_i + b) \geq 1, \forall_i \quad (4)$$

where w is the weight vector, x_i is the input feature vector, y_i is the corresponding class label, b is the bias term.

SVMs are well-suited for structured data and offer robust performance with different kernel functions. Figure 6 shows the architecture of an SVR model. On the left, the SVR network structure is depicted with an input layer receiving feature vectors $(x_1, x_2, \ldots, x_n)$, which are transformed through a kernel function $k(x_i,x)$ in the hidden layer. These kernel evaluations are aggregated in the output layer along with a bias term b, resulting in the predicted output y. The right side shows the concept of univariate linear SVR, where the goal is to fit a linear function $f(x) = wx + b$ such that the deviations from actual values lie within a defined margin ϵ. Points outside this margin incur an error penalty ε, forming the basis for the SVR loss function.

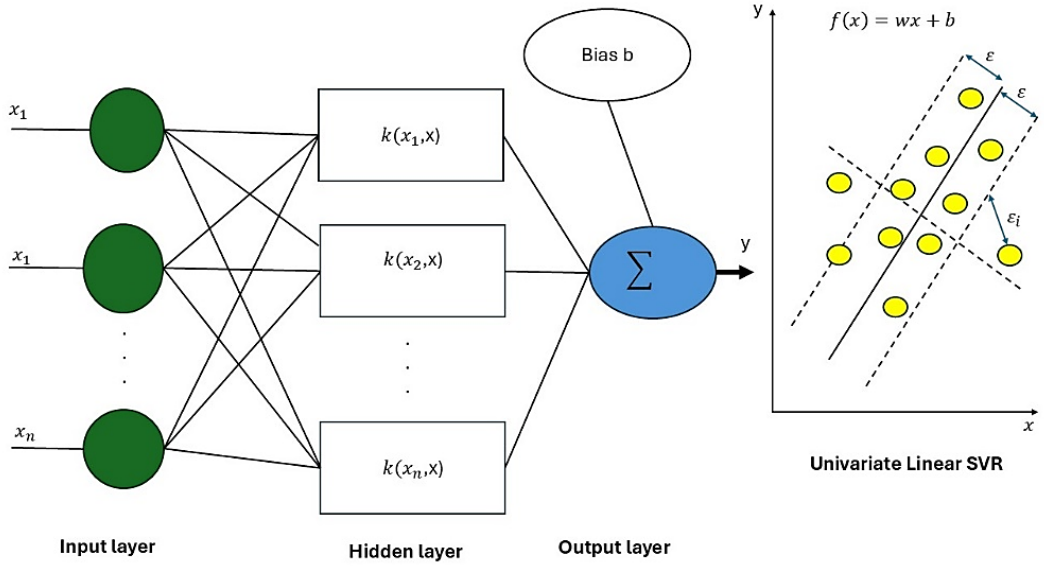

Figure 6. Architecture of a Support Vector Regression (SVR) model

2.2.5. Stochastic Gradient Descent (SGD)

SGD is an optimization algorithm used to minimize the loss function by iteratively updating model parameters based on gradients from small batches or individual training samples [36]. It is formulated as Eq. (5).

$$\Theta = \theta - \eta \nabla L\left(\theta; x_{i \mid y_i}\right) \quad (5)$$

where θ is model parameters, η is learning rate, ∇L is gradient of the loss function with respect to θ.

SGD is especially useful in large-scale or streaming data scenarios due to its computational efficiency [29, 30]. Figure 7 visualizes the gradient descent optimization process used in machine learning model training. The graph shows the relationship between model weight and loss reduction. Starting from an initial weight, the model updates its parameters toward the negative gradient, iteratively minimizing the loss function. The global minimum is achieved when the gradient becomes zero, indicating optimal model parameters with the lowest possible loss.

https://doi.org/10.31436/iiumej.v26i3.3730

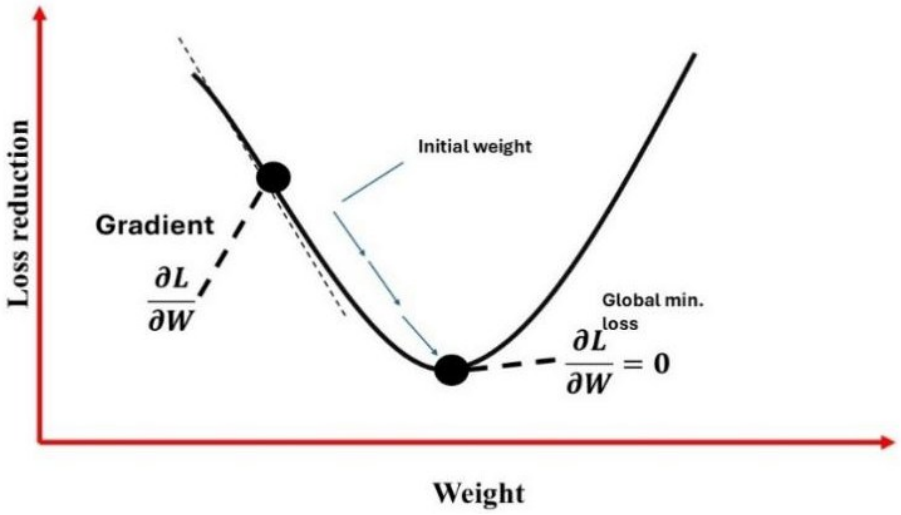

Figure 7. Gradient descent optimization process used in machine learning model training

2.2.6. K-Nearest Neighbors (KNN)

KNN is a non-parametric, instance-based learning method. It predicts the output based on the majority vote (classification) or average (regression) of the k-closest training samples [37] represented by Eq. (6).

$$F(x) = \frac{1}{k}\sum_{i \in kNN} y_i \tag{6}$$

where y_i is the labels of the nearest neighbors. KNN is intuitive and requires no training phase, but performance can be sensitive to feature scaling and distance metrics. Figure 8. Illustration of the K-Nearest Neighbors (K-NN) classification process. The left panel ("Before") shows a new data point introduced into a feature space defined by variables, surrounded by known samples from Category A and Category B. Using the K-NN algorithm, the right panel ("After") depicts how the new data point is classified based on its proximity to the majority.

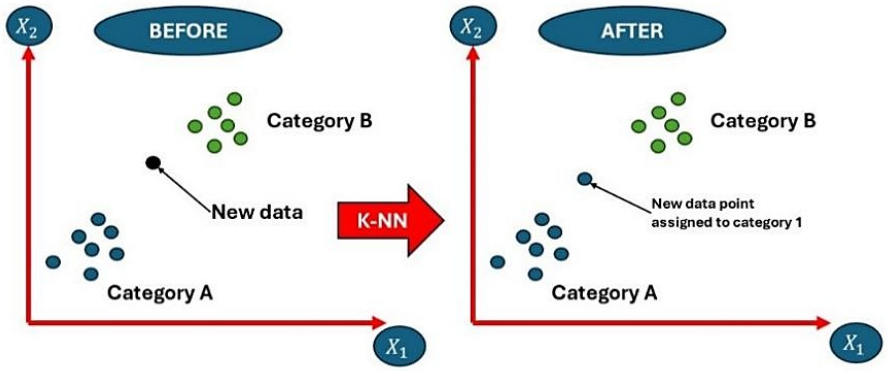

Figure 8. K-Nearest Neighbors (K-NN) classification process

2.2.7. Decision Trees (DT)

Decision Trees are recursive models that partition the data based on feature splits that maximize information gain or minimize impurity [38]. The Gini Index is a common criterion used for classification tasks represented in Eq. (7).

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \tag{7}$$

https://doi.org/10.31436/iiumej.v26i3.3730

where p_i is the proportion of class i in a node, and C is the number of classes. Trees are simple to interpret and can model non-linear relationships effectively. Figure 9 is a structure of a decision tree model. The diagram illustrates the hierarchical flow of decisions starting from the root node (top red circle), followed by multiple split nodes (intermediate red circles) representing decision rules based on input features. The process continues until reaching the leaf nodes (yellow circles), which represent final predictions or class labels. This structure enables the model to partition the data space for classification or regression tasks recursively.

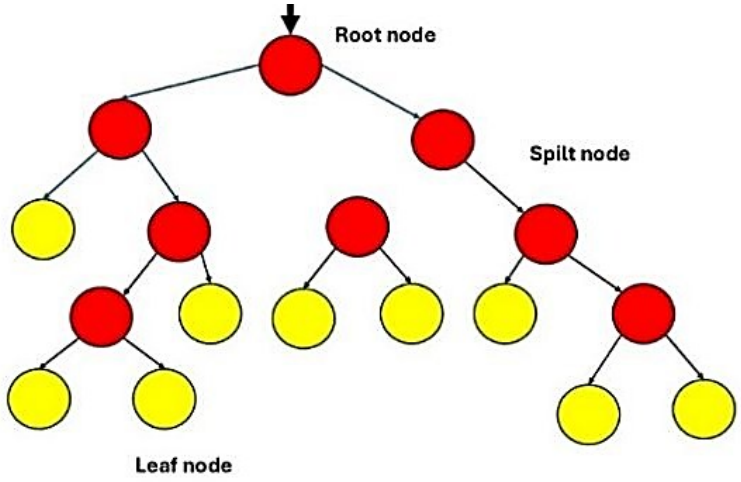

Figure 9. Structure of a decision tree model

2.2.8. Random Forest (RF)

Random Forest is an ensemble of decision trees built on bootstrapped samples of the dataset. Predictions are aggregated through majority voting or averaging [39]. The ensemble output is given by Eq. (8):

$$y^{-} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \tag{8}$$

where T is the number of trees, $h_t(x)$ output from the t^{th} tree.

Random Forest improves generalization by reducing overfitting, a common issue in single decision tree models. Figure 10. Illustration of a decision tree using logical conditions for binary classification. The tree begins with a root condition on $Variable_a \geq x$. If true, the data point is classified as TRUE (green node); if false, the tree proceeds to evaluate $Variable_b \leq y$. Further branching leads to additional conditions until a final classification of TRUE or FALSE (green or red leaf node) is reached. This logical flow demonstrates how decision rules are applied in a tree-based model to reach categorical outcomes.

https://doi.org/10.31436/iiumej.v26i3.3730

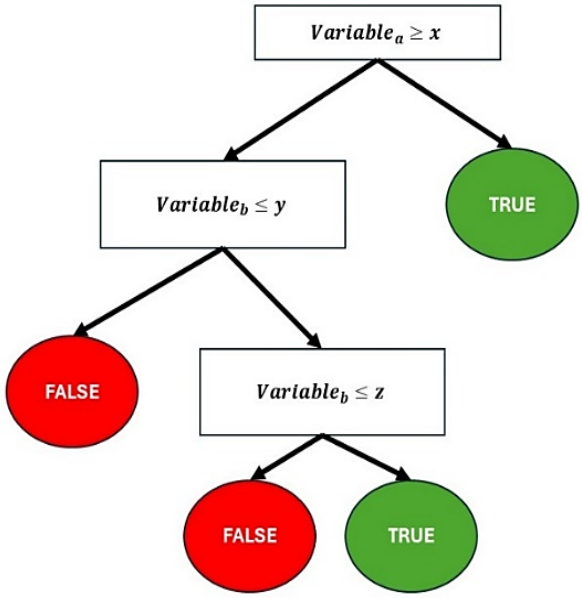

Figure 10. Decision tree using logical conditions for binary classification

The choice of algorithm in this study was guided by theoretical characteristics and empirical performance in modeling SCGC properties. Hyperparameter tuning was conducted using grid search and cross-validation to achieve optimal configurations. These algorithms offer a spectrum of benefits, ranging from interpretability (e.g., CN2, Decision Trees) to high accuracy and scalability (Gradient Boosting, Random Forest, SVM), thereby providing a robust framework for predictive analysis of SCGC mechanical behavior.

2.3. Response Surface Methodology (RSM), Hoffman & Gardener, and Analysis of Variance (ANOVA)

RSM statistical use for modeling and optimizing processes in which multiple input variables influence a quantitative outcome or response. It provides a systematic methodology for understanding the interactions among process variables and for determining the optimal operating conditions of complex systems [40]. RSM constructs an empirical model, commonly a second-degree polynomial, that approximates the actual relationship between the independent variables and the output response, as expressed in Eq. (9).

$$Y = \beta_0 + \sum \beta_i Xi + \sum \beta_{ii} X_i^2 + \sum \beta_{ij} X_i X_j + \varepsilon \quad (9)$$

where Y is the response variable, X_i are the input factors, β terms are the regression coefficients, and ε represents the experimental error [4].

In this study, RSM was incorporated alongside ML models (KNN and SVM to leverage the strengths of both analytical and data-driven approaches. While ML algorithms are highly effective for capturing complex, non-linear patterns in experimental data, they often lack transparency regarding how input variables influence the response. RSM addresses this gap by offering an interpretable, regression-based framework that models variable interactions and supports optimization through structured mathematical formulations. Therefore, the integration of RSM is not aimed at outperforming ML in predictive capacity, but rather at enhancing the interpretability, parametric sensitivity analysis, and optimization of mix design parameters. As illustrated in Figure 11, the accompanying decision-making diagram, both approaches operate on the same dataset but fulfill complementary roles. RSM provides equation-based insights and

design guidance, while ML supports robust pattern recognition and generalization across a wide input space.

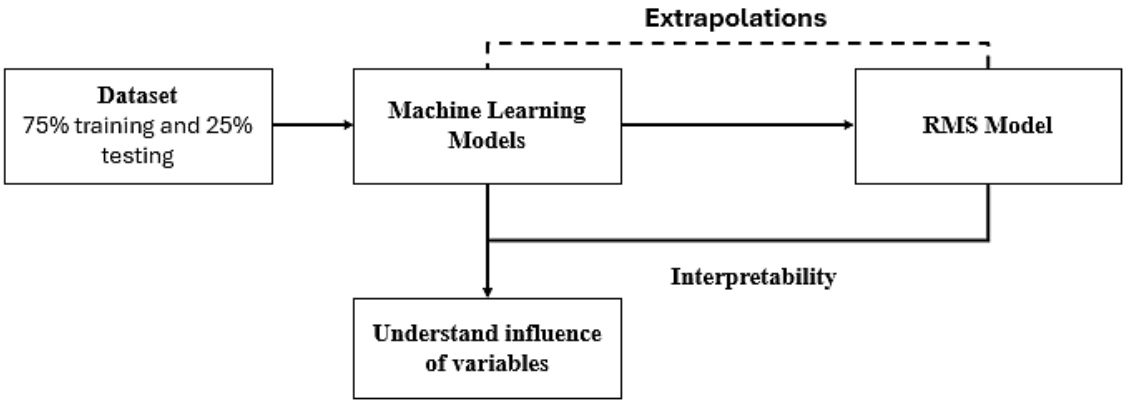

Figure 11. Methodological framework showing the complementary roles of RSM and machine learning

ANOVA is a statistical method used to break down the overall variability in a response variable into separate components, such as those caused by experimental conditions and inherent randomness. In the context of RSM, ANOVA plays a key role in evaluating the significance and adequacy of the regression model developed [41]. It tests the null hypothesis that the means of various factor levels are equal and provides an F-ratio to determine if the regression model explains a significant portion of the variance in the data. The formula is as in Eq. (10)

$$F = \frac{Mean\ Square\ Regression\ (MSR)}{Mean\ Square\ Error\ (MSE)} \tag{10}$$

A low p-value (typically < 0.05) associated with the F-statistic indicates that the model terms significantly contribute to the response prediction. ANOVA also provides insight into the contribution of linear, interaction, and quadratic terms, allowing researchers to refine the model structure by retaining only statistically significant terms [41].

RSM and ANOVA form a robust and complementary framework for prediction and optimization in experimental research. RSM enables the construction of mathematical models that capture how input parameters influence the output response, whereas ANOVA provides a means to statistically verify the model's validity, effectiveness, and predictive strength This integrated approach enhances decision-making by enabling researchers to identify the main and interaction effects of key process variables, assess model goodness-of-fit and lack-of-fit, and determine optimal factor combinations for achieving target performance outcomes. Moreover, by systematically exploring the experimental space, the RSM–ANOVA combination reduces the required trials, maximizing information gain with minimal resource use. In the context of materials science, particularly in the formulation of SCGC, this dual methodology proves especially valuable for fine-tuning mix designs and curing conditions to achieve superior mechanical properties cost- and time-efficiently.

The same 75% training and 25% testing data split was consistently applied to all modeling approaches, including the machine learning algorithms and the RSM models, to ensure a fair and unbiased comparison across modeling techniques. This uniform data partitioning strengthens the validity of comparative performance evaluations and mitigates potential biases introduced by differing datasets.

In addition to RSM and ANOVA, a sensitivity analysis was conducted using the Hoffman & Gardener method to evaluate how each input variable affects the predicted mechanical performance outcomes. Although not a standard sensitivity approach in concrete modeling, this method has been successfully applied in recent machine learning studies involving concrete composites [42]. It involves systematically perturbing one input at a time while keeping the others constant and evaluating the resulting change in the output prediction from trained models. [43].

Although the method involves considerable computational effort, it offers significant benefits in complex modeling tasks where identifying the influence of input variables is essential for optimization and informed decision-making. In this research, an initial sensitivity investigation was performed using the developed dataset to assess how each independent variable contributes to the outcome Y. A one-factor-at-a-time approach was applied, in which a single input is varied. In contrast, all others remain fixed, enabling a focused evaluation of individual variable influence. The Sensitivity Index (SI) for each input was computed using the Hoffman & Gardner technique [44], defined by the following Eq. (11).

$$SI(X_N) = \frac{Y(X_{max}) - Y(X_{min})}{Y(X_{max})} \tag{11}$$

A sensitivity value equal to 1.0 signifies maximum responsiveness of the output to that input variable. In contrast, a value below 0.01 reflects negligible influence, suggesting that the model's predictions remain largely unaffected by variations in that parameter.

2.4. Performance Accuracy Evaluation of Machine Learning Models

A set of statistical performance metrics was employed to assess the predictive capabilities of the developed ML models for estimating the mechanical properties of SCGC.

2.4.1. Error-Based Metrics

Sum of Squared Errors (SSE) quantifies the total squared deviation between actual and predicted values, serving as a cumulative indicator of prediction error. While useful for comparative analysis, SSE is scale-dependent and less intuitive when comparing across datasets [45]. MAE (Mean Absolute Error) measures the average magnitude of prediction errors, offering a direct and easily interpretable estimate of how far predictions deviate from actual values [46] represented by Eq. (12).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - y_i^-| \tag{12}$$

The Mean Squared Error (MSE) is a widely used metric for evaluating the accuracy of predictive models. It is calculated by squaring the prediction errors (the differences between observed and predicted values) and then averaging these squared errors. This squaring process amplifies the influence of larger deviations, making MSE particularly sensitive to outliers [47]. The formula represented by Eq. (13).

Root Mean Squared Error (RMSE) is a widely used metric for evaluating the performance of models, particularly in fields like hydrology, geoscience, and machine learning. calculates the square root of the mean squared deviations between actual and predicted values, offering a concise indicator of overall prediction accuracy [48]. The formula represented by Eq. (14).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^-)^2 \tag{13}$$

https://doi.org/10.31436/iiumej.v26i3.3730

$$RMSE = \sqrt{MSE} \tag{14}$$

These metrics (MAE, MSE, $RMSE$) provide complementary perspectives on error behavior and are essential for understanding average and extreme prediction performance.

2.4.2. Normalized Performance Measures

Error (%) expresses prediction error as a percentage relative to the predicted value, enabling model comparisons across variables or datasets with different scales. The formula is represented by Eq. (15). Accuracy (%) is calculated as the complement of error, representing the proportion of accurate predictions in percentage terms. The formula is represented by Eq. (16).

$$\text{Error}(\%) = \frac{\text{RMSE}}{\bar{y_i}} \times 100 \tag{15}$$

$$\text{Accuracy}(\%) = 100 - \text{Error} \tag{16}$$

Although frequently associated with classification tasks, Error (%) and Accuracy (%) can also be adapted for use in regression settings to provide scale-independent performance indicators.

2.4.3. Coefficient of Determination (R^2)

R^2 indicates the proportion of variance in the observed data that the model explains. It ranges from 0 to 1, with values closer to 1 indicating a better fit. Unlike absolute error metrics, R^2 evaluates model goodness-of-fit in terms of variance reduction, and the formula represented by Eq. (17).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y_i})^2}{\sum(y_i - \bar{y})^2} \tag{17}$$

Here, y_i are the observed values, $\hat{y_i}$ are the predicted values, and $\bar{y}$ is the mean of observed values. A high R^2 implies that the model captures the underlying patterns of the data effectively [49].

2.4.4. Multi-Metric Evaluation Strategy

Given the limitations and strengths of each metric, a multi-metric evaluation framework was adopted. For instance, a model may achieve a high R^2 while yielding large MAE or RMSE if influenced by outliers. Similarly, low error percentages do not necessarily guarantee high explanatory power if model predictions fail to capture data variance. Therefore, combining these metrics enables a balanced and comprehensive assessment of performance. This performance evaluation framework ensures that the selected ML models for SCGC prediction are statistically robust and practically relevant across varying strength parameters.

3. RESULTS AND DISCUSSION

3.1. Model Performance

3.1.1. Gradient Boosting (GB) model

GB models were implemented using the Scikit-learn library. The models employed a constant rate of 0.1 and a minimum sample split threshold of 2. A series of nine model setups

https://doi.org/10.31436/iiumej.v26i3.3730

was tested by progressively varying the number of decision trees (from 1 to 3) and their corresponding depth levels (from 1 to 3). The influence of these parameters on prediction error is illustrated in Figure 12, where a consistent decrease in error percentage is observed with greater model complexity. Specifically, increasing tree depth and count significantly enhanced model accuracy, particularly in reducing prediction error for Ff and Ft.

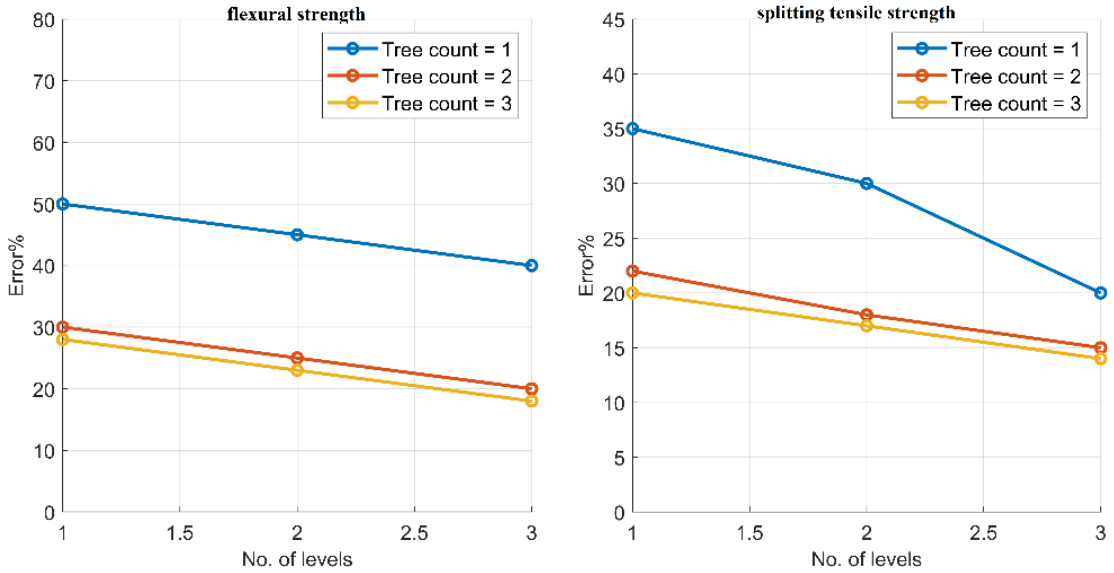

Figure 12. Gradient Boosting (GB) model performance for predicting Ff and Ft

The optimal GB model configuration was identified as two trees with a maximum depth of three levels. Under this configuration, the GB models achieved average prediction accuracies of 90% for Ff and 88% for Ft. Figure 13 and Figure 14 illustrate the correlation between the predicted strength values and the corresponding experimental results. In the training phase, the model attained an R^2 of 0.97 for Ff and 0.95 for Ft, with a near-ideal regression slope (y = 0.99x), indicating high predictive alignment. In the validation phase, the R^2 remained robust at 0.96, affirming the model's generalization capability and reliability in estimating the tensile-related properties of SCGC.

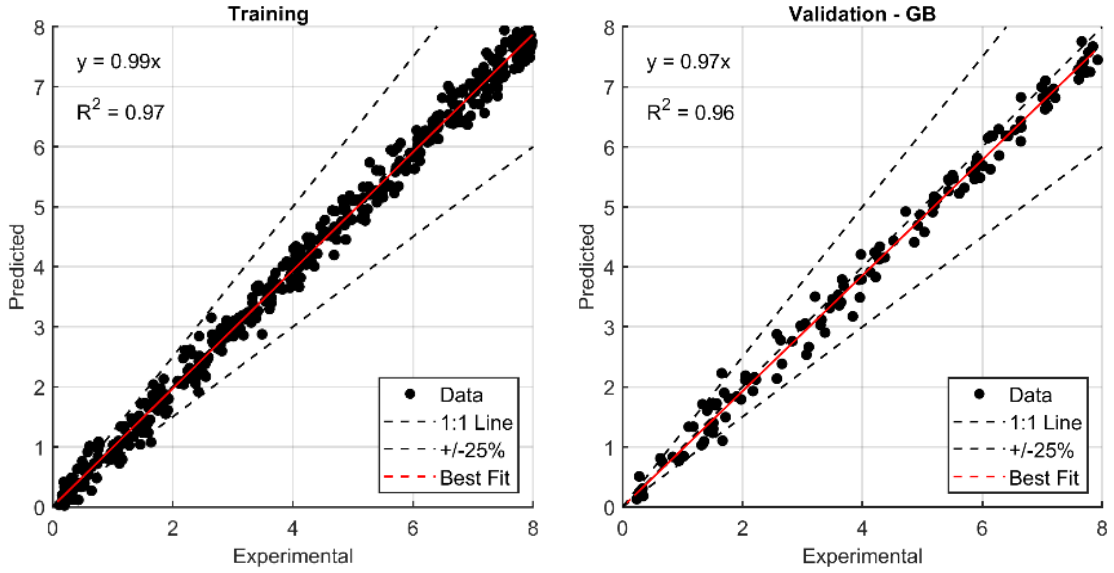

Figure 13. Regression analysis comparing predicted and experimental values for Ff using the Gradient Boosting model

https://doi.org/10.31436/iiumej.v26i3.3730

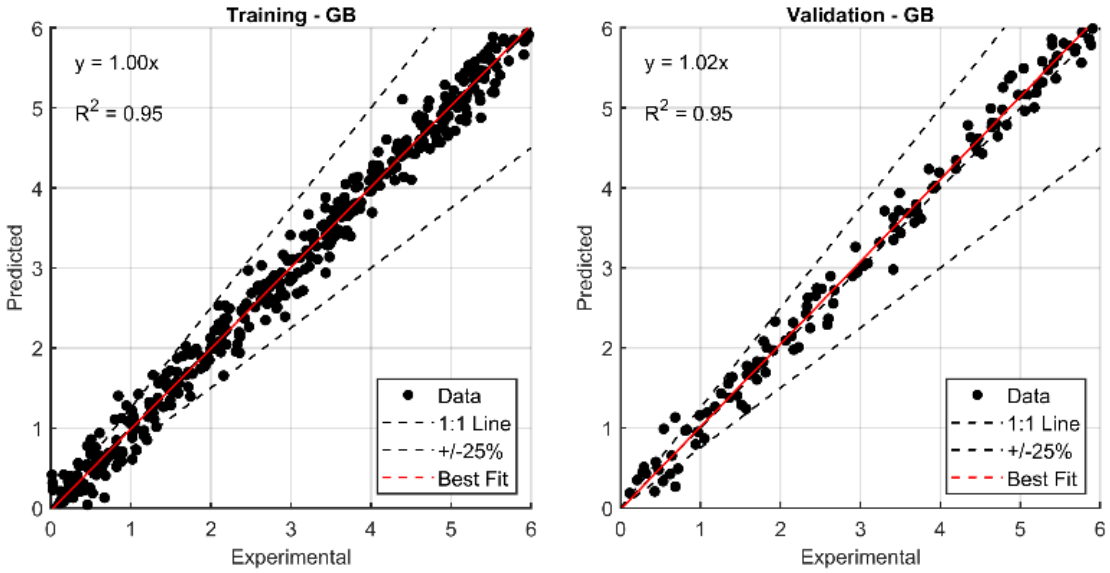

Figure 14. Regression analysis comparing predicted and experimental values for Ft using the Gradient Boosting model

3.1.2. CN2 model

Nine CN2-based rule induction models were constructed to model the mechanical performance of SCGC, Ff, and Ft. The assessment was based on Laplace accuracy, using a beam width of 1.0 and setting the minimum coverage for rule generation to 1.0. Model refinement was done by gradually extending the maximum rule length from 1 to 5. As depicted in Figure 15, extending the rule length consistently improved prediction accuracy, with the configuration at a length of 5.0 yielding the lowest error rates for both target properties.

Each developed CN2 model generated a comprehensive set of interpretable "IF–THEN" rules 42 for Ff, and 31 for Ft. These rules captured complex relationships between the input mix design parameters and the corresponding strength outcomes.

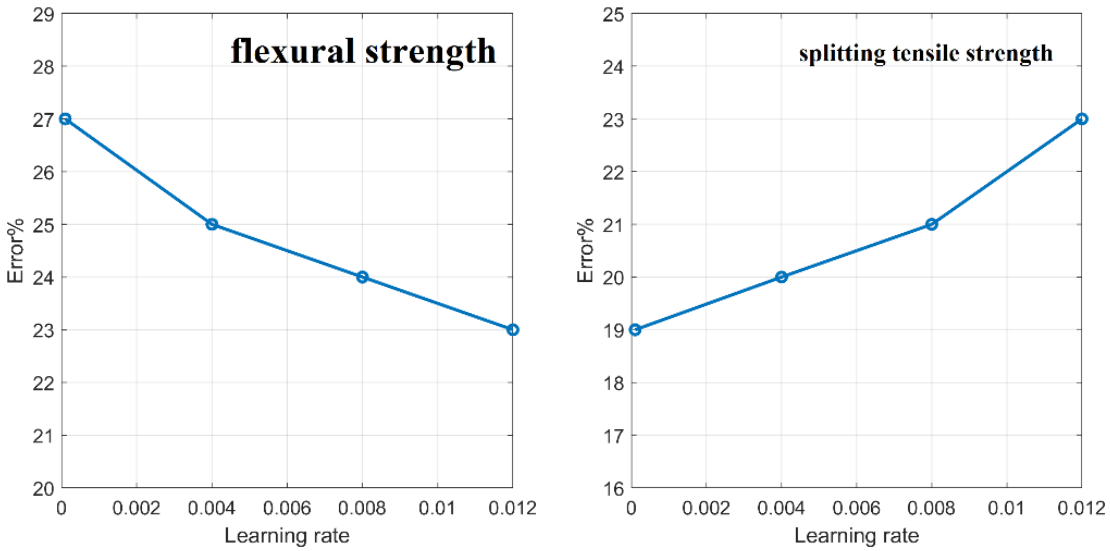

Figure 15. Error (%) reduction in predicting Ff and Ft using the CN2 model as the rule length increases from 1 to 5

https://doi.org/10.31436/iiumej.v26i3.3730

The regression plots in Figs. 16 and 17 demonstrate the close agreement between the predicted and experimental strength values across all three strength types. For Ff and Ft, the training and validation phases achieved high coefficients of determination ($R^2 \approx 0.98$ and 0.94), with nearly ideal regression slopes. These results validate the effectiveness of the CN2 model in capturing the governing patterns of SCGC mechanical behavior, while offering transparent decision rules for practical implementation in sustainable concrete mix design.

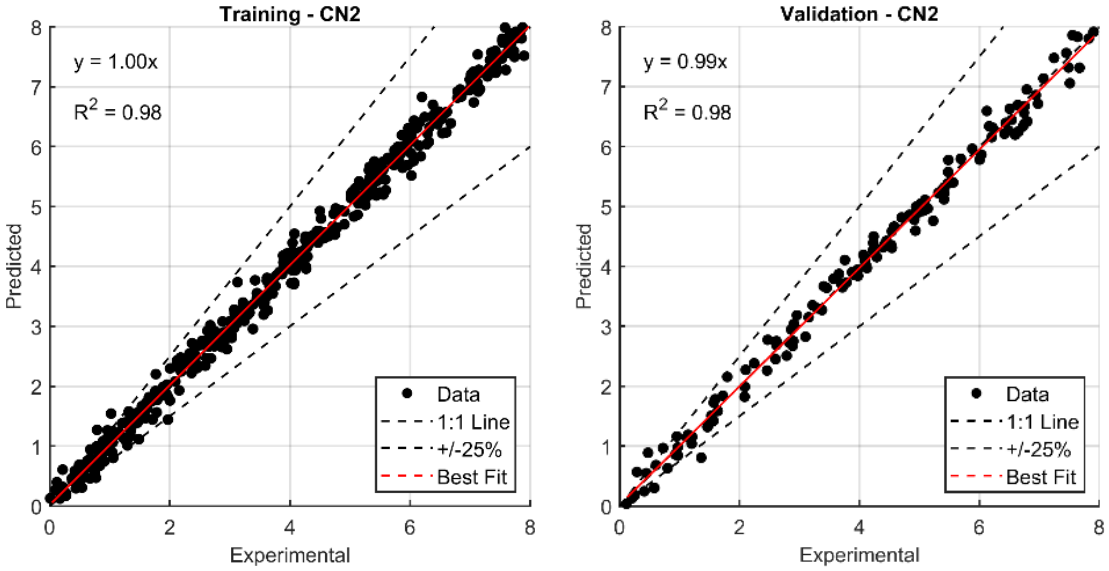

Figure 16. Regression analysis of CN2 model predictions versus experimental results for flexural strength

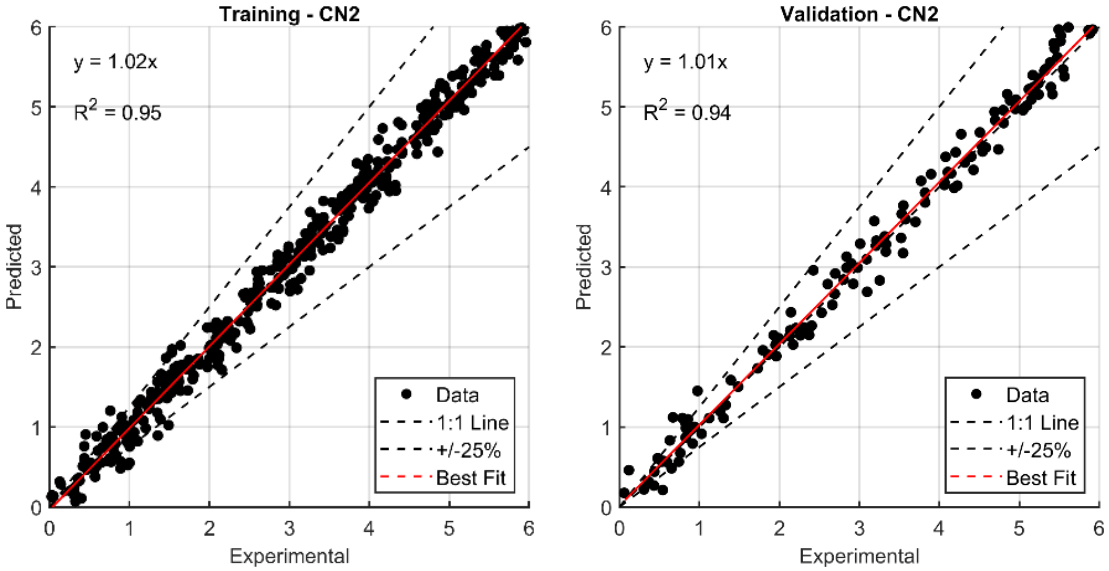

Figure 17. Regression analysis of CN2 model predictions versus experimental results for splitting tensile strength

3.1.3. Support Vector Machine (SVM) model

SVM models were employed to estimate the mechanical properties of SCGC, using a polynomial kernel function. Model optimization was performed by assigning a regularization constant (C) of 100, setting the loss margin for regression to 0.10, and applying a convergence threshold of 1.0. Polynomial degrees from 1 (linear) to 3 (cubic) were methodically explored

https://doi.org/10.31436/iiumej.v26i3.3730

across multiple configurations to assess how complexity influences predictive performance. As shown in Figure 18, increasing the kernel degree led to a noticeable reduction in prediction error for flexural and tensile strength, with the quadratic kernel (degree 2) providing the best balance between complexity and accuracy.

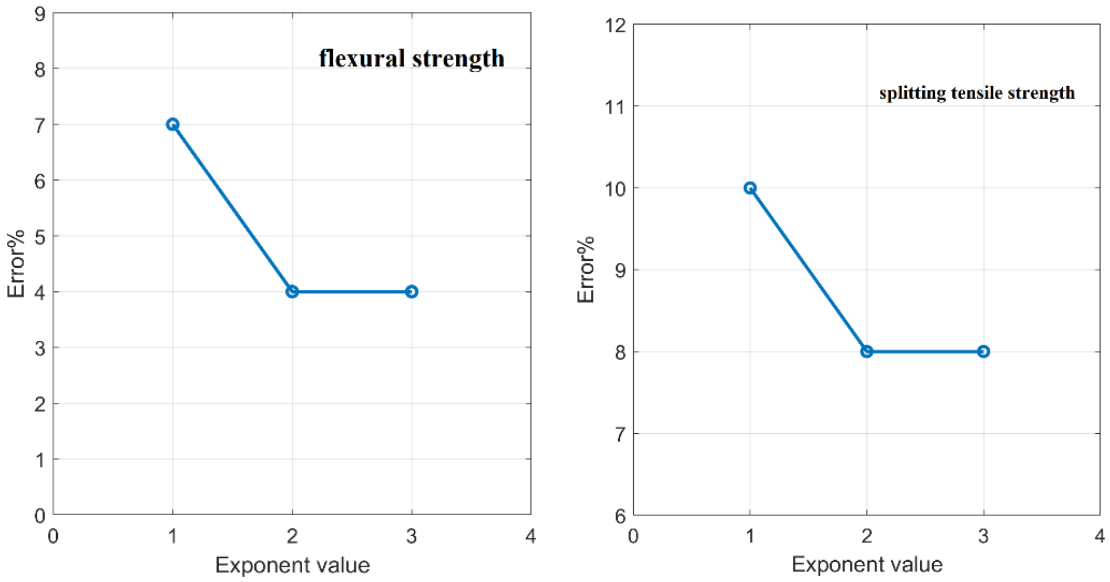

Figure 18. Prediction error (%) for Ff and Ft using SVM models with polynomial kernels of increasing degree

The SVM models demonstrated excellent predictive capacity across all target properties, with average accuracies of 99% and 99% for Ff and Ft, respectively. Regression plots in Figs. 19 and 20 confirm the strong correlation between experimental and predicted values. The models achieved R^2 values of up to 0.99 for each strength property, with regression lines closely aligned with the ideal 1:1 line. These results affirm the robustness and high precision of the SVM models in estimating SCGC strength performance, particularly when using a second-degree polynomial kernel.

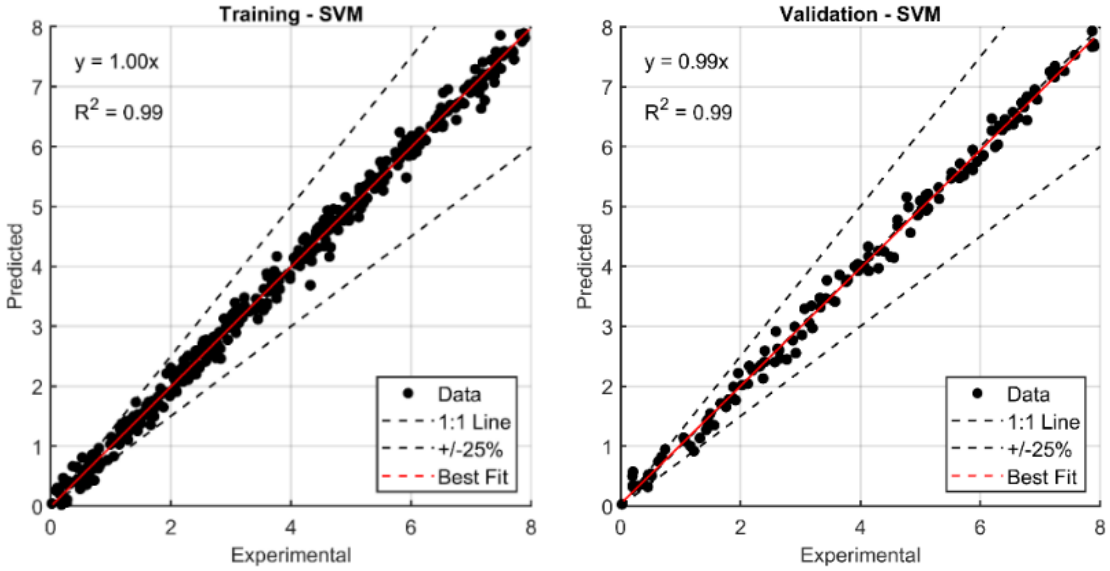

Figure 19. Regression plots comparing predicted and experimental results for Ff using the optimized SVM model (quadratic kernel)

https://doi.org/10.31436/iiumej.v26i3.3730

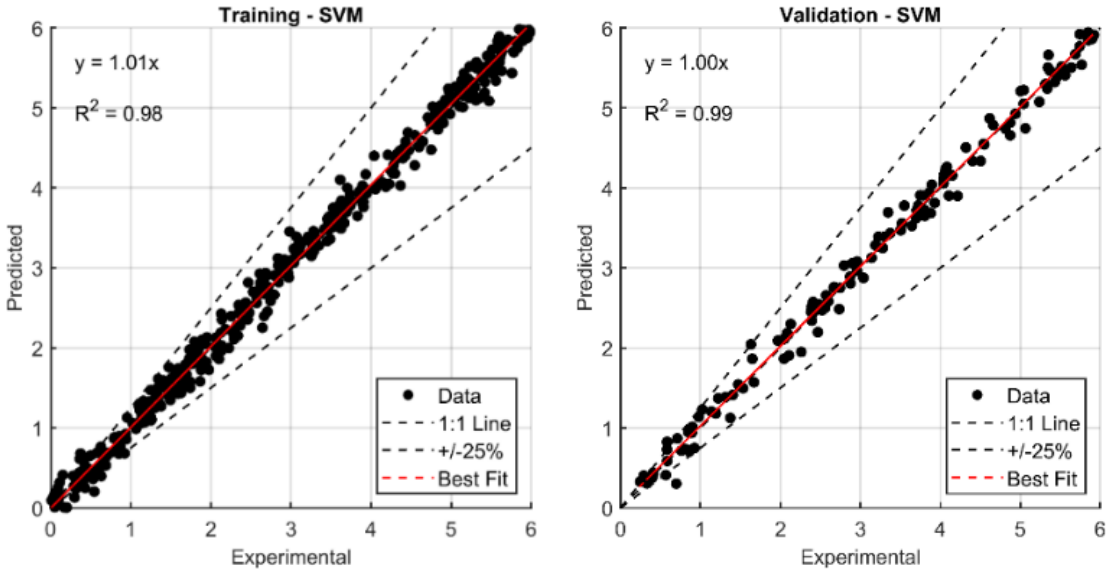

Figure 20. Regression plots comparing predicted and experimental results for Ft using the optimized SVM model (quadratic kernel)

3.1.4. Stochastic Gradient Descent (SGD model results

SGD models were developed using the modified Huber loss and elastic net regularization, with a blending ratio set at 0.01 and a regularization coefficient of 0.001. The models were evaluated across a learning rate range that began at 0.01 and gradually decreased to 0.001. As shown in Figure 21, learning rate adjustments had a significant impact on prediction error, with the lowest error observed at the smallest learning rate tested.

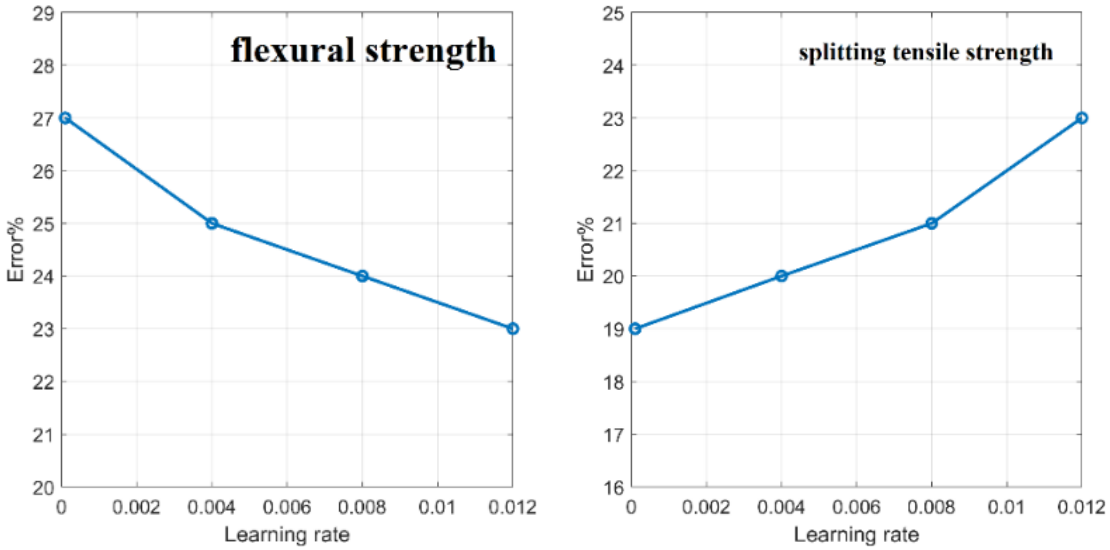

Figure 21. Effect of learning rate on prediction error (%) for Ff and Ft using the SGD model

Nine SGD model configurations were analyzed to estimate flexural and splitting tensile strength. Figure 21 and 22 present regression plots of predicted versus experimental values for both training and validation phases. The R^2 values ranged between 0.89 and 0.93 for flexural and tensile strengths. The results confirm the model's moderate yet consistent performance across the three strength categories, with good alignment along the 1:1 prediction line.

https://doi.org/10.31436/iiumej.v26i3.3730

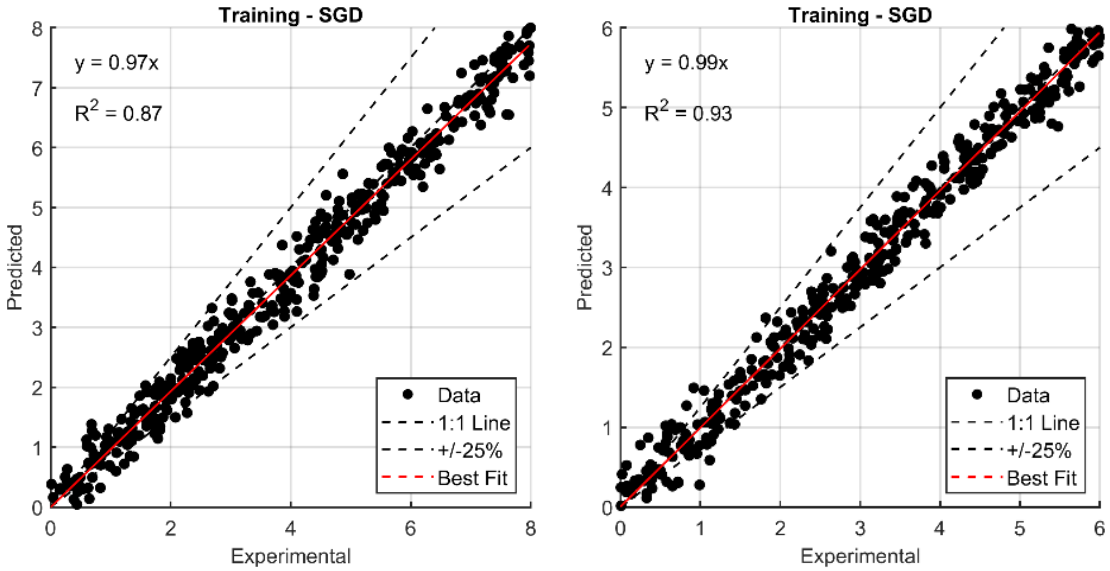

Figure 22. Predicted vs. experimental plots for Ff using the optimized SGD model

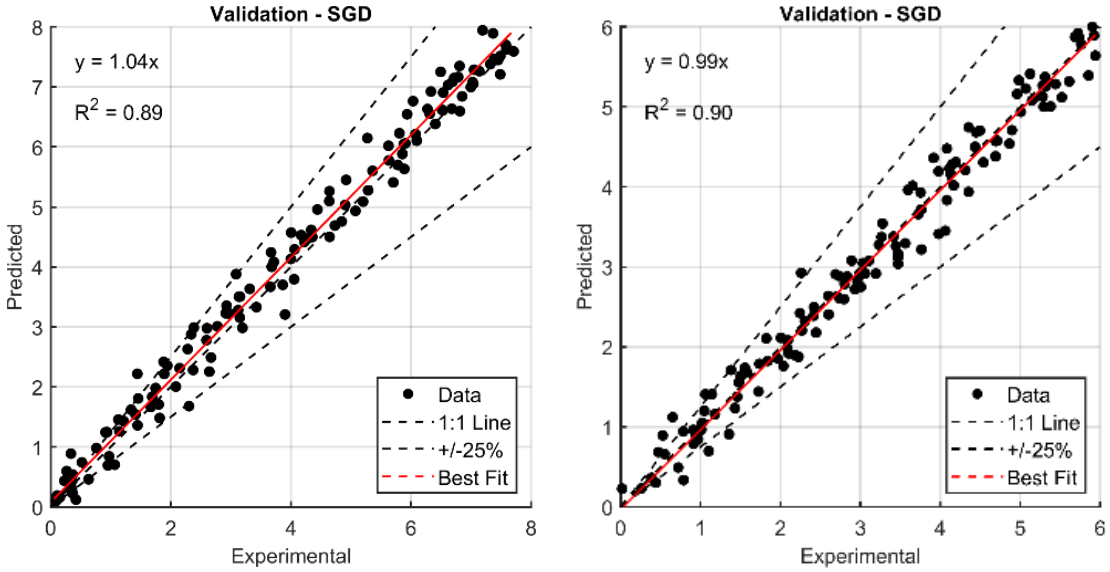

Figure 23. Predicted vs. experimental plots for Ft using the optimized SGD model

3.1.5. K-Nearest Neighbors (K-NN) model results

The K-Nearest Neighbors (K-NN) models were developed to predict the flexural and splitting tensile strength of SCGC. The optimal model configuration was determined using a single neighbor ($k = 1$) and the Euclidean distance metric, where each prediction was weighted according to the inverse of the distance from the query point. This setup allows the model to prioritize the most similar past observations, making it highly responsive to local patterns within the dataset. K-NN exhibited the highest predictive performance among all the machine learning algorithms evaluated in this study. The average prediction accuracies reached 99% for Ff, and 94% for Ft, indicating the model's exceptional capability in capturing the nonlinear interactions among input variables, as in Figs. 24 and 25. Its non-parametric and instance-based learning nature makes it particularly suitable for datasets with high variability and strong local dependencies, such as those found in SCGC mix designs.

https://doi.org/10.31436/iiumej.v26i3.3730

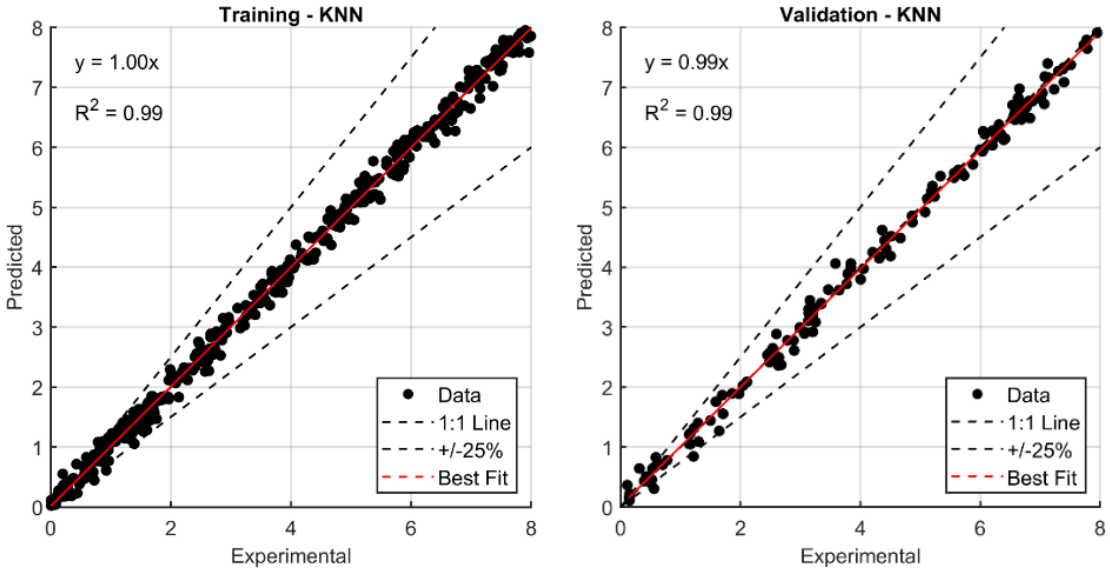

Figure 24. Predicted vs. experimental plots for Ft using the optimized K-NN model

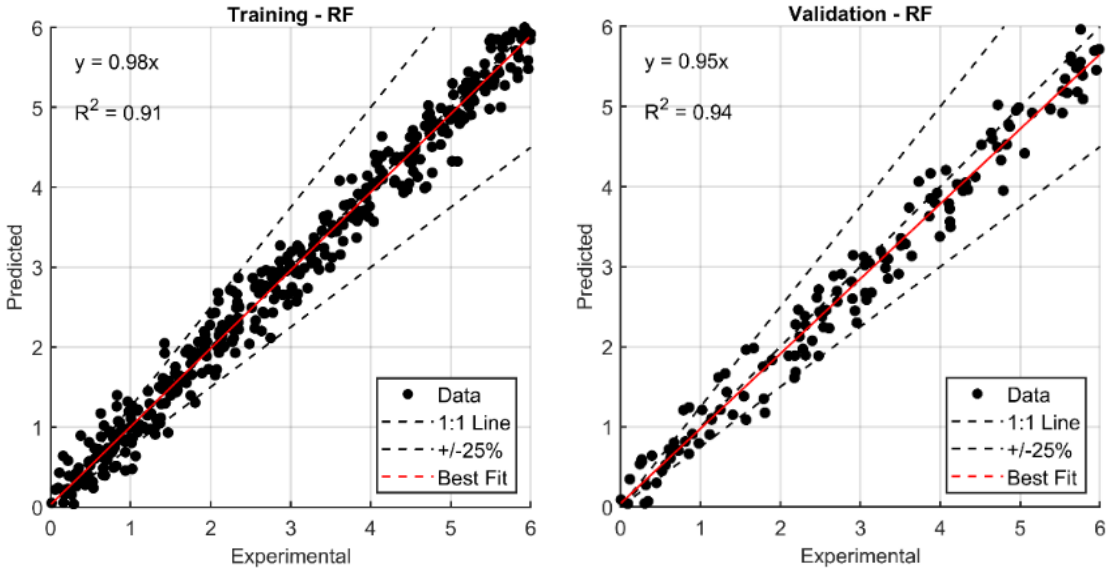

Figure 25. Predicted vs. experimental plots for Ft using the optimized K-NN model

3.1.6. Decision Tree (DT) model

DT models were built using a tiered node-splitting method. Each model was designed with at least 2.0 instances per leaf and a minimum split requirement of 5.0, balancing clarity of the decision rules with control over overfitting. Tree depth gradually increased, starting from a simple structure with one level and expanding up to five levels. This progressive layering allowed the algorithm to balance model complexity and predictive accuracy. The impact of tree depth on model performance is shown in Figure 26, where a consistent reduction in prediction error (%) is observed for both Ff and Ft as the number of layers increases. The most notable performance gains occur between depths 1 and 3, with diminishing returns beyond this point, indicating an optimal complexity level near five layers.

https://doi.org/10.31436/iiumej.v26i3.3730

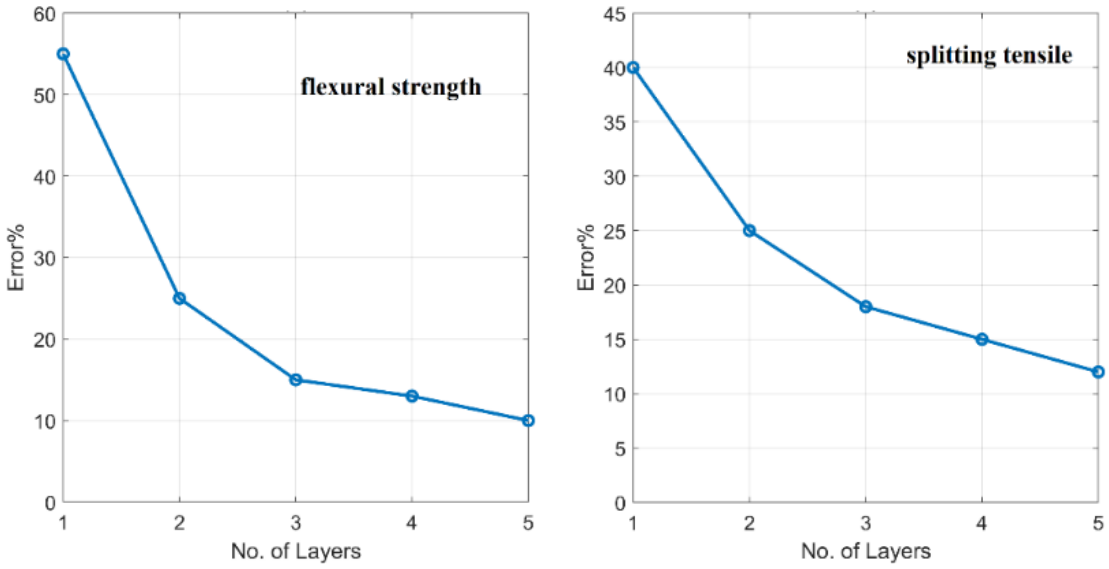

Figure 26. Error (%) reduction in predicting Ff and Ft

The final optimized models yielded strong performance, achieving average prediction accuracies of 96% and 97% for Ff and Ft, respectively, as in Figs. 27 and 28. These results confirm that the DT model is a highly interpretable and moderately high-performing tool, suitable for practical applications where model explainability is essential, such as field-based strength estimation and sustainable mix design optimization.

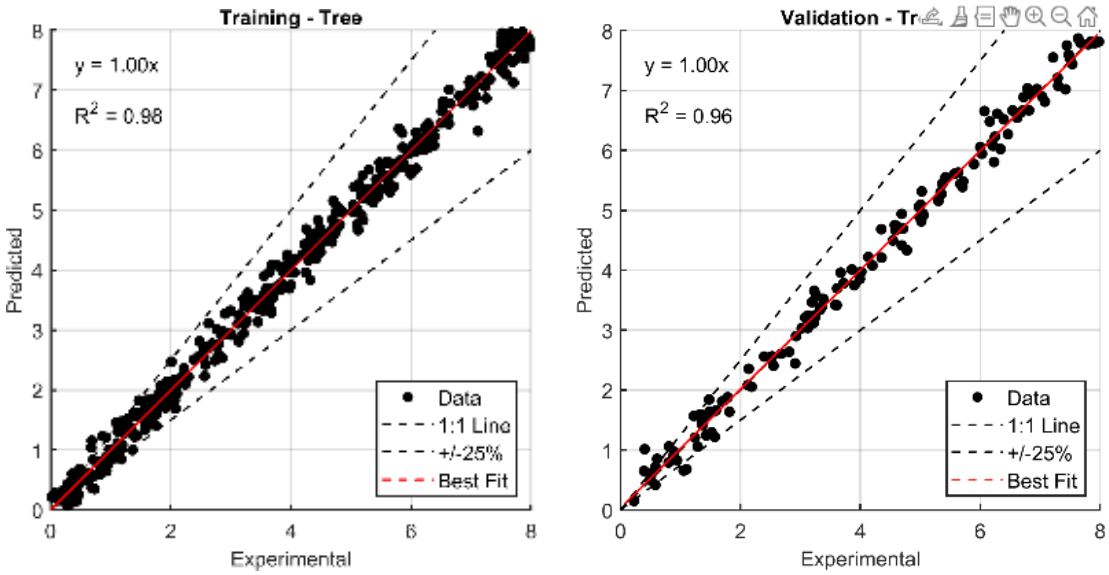

Figure 27. Predicted versus experimental values for Ff using the optimized Decision Tree model

https://doi.org/10.31436/iiumej.v26i3.3730

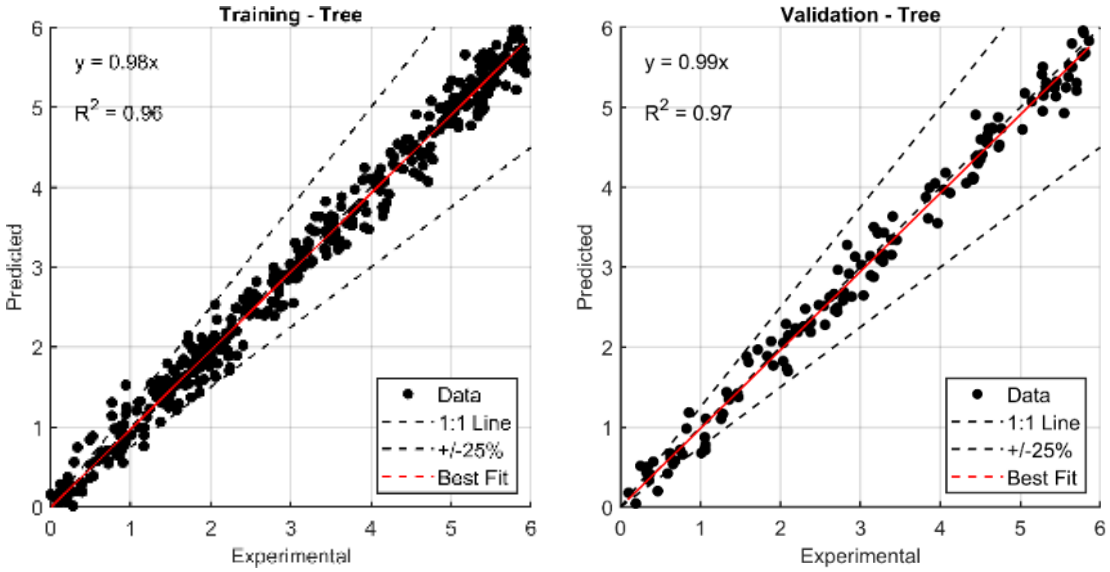

Figure 28. Predicted versus experimental values for Ft using the optimized Decision Tree model

3.1.7. Random Forest (RF) model

The Random Forest (RF) models were developed to predict the flexural and splitting tensile strength of SCGC. These models were generated through a progressive configuration process starting with a single tree of depth one and expanding to three trees with three depth levels. This approach allowed the ensemble to grow in complexity while enabling evaluation of performance improvements across multiple configurations.

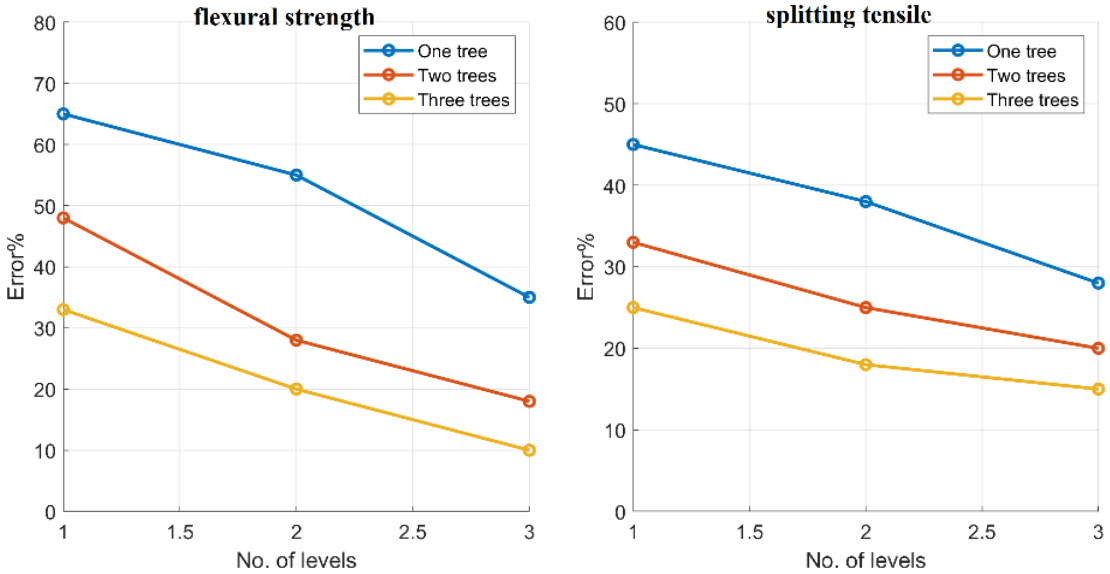

Figure 29. Error (%) reduction trends for Ff and Ft

As depicted in Figure 29, increasing the number of trees and their depth led to a noticeable reduction in prediction error (%) for all target properties. The most substantial improvements occurred in the lower range (from 1 to 2 layers), with diminishing error rates continuing to the third tree level. These results identified the optimal model configuration as three trees with three depth levels, offering the best balance between accuracy and computational efficiency.

https://doi.org/10.31436/iiumej.v26i3.3730

Figure 30 and Figure 31 display the regression results, illustrating how the predicted values align with the experimental data in the training and validation sets for all evaluated strength properties. For Ff, the model achieved (R^2 training = 0.95 and R^2 validation = 0.93), with regression slopes very close to unity ($y = 1.02x$ and $y = 0.98x$), and for Ft, where R^2 ranged from 0.91 to 0.94, confirming the model's capacity to capture the underlying trends without excessive overfitting.

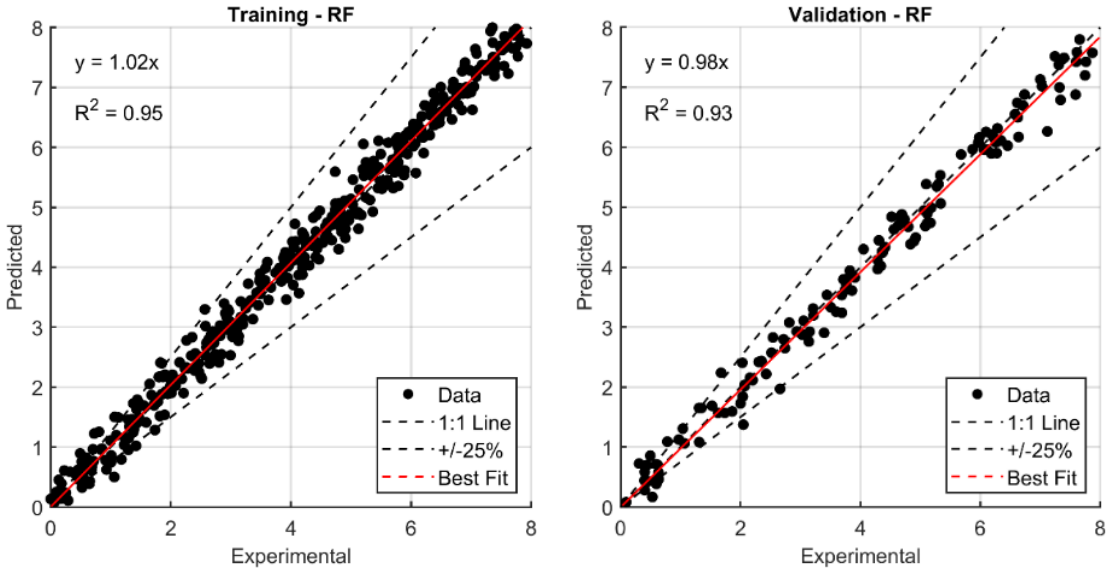

Figure 30. Regression plots for Ff, comparing predicted and experimental values using the optimized Random Forest model

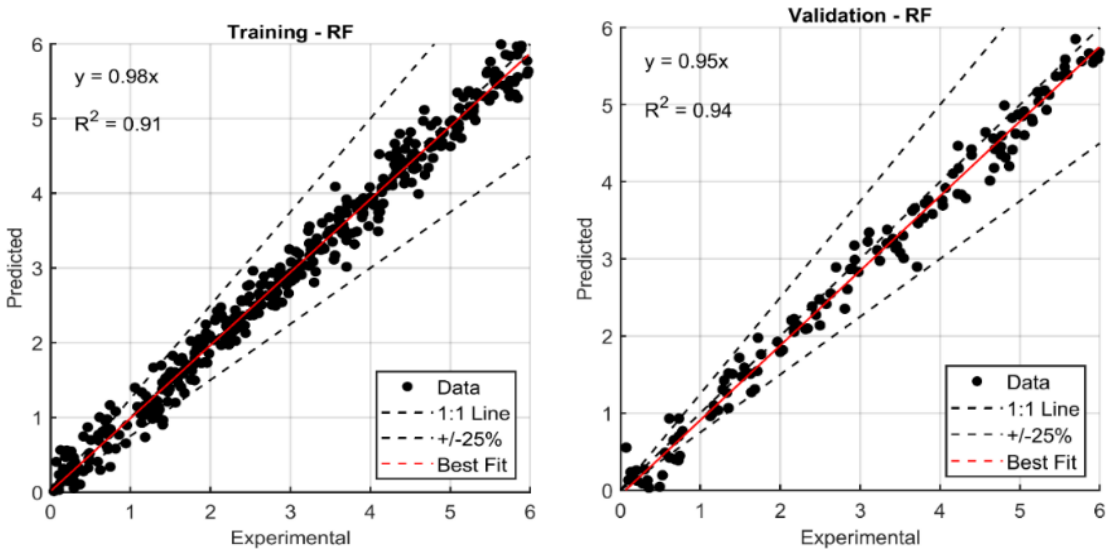

Figure 31. Regression plots for Ft comparing predicted and experimental values using the optimized Random Forest model

Although slightly less accurate than K-NN and SVM models, the RF model offers the advantage of ensemble averaging, which mitigates overfitting and improves robustness against noisy data. Furthermore, RF allows feature importance extraction, making it a valuable tool for identifying the most influential parameters in SCGC mix design optimization.

The SVM and K-NN models emerged as the most accurate and consistent approaches for predicting Ff and Ft. Both models achieved outstanding results, with R^2 values of 0.99 and

RMSE as low as 0.11 MPa, and prediction accuracy reaching 96% on the validation datasets. Figure 32 presents the Ff model performance, where SVM and K-NN clearly outperformed others: the Naïve Bayes (NB) model, by contrast, showed the weakest results with R^2 of 0.76, RMSE of 0.76 MPa, and only 74% accuracy. Similarly, Figure 33 shows the evaluation for splitting tensile strength (Ft), where SVM and K-NN again led the ranking with R^2 values of 0.99, RMSE values of 0.10 MPa, and validation accuracies of 96% and 95%, respectively. While CN2 also provided competitive performance with Ff and Ft prediction accuracies of 95% and 91%, and DT models offered R^2 values of 0.96 and 0.97 for Ff and Ft, respectively, SVM and K-NN offered the best combination of predictive power, generalization, and low error metrics. These findings affirm their suitability for advanced modeling of SCGC mechanical behavior and support their use in sustainable concrete design optimization.

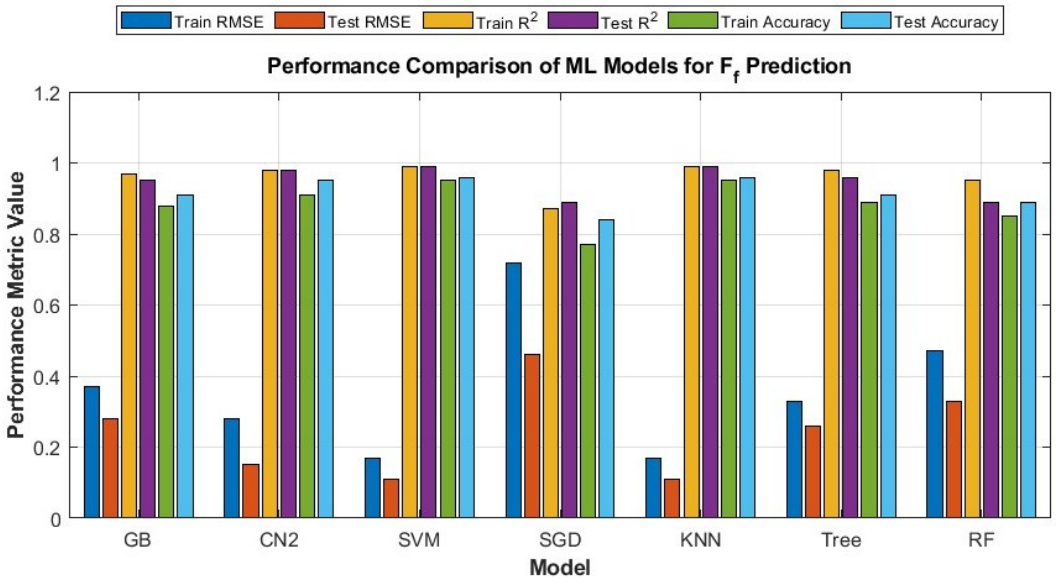

Figure 32. Comparison of machine learning model performance in predicting Ff of SCGC

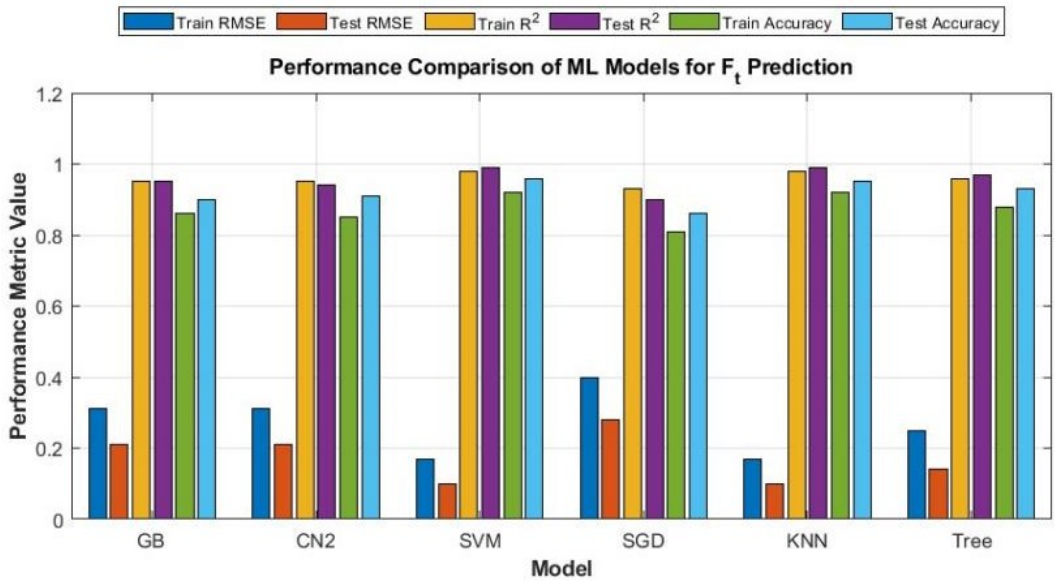

Figure 33. Performance evaluation of developed machine learning models for predicting splitting tensile strength (F_t) of SCGC

https://doi.org/10.31436/iiumej.v26i3.3730

3.1.8. Learning Curve Analysis for Model Generalization

To strengthen the reliability of the reported predictive performance and address the reviewer's request, learning curve analyses were conducted for both Ff and Ft prediction tasks. These analyses assess how the model's predictive accuracy evolves with increasing training data size, using a fixed 75/25 data split. The models evaluated were SVM with KNN. At each analysis step, the training set was incrementally increased from approximately 30 to 408 samples. Both models were retrained from scratch at each training size, and the coefficient of determination (R^2) was calculated on the fixed test set to measure generalization performance.

As shown in Figure 34, the SVM model for Ff prediction achieved high accuracy quickly, reaching an R^2 above 0.99 with only around 100 training samples. Its performance remained consistent across all training sizes, indicating strong generalization. Meanwhile, the KNN model exhibited a more gradual improvement, with R^2 increasing from approximately 0.85 to 0.93 as the training size expanded. This reflects KNN's dependency on a sufficiently large dataset to optimize local predictions.

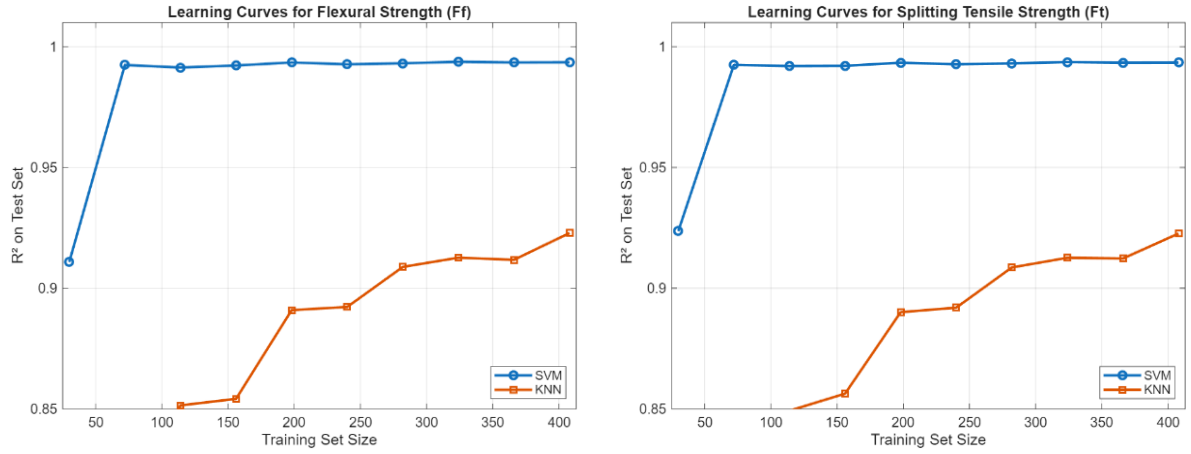

Figure 34. Learning curves of SVM and KNN models for predicting Ff and Ft

A similar trend was observed in the splitting tensile strength (Ft) prediction. The SVM model again reached a stable R^2 of nearly 0.99 early on, while the KNN model improved progressively with additional data, ultimately achieving a comparable R^2 to that of the Ff case. The consistent growth of both curves and the absence of abrupt fluctuations confirm that both models benefit from increased data volume and maintain good generalization without signs of overfitting.

These results validate the earlier reported model performance and demonstrate that the high R^2 values (approaching 0.99) are reproducible under a held-out test condition. The learning curve analysis provides clear visual evidence of the robustness of the SVM model and the reliability of KNN when supported by sufficient training data.

3.2. Response Surface Methodology (RSM) Analysis, Diagnostic Validation, and Hoffman and Gardener Sensitivity Analysis

To complement machine learning techniques with interpretable mathematical modeling, RSM was employed to estimate Ff and Ft of SCGC. RSM enables the derivation of second-order polynomial equations that describe the relationships between key input parameters GGBS, FA, NaOH, Na_2SiO_3, and curing age and the target strength properties. The approach also provides a suite of statistical diagnostics and visualization tools that validate model quality and generalizability.

https://doi.org/10.31436/iiumej.v26i3.3730

For Ff, the RSM model achieved an R^2 of 0.9879, with an adjusted R^2 of 0.9796 and a predicted R^2 of 0.9727. The minimal difference between adjusted and predicted values (0.0069) confirms strong internal consistency. The model's Adequate Precision was calculated at 85.318, significantly higher than the minimum threshold of 4, indicating excellent signal strength. Similarly, the RSM model for Ft showed outstanding performance, with an R^2 of 0.9878, adjusted R^2 of 0.9863, and predicted R^2 of 0.9801, resulting in a minimal difference of 0.0062, well within acceptable bounds. Although the Adequate Precision for Ft was not shown graphically, it was confirmed in the regression output to exceed the minimum value, reinforcing model robustness.

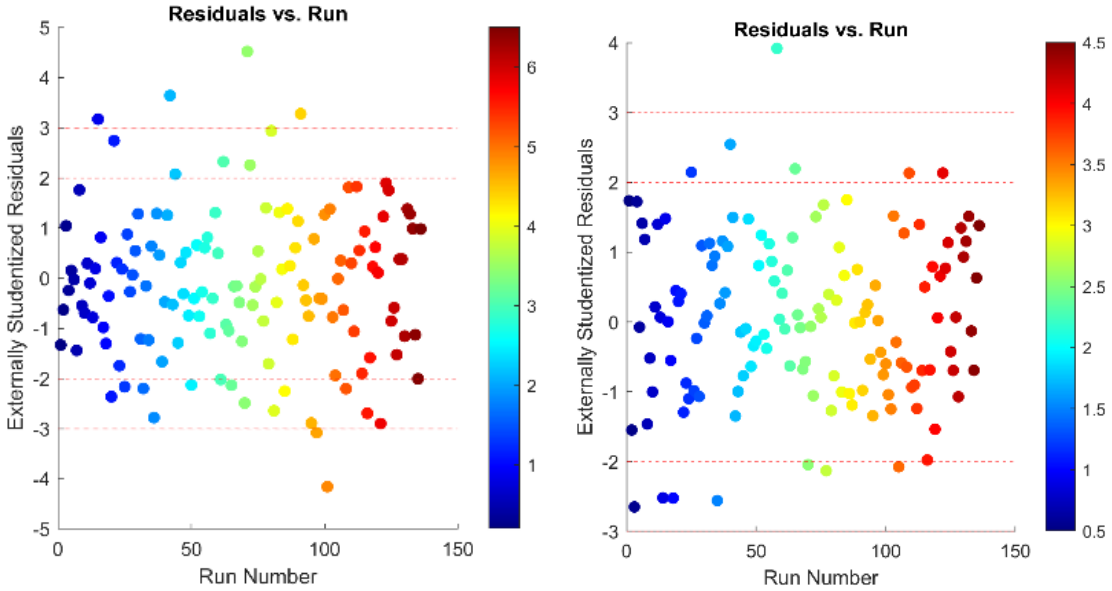

Figure 35. Residuals versus experimental run order for Ff and Ft

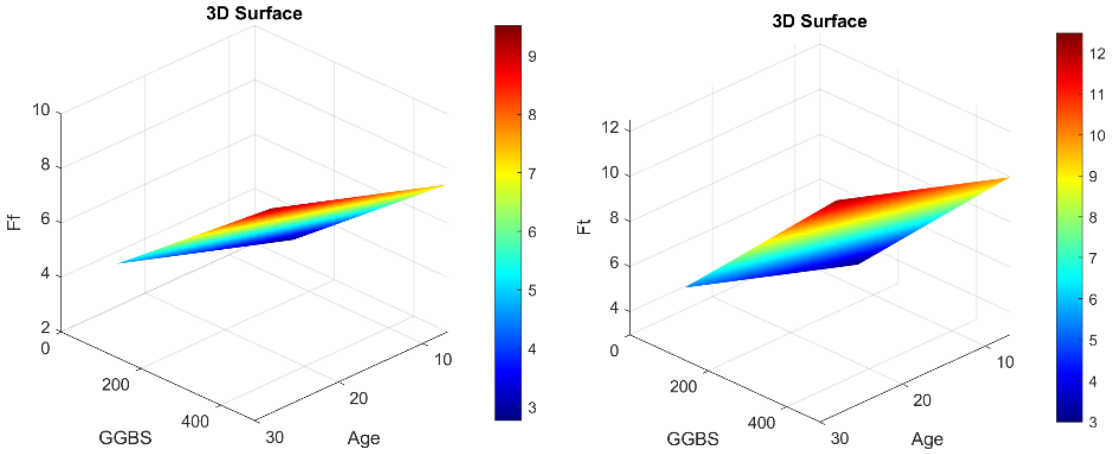

Figure 36. 3D response surface showing the interaction between GGBS and Age on Ff and Ft

The residuals vs. run number plots, Figure 35, show randomly scattered externally studentized residuals, indicating that residuals are independent, homoscedastic, and free from run-order bias in both models. This randomness supports the assumption that model errors are normally distributed and independent of the experimental order. The 3D surface plot of GGBS and Age (Figure 36). illustrates a clear rising plane, confirming their joint positive influence on Ft.

https://doi.org/10.31436/iiumej.v26i3.3730

Regarding influence diagnostics, the Cook's Distance plots for Ff and Ft, as in Figure 37, demonstrate that all data points fall well below the critical threshold. This indicates that no individual observation has a disproportionately large impact on the regression output in either model. The uniformity across these plots affirms the structural stability and fairness of the models.

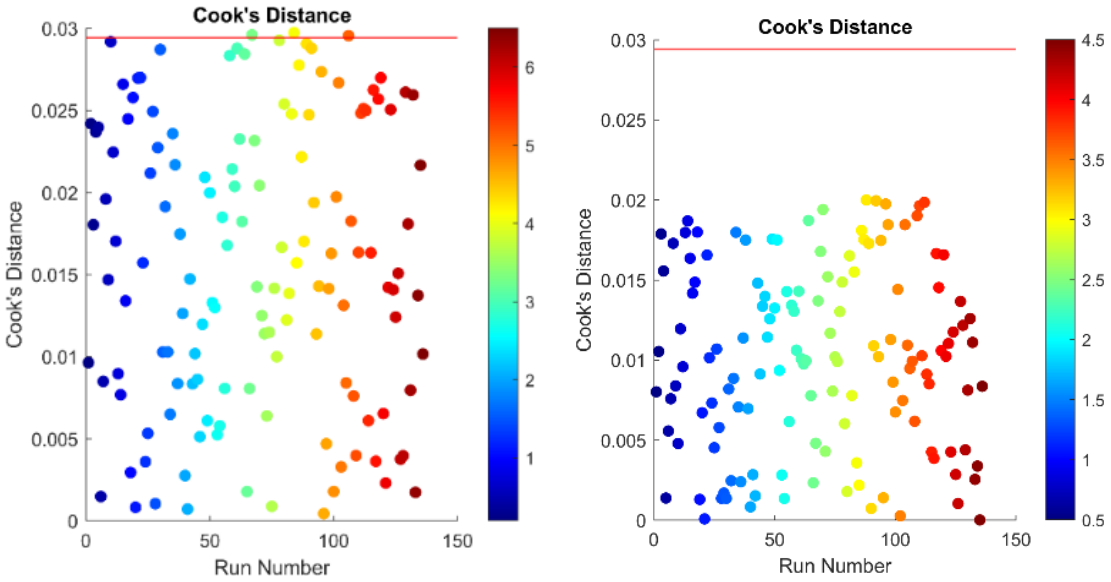

Figure 37. Cook's Distance plot for Ff, Ft model

The predicted vs. actual plots in Figure 38 for both strength models exhibit a tight clustering of points along the ideal 1:1 reference line. This indicates a strong correlation between experimental and predicted values and confirms the models' high predictive capability. These plots show the model's accuracy in replicating real-world outcomes.

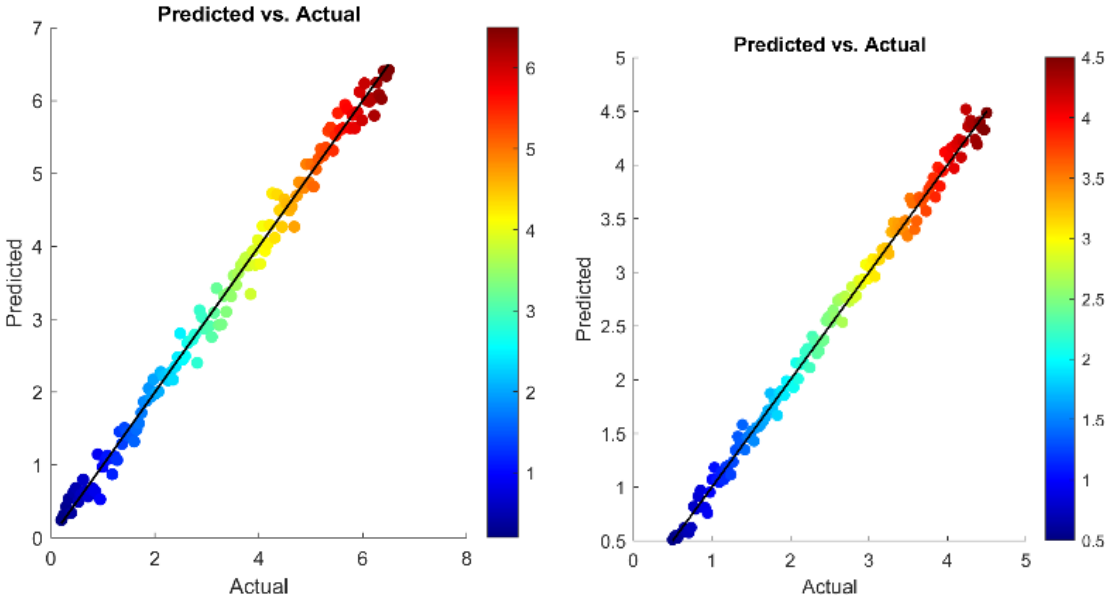

Figure 38. Predicted versus actual flexural strength and tensile strength values

Leverage diagnostics also support model validity. In both the Ff model and the Ft model, as shown in Figure 39, all leverage values remain well within the safe zone, confirming that no single observation has an excessive statistical influence on the fitted equation. The distribution of leverage values suggests that the experimental design is well-balanced.

https://doi.org/10.31436/iiumej.v26i3.3730

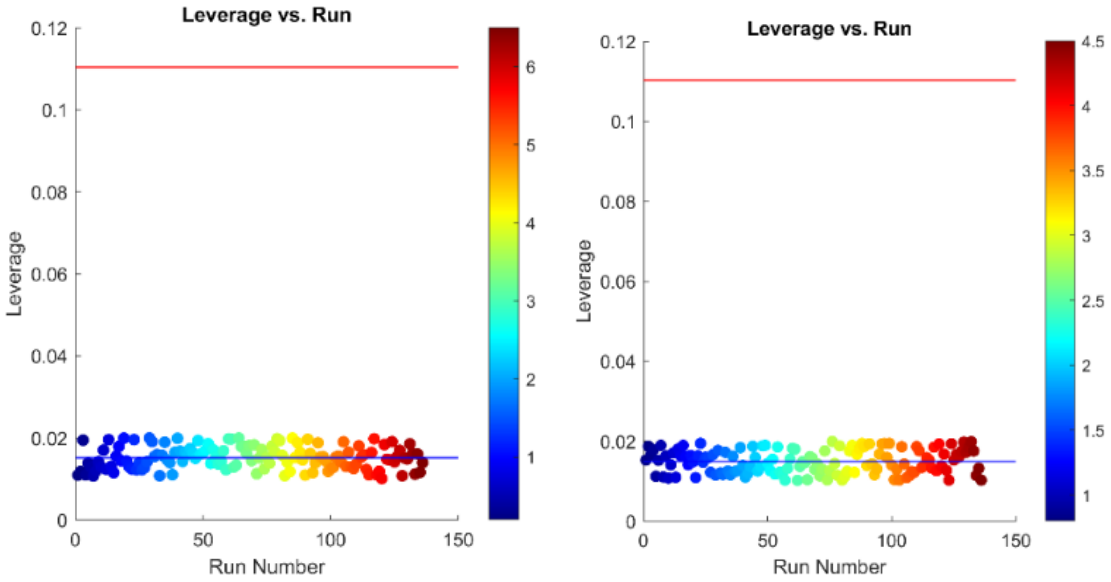

Figure 39. Leverage values plotted against run number

Lastly, the DFBETAS plots for the intercept term, Figure 40, show that the sensitivity of the regression intercept to any specific observation is minimal. All data points fall within the ±0.2 threshold, ensuring that the model coefficients remain stable and unaffected by isolated experimental runs.

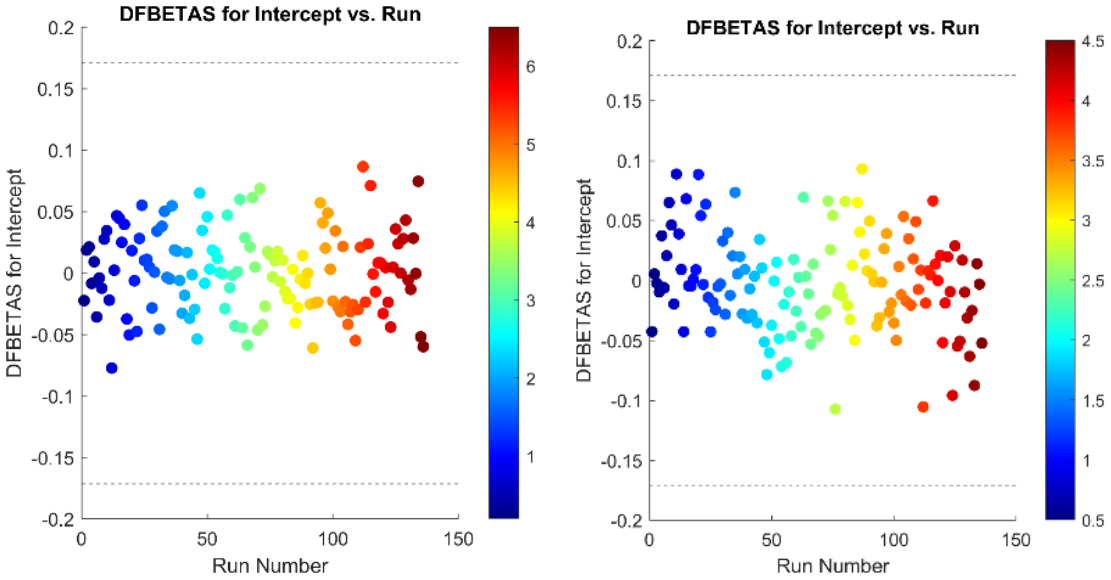

Figure 40. DFBETAS for the intercepting term plotted versus run number

These combined diagnostics confirm that the RSM models for Ff and Ft are statistically sound, structurally stable, and capable of effectively generalizing to unseen data. The consistently high R^2 values, low error distributions, and absence of influential outliers support the conclusion that RSM provides a transparent and reliable framework for modeling the mechanical performance of geopolymer concrete based on compositional and curing parameters.

The equations derived from RSM, expressed using actual factor values, provide practical tools for predicting strength outcomes in SCGC. These closed-form models account for main

https://doi.org/10.31436/iiumej.v26i3.3730

effects, interaction effects, and nonlinear contributions of critical mix parameters presented in Eq. (18) and Eq. (19).

$$\text{Flexural Strength (Ff)} = 5.62342 + 0.014130\,\text{GGBS} + 1.000000\,\text{FA} + 0.016490\,\text{NaOH} + 1.000000\,\text{Na2SiO3} + 0.034523\,\text{Age} + 0.000010\,(\text{GGBS}\cdot\text{FA}) + 0.000060\,(\text{GGBS}\cdot\text{NaOH}) + 1.000000\,(\text{GGBS}\cdot\text{Na2SiO3}) + 0.000024\,(\text{GGBS}\cdot\text{Age}) + 1.000000\,(\text{FA}\cdot\text{NaOH}) + 1.000000\,(\text{FA}\cdot\text{Na2SiO3}) + 1.000000\,(\text{FA}\cdot\text{Age}) + 0.000912\,(\text{NaOH}\cdot\text{Na2SiO3}) + 0.000019\,(\text{NaOH}\cdot\text{Age}) + 1.000000\,(\text{Na2SiO3}\cdot\text{Age}) + (\text{GGBS})^2 + (\text{FA})^2 + (\text{NaOH})^2 + (\text{Na2SiO3})^3 + (\text{Age})^2 \quad (18)$$

$$\text{Tensile strength Ft} = 4.86932 + 0.009817\,\text{GGBS} + 1.000000\,\text{FA} + 0.016063\,\text{NaOH} + 1.000000\,\text{Na2SiO3} + 0.017814\,\text{Age} + 3.07324 \times 10 - 6\,(\text{GGBS}\cdot\text{FA}) + 0.000050\,(\text{GGBS}\cdot\text{NaOH}) + 1.000000\,(\text{GGBS}\cdot\text{Na2SiO3}) + 0.000037\,(\text{GGBS}\cdot\text{Age}) + 1.000000\,(\text{FA}\cdot\text{NaOH}) + 1.000000\,(\text{FA}\cdot\text{Na2SiO3}) + 1.000000\,(\text{FA}\cdot\text{Age}) + 0.000847\,(\text{NaOH}\cdot\text{Na2SiO3}) + 0.000026\,(\text{NaOH}\cdot\text{Age}) + 1.000000\,(\text{Na2SiO3}\cdot\text{Age}) + (\text{GGBS})^2 + (\text{FA})^2 + (\text{NaOH})^2 + (\text{Na2SiO3})^3 + 0.000468\,(\text{Age})^2 \quad (19)$$

Figure 41 illustrates the Hoffman analysis conducted for the two primary strength parameters, Ff and Ft. The study provides a quantitative breakdown of the contribution of each input factor, FA, GGBS, NaOH, Na_2SiO_3, and curing Age, to the variation observed in the model outputs.

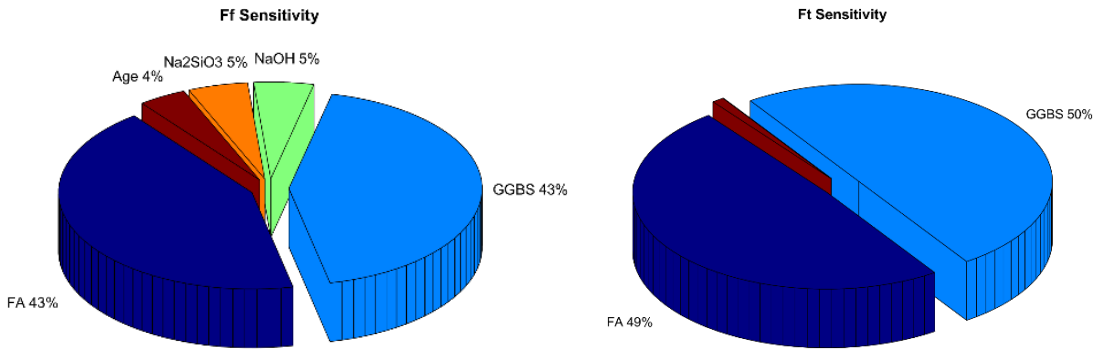

Figure 41. Hoffman and Gardner's sensitivity analysis

Binder constituents had the most significant influence on flexural and tensile strength development in SCGC mixtures. Specifically, fly ash and GGBS displayed equally dominant effects on flexural strength, highlighting their critical role in load-bearing resistance. In contrast, alkaline activators such as sodium hydroxide and sodium silicate contributed only marginally, with curing age also exerting a minor impact. When examining tensile strength, the dominance of binder materials became even more pronounced, with fly ash and GGBS accounting for nearly the entire influence. At the same time, the remaining variables showed virtually no contribution. This trend illustrates that achieving optimal strength outcomes in geopolymer concrete relies heavily on the proportioning and quality of binder materials. The negligible impact of other mix parameters suggests that efforts to refine strength performance can be streamlined by prioritizing binder optimization, reducing complexity in the design process, and offering more consistent control over mechanical outcomes in sustainable concrete applications.

https://doi.org/10.31436/iiumej.v26i3.3730

4. LIMITATION OF THIS STUDY

One major constraint lies in the dataset size and composition. Although 544 experimental records were compiled from reliable sources, the data originated from different laboratories with varying testing protocols, material purities, and environmental conditions. Such variability may introduce hidden biases and affect model training and validation consistency. Moreover, the study focused solely on predicting two mechanical properties, Ff and Ft, without addressing other essential performance metrics such as modulus of elasticity, shrinkage, creep, or long-term durability indicators. These properties are equally important in practical applications and warrant inclusion in future research to establish a more holistic prediction framework.

Additionally, while ML models like K-NN and SVM offered excellent predictive accuracy, they remain limited in interpretability, especially for stakeholders unfamiliar with data science. The "black-box" nature of these models can make extracting clear decision rules for field applications challenging. Although Response Surface Methodology (RSM) improved model transparency, its predictive capability is confined to the parameter range used during model fitting, making extrapolation beyond the tested domain unreliable. Furthermore, the sensitivity analysis adopted in this study primarily captured linear relationships, which may overlook complex, higher-order interactions among input variables. To strengthen the applicability of such models in real-world construction scenarios, future work should integrate larger, more diverse datasets, explore broader mechanical and durability parameters, and adopt advanced interpretability tools or global sensitivity analysis techniques.

5. CONCLUSION

This study provides a robust and comprehensive framework for predicting the flexural strength (Ff) and splitting tensile strength (Ft) of Self-Compacting Geopolymer Concrete (SCGC) using a dual approach that integrates machine learning (ML) techniques and Response Surface Methodology (RSM). Addressing a notable gap in the current literature, where compressive strength has been extensively studied while other mechanical properties like Ff and Ft remain underexplored, this research contributes to the growing domain of data-driven concrete modeling aimed at sustainable construction solutions.

A diverse suite of eight machine learning algorithms, including K-Nearest Neighbors (K-NN), Support Vector Machines (SVM), CN2 rule induction, Random Forest (RF), Gradient Boosting (GB), Decision Trees (DT), Naïve Bayes (NB), and Stochastic Gradient Descent (SGD), was evaluated. Among these, K-NN and SVM exhibited the highest predictive accuracy, with R^2 values reaching 0.99 and root mean squared errors (RMSE) as low as 0.10 for both Ff and Ft. These results confirm their suitability for modeling the nonlinear and multivariate relationships characteristic of geopolymer concrete mix behavior. While CN2, DT, GB, and RF models also showed strong performance ($R^2 \geq 0.93$), the Naïve Bayes model underperformed, underscoring the need for algorithms to accommodate complex feature interdependencies.

Complementing the machine learning analysis, RSM provided interpretable polynomial equations capable of estimating strength outcomes based on actual mix parameters. These models achieved high statistical validity, with R^2 values exceeding 0.98 and strong agreement between adjusted and predicted R^2 scores, indicating excellent model generalization. The models passed rigorous diagnostics, including leverage, Cook's Distance, and DFBETAS, confirming the structural integrity of the regression outputs.

Furthermore, Hoffman and Gardner's sensitivity analysis revealed that binder materials, Fly Ash (FA) and Ground Granulated Blast Furnace Slag (GGBS), are the most influential factors for Ff and Ft, collectively contributing over 90% to strength variation. In contrast, sodium-based activators (NaOH, Na_2SiO_3) and curing age had minimal impact, simplifying the optimization process by focusing design efforts on binder proportions.

Finally, integrating advanced ML algorithms with interpretable RSM models offers a powerful toolkit for predicting and optimizing SCGC mechanical performance. This hybrid framework enhances the accuracy and reliability of strength estimation, improving practical decision-making in mix design for sustainable concrete technologies. Future studies may build upon this foundation by incorporating additional performance parameters such as durability, shrinkage, and creep, as well as deploying these models within digital twin platforms for real-time monitoring and adaptive construction management.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by Universitas Muhammadiyah Yogyakarta (UMY).

REFERENCES

[1] Heshmati, M., Sheikh, M. N., & Hadi, M. N. S. (2025). A comprehensive review of the fresh and hardened characteristics of self-compacting geopolymer concrete. Journal of Sustainable Cement-Based Materials, 14(4), 816–835. https://doi.org/10.1080/21650373.2025.2461182

[2] Gopalakrishna, B., & Pasla, D. (2024). Study of Ambient Cured Fly Ash-GGBS-Metakaolin-Based Geopolymers Mortar. Lecture Notes in Civil Engineering, 440. https://doi.org/10.1007/978-981-99-7464-1_3

[3] Shahrokh, S., Armaghan, S., & Danandeh, A. M. (2025). Service Life Prediction and Life Cycle Assessment of Pozzolanic Concretes. 54(4), 29–36. 10.22034/ceej.2024.61027.2338

[4] Inqiad, W. Bin, Siddique, M. S., Alarifi, S. S., Butt, M. J., Najeh, T., & Gamil, Y. (2023). Comparative analysis of various machine learning algorithms to predict 28-day compressive strength of Self-compacting concrete. Heliyon, 9(11). https://doi.org/10.1016/j.heliyon.2023.e22036

[5] Sukontasukkul, P., Intarabut, D., Phoo-ngernkham, T., Suksiripattanapong, C., Zhang, H., & Chindaprasirt, P. (2023). Self-compacting steel fibers reinforced geopolymer: Study on mechanical properties and durability against acid and chloride attacks. Case Studies in Construction Materials, 19. https://doi.org/10.1016/j.cscm.2023.e02298

[6] Sun, C., Wang, K., Liu, Q., Wang, P., & Pan, F. (2023). Machine-Learning-Based Comprehensive Properties Prediction and Mixture Design Optimization of Ultra-High-Performance Concrete. Sustainability (Switzerland), 15(21). https://doi.org/10.3390/su152115338

[7] Ahmad, A., Ahmad, W., Chaiyasarn, K., Ostrowski, K. A., Aslam, F., Zajdel, P., & Joyklad, P. (2021). Prediction of geopolymer concrete compressive strength using novel machine learning algorithms. Polymers, 13(19). https://doi.org/10.3390/polym13193389

[8] Saha, P., Sapkota, S. C., Das, S., & Kwatra, N. (2024). Prediction of fresh and hardened properties of self-compacting concrete using ensemble soft learning techniques. Multiscale and Multidisciplinary Modeling Experiments and Design, 7(5), 4923–4945. https://doi.org/10.1007/s41939-024-00423-5

[9] de-Prado-Gil, J., Palencia, C., Silva-Monteiro, N., & Martínez-García, R. (2022). To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models. Case Studies in Construction Materials, 16. https://doi.org/10.1016/j.cscm.2022.e01046

[10] Biswas, R., Kumar, M., Singh, R. K., Alzara, M., El Sayed, S. B. A., Abdelmongy, M., Yosri, A. M., & Yousef, S. E. A. S. (2023). A novel integrated approach of RUNge Kutta optimizer and ANN for estimating compressive strength of self-compacting concrete. Case Studies in Construction Materials, 18. https://doi.org/10.1016/j.cscm.2023.e02163

[11] Dong, B., Zhao, S., Zhang, Y., Zhang, Y., Yu, Y., & Yang, J. (2025). Physics-based probabilistic analysis of corrosion initiation in alkali-activated slag concrete assisted by machine learning. Construction and Building Materials, 471. https://doi.org/10.1016/j.conbuildmat.2025.140661

[12] Iqbal, M., Nazar, S., Yang, J., & Mahmoud, H. A. (2025). Integrated machine learning and response surface methodology for comprehensive rheological characterization of low-carbon binders. Case Studies in Construction Materials, 22. https://doi.org/10.1016/j.cscm.2025.e04475

[13] Parhi, S. K., Dwibedy, S., & Panigrahi, S. K. (2024). AI-driven critical parameter optimization of sustainable self-compacting geopolymer concrete. Journal of Building Engineering, 86. https://doi.org/10.1016/j.jobe.2024.108923

[14] Aneja, S., Sharma, A., Gupta, R., & Yoo, D. Y. (2021). Bayesian regularized artificial neural network model to predict strength characteristics of fly-ash and bottom-ash based geopolymer concrete. Materials, 14(7). https://doi.org/10.3390/ma14071729

[15] Jiang, H., Liu, G., Alyami, H., Alharbi, A., Jameel, M., & Khadimallah, M. A. (2022). Intelligence decision mechanism for prediction of compressive strength of self-compaction green concrete via neural network. Journal of Cleaner Production, 340. https://doi.org/10.1016/j.jclepro.2022.130580

[16] Mazumder, E. A., & Prasad Meesaraganda, L. V. (2023). Probabilistic Estimation for Mechanical Properties of Self-Compacting Geopolymer Concrete Using Machine Learning Technique. Arabian Journal for Science and Engineering, 48(10). https://doi.org/10.1007/s13369-023-07866-x

[17] Aliabdo, A. A., Abd Elmoaty, A. E. M., & Salem, H. A. (2016). Effect of water addition, plasticizer and alkaline solution constitution on fly ash based geopolymer concrete performance. Construction and Building Materials, 121, 694–703. https://doi.org/10.1016/J.CONBUILDMAT.2016.06.062

[18] Phoo-Ngernkham, T., Phiangphimai, C., Damrongwiriyanupap, N., Hanjitsuwan, S., Thumrongvut, J., & Chindaprasirt, P. (2018). A Mix Design Procedure for Alkali-Activated High-Calcium Fly Ash Concrete Cured at Ambient Temperature. Advances in Materials Science and Engineering, 2018. https://doi.org/10.1155/2018/2460403

[19] Arun, B. R., Srishaila, J. M., S, M. K., Algur, V., K, K., & Tanu, H. (2024). Prediction of machine learning application in the development of novel sustainable self-compacting geopolymer concrete. Research on Engineering Structures and Materials, x(xxxx), 1–22. https://doi.org/10.17515/resm2024.354ma0716rs

[20] Lavanya, G., & Jegan, J. (2015). Durability Study on High Calcium Fly Ash Based Geopolymer Concrete. Advances in Materials Science and Engineering, 2015. https://doi.org/10.1155/2015/731056

[21] Deilami, S., Aslani, F., & Elchalakani, M. (2019). An experimental study on the durability and strength of SCC incorporating FA, GGBS and MS. Proceedings of the Institution of Civil Engineers: Structures and Buildings, 172(5). https://doi.org/10.1680/jstbu.17.00129

[22] Rangan, B., & Hardjito, D. (2005). Studies on fly ash-based geopolymer concrete. Proc. 4th World …, November.

[23] Nishanth, L., & Patil, D. N. N. (2022). Experimental evaluation on workability and strength characteristics of self-consolidating geopolymer concrete based on GGBFS, flyash and alccofine. Materials Today: Proceedings, 59. https://doi.org/10.1016/j.matpr.2021.10.200

[24] Vijaya Prasad, B., Paul Daniel, A. P., Anand, N., & Yadav, S. K. (2022). Strength and microstructure behaviour of high calcium fly ash based sustainable geo polymer concrete.

Journal of Engineering, Design and Technology, 20(2). https://doi.org/10.1108/JEDT-03-2021-0178

[25] Alam, M. F., Shubham, K., Kumar, S., & Srivastava, A. K. L. (2024). Enhancing high-strength self-compacting concrete properties through Nano-silica: analysis and prediction of mechanical strengths. Journal of Building Pathology and Rehabilitation, 9(1). https://doi.org/10.1007/s41024-024-00386-7

[26] Tho-In, T., Sata, V., Chindaprasirt, P., & Jaturapitakkul, C. (2012). Pervious high-calcium fly ash geopolymer concrete. Construction and Building Materials, 30. https://doi.org/10.1016/j.conbuildmat.2011.12.028

[27] Bheel, N., Awoyera, P., Tafsirojjaman, T., Hamah Sor, N., & sohu, S. (2021). Synergic effect of metakaolin and groundnut shell ash on the behavior of fly ash-based self-compacting geopolymer concrete. Construction and Building Materials, 311. https://doi.org/10.1016/j.conbuildmat.2021.125327

[28] Pirhadi, N., Tang, X., Yang, Q., & Kang, F. (2019). A new equation to evaluate liquefaction triggering using the response surface method and parametric sensitivity analysis. Sustainability (Switzerland), 11(1). https://doi.org/10.3390/su11010112

[29] Fontanari, T., Fróes, T. C., & Recamonde-Mendoza, M. (2022). Cross-validation Strategies for Balanced and Imbalanced Datasets. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13653 LNAI. https://doi.org/10.1007/978-3-031-21686-2_43

[30] Li, Z., Pei, T., Ying, W., Srubar III, W. V., Zhang, R., Yoon, J., Ye, H., Dabo, I., & Radlińska, A. (2024). Simulation-Based Transfer Learning for Concrete Strength Prediction. In Rilem Bookseries (Vol. 48). https://doi.org/10.1007/978-3-031-53389-1_98

[31] Shaikhina, T., & Khovanova, N. A. (2017). Handling limited datasets with neural networks in medical applications: A small-data approach. Artificial Intelligence in Medicine, 75, 51–63. https://doi.org/10.1016/j.artmed.2016.12.003

[32] Arthi, K., Sankaradass, V., Parveen, N., & Muralidharan, J. (2023). Methods of cross-validation and bootstrapping. In Toward Artificial General Intelligence: Deep Learning, Neural Networks, Generative AI. https://doi.org/10.1515/9783111323749-007

[33] Hong, J.-S., Lee, J., & Sim, M. K. (2024). Concise rule induction algorithm based on one-sided maximum decision tree approach. Expert Systems with Applications, 237. https://doi.org/10.1016/j.eswa.2023.121365

[34] Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Computing, 25(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6

[35] Jin, Z., Hei, T. G., & Jin, K. (2024). Application of support vector machines for categorizing biological and medical data. Proceedings of SPIE - The International Society for Optical Engineering, 13105. https://doi.org/10.1117/12.3026871

[36] Saha, A., Yadav, R. K., & Dwivedi, R. K. (2024). Efficient Convergence Analysis of Stochastic Gradient Descent for Large-Scale Optimization. 2024 International Conference on Optimization Computing and Wireless Communication, ICOCWC 2024. https://doi.org/10.1109/ICOCWC60930.2024.10470675

[37] Liang, C., Xia, S., & Chen, Z. (2019). Improvement k-Nearest Neighbor Classification Algorithm Based on Reference Points. Jisuanji Gongcheng/Computer Engineering, 45(2). https://doi.org/10.19678/j.issn.1000-3428.0049901

[38] Rutkowski, L., Jaworski, M., & Duda, P. (2020). Decision Trees in Data Stream Mining. In Studies in Big Data (Vol. 56). https://doi.org/10.1007/978-3-030-13962-9_3

[39] Tran, T. T., Phan, N. Q., & Huynh, H. X. (2025). Random Forest Model Parameters Optimization. In Communications in Computer and Information Science: Vol. 2191 CCIS. https://doi.org/10.1007/978-981-97-9616-8_19

[40] Podder, S., & Mukherjee, S. (2024). RESPONSE SURFACE METHODOLOGY (RSM) AS A TOOL IN PHARMACEUTICAL FORMULATION DEVELOPMENT. Asian Journal of Pharmaceutical and Clinical Research, 17(11), 18–25. https://doi.org/10.22159/ajpcr.2024v17i11.52149

[41] Mahmoodi, V., & Sargolzaei, J. (2014). Optimization of photocatalytic degradation of naphthalene using nano-TiO2/UV system: statistical analysis by a response surface methodology. Desalination and Water Treatment, 52(34–36). https://doi.org/10.1080/19443994.2013.861774

[42] Onyelowe, K. C., Kamchoom, V., Hanandeh, S., Ebid, A. M., Llamuca Llamuca, J. L., Cayán Martínez, J. C., Rose, E., Awoyera, P., & Avudaiappan, S. (2025). Predicting the strengths of basalt fiber reinforced concrete mixed with fly ash using AML and Hoffman and Gardener techniques. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-96420-w

[43] Paleari, L., Movedi, E., Zoli, M., Burato, A., Cecconi, I., Errahouly, J., Pecollo, E., Sorvillo, C., & Confalonieri, R. (2021). Sensitivity analysis using Morris: Just screening or an effective ranking method? Ecological Modelling, 455. https://doi.org/10.1016/j.ecolmodel.2021.109648

[44] Hoffman, F. O., & Gardner, R. H. (1983). Evaluation of uncertainties in radiological assessment models. Radiological Assessment: A Textbook on Environmental Dose Analysis.

[45] Wu, Y., & Liu, S. (2014). A suggestion for computing objective function in model calibration. Ecological Informatics, 24. https://doi.org/10.1016/j.ecoinf.2014.08.002

[46] Robeson, S. M., & Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. Plos One, 18(2 February). https://doi.org/10.1371/journal.pone.0279774

[47] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7. https://doi.org/10.7717/PEERJ-CS.623

[48] Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1). https://doi.org/10.3354/cr030079

[49] Turney, S. (2022). Coefficient of determination (R2): Calculation and interpretation. Scribbr.