



Journal of Materials and Engineering Structures

Research Paper

Prediction and Feature Analysis of Self-Compacting Concrete Strength Using Machine Learning with Recycled Coarse Aggregate and Supplementary Cementitious Materials

Kavendra Pulkit ^{a,*}, Shivi Nigam ^b

^a Assistant Manager, Bureau Veritas India Pvt Limited, Noida, Uttar Pradesh, India 201301

^b Engineer, Bureau Veritas India Pvt limited, Noida, Uttar Pradesh, India 201301

ARTICLE INFO

Article history -

Received : 6 July 2025

Revised : 2nd September 2025

Accepted : 5 September 2025

Keywords:

Machine Learning

Gradient Boosting

K Fold Cross Validation

Hyperparameter Tuning

ABSTRACT

This study presents a machine learning based approach to predict the CS of SCC incorporating RCA and supplementary cementitious materials (SCMs), aimed at promoting sustainable construction practices. A dataset comprising 337 SCC mixes with varying input parameters such as cement (40–635 kg/m³), SCMs (0–592 kg/m³), water to binder ratio (0.25–0.45), and curing age (7–120 days) was utilized. Five regression models were evaluated- Extra Trees, AdaBoost, SVR Gradient Boosting, and ANN. Among them, Gradient Boosting achieved the highest accuracy with $R^2 = 0.8876$, RMSE = 4.032 MPa, and MAE=2.6382 MPa. Extra Trees followed closely with $R^2=0.8309$ and RMSE=4.9444 MPa. SHAP interpretability analysis revealed that cement content and curing age were the most influential parameters in predicting CS. The study confirms the effectiveness of ML models in replacing time consuming lab tests and provides an interpretable, data driven framework for optimizing concrete mix design using industrial waste materials.

1 Introduction

The increasing volume of construction and demolition (C&D) waste poses a significant global sustainability issue, as it makes up almost half of the world's total solid waste and plays a major role in environmental harm and health risks [1]. Sustainable practices like material reuse and recycling have become more popular as they help lessen environmental impacts by promoting efficient resource recovery. The integration of circular economy principles focused on sustainable production and consumption has demonstrated potential in reducing waste and improving economic resilience [2]. Furthermore, the development of policies and legislative support is essential, as financial incentives and regulations can encourage waste

* Corresponding author.

E-mail address: kavendra_61900115@nitkkr.ac.in

minimization [3]. Additionally, raising awareness and educating stakeholders can enhance the effectiveness of implementation outcomes [4]. Even with these advancements, practical issues such as regulatory inconsistencies and local economic limitations need to be tackled to achieve sustainable and efficient management of construction and demolition waste worldwide.

Cement production has a considerable environmental impact, mainly because of its high energy use and significant carbon dioxide (CO₂) emissions. The production of cement accounts for about 5% of global industrial energy consumption, with emissions varying between 0.65 and 0.92 tons of CO₂ for each ton of cement produced [5]. This process contributes to greenhouse gas emissions and results in ecological degradation due to resource extraction and pollution.

The cement industry is one of the largest sources of global greenhouse gas emissions because of its energy-intensive manufacturing process. This is particularly evident during the kiln and calcination stages, where calcium carbonate breaks down into calcium oxide, resulting in substantial CO₂. This process produces nitrogen oxides (NO_x) and sulfur dioxide (SO₂), which worsen environmental and health issues [6]. In response, strategies for mitigation such as carbon capture technologies, the use of low carbon materials, and the adoption of blended cements that include materials like limestone have been developed, which could reduce carbon footprints by as much as 6.4% [7]. Furthermore, policies that promote circular economy practices and encourage environmentally friendly production methods are crucial for minimizing the environmental impact of cement manufacturing [8]. Supplementary Cementitious Materials (SCMs), which come from industrial by-products such as fly ash, silica fume, and rice husk ash, provide a practical method to decrease reliance on cement and the emissions linked to it [9]. The incorporation of SCMs aids in waste management by repurposing materials that would otherwise be discarded [10] and improves the strength and durability of concrete, with studies indicating optimal performance at a 30% replacement of fly ash [11]. Additionally, SCMs enhance workability, facilitating concrete placement [12]. However, issues like quality variability and restricted market supply continue to pose challenges to their broader use [13]. SCC represents an important advancement in the field of sustainable construction. SCC is a flowable, non-segregating mixture that effectively fills formwork and surrounds reinforcement without the need for vibration, providing good workability in intricate or crowded reinforcement regions [14]. The fluid nature, along with the use of superplasticizers and mineral admixtures such as fly ash and slag, improves durability and surface quality, while also decreasing labor, noise, and health risks [15]. Furthermore, the compatibility of SCC with industrial by-products plays a role in promoting environmental sustainability [16]. Although precise mix design and quality control are necessary, SCC is being used more frequently in precast and architectural applications because it can adapt to complex shapes and provide excellent finish and performance.

These advances in SCMs and SCC highlight progress in sustainable concrete technology. However, previous works mostly studied SCMs or SCC in isolation, and limited attention has been given to their combined use with recycled aggregates in predictive modeling frameworks. This creates an opportunity to evaluate such interactions using modern ML approaches.

Parallel to this, comprehensive reviews have analyzed the integration of recycled plastic waste into concrete, highlighting both its potential in reducing environmental pollution and the mechanical challenges such as reductions in compressive and tensile strength [17]. As sustainable construction practices progress, the combination of SCC and SCM is expected to significantly contribute to minimizing the environmental impact of the concrete industry. Recycled concrete aggregate (RCA), obtained from construction and demolition (C&D) waste, presents a viable method for promoting sustainable construction. It helps decrease the need for natural aggregates and diverts significant quantities of waste from landfills. The mechanical properties of recycled concrete aggregates (RCA) are typically lower than those of natural aggregates. Research indicates a reduction in compressive strength of up to 8.31% when 100% RCA is used [18]. However, the best outcomes are seen with approximately 40% RCA replacement, where the loss of strength is minimal and performance remains stable [19]. Different enhancement methods, including chemical surface modification and mechanical abrasion, have been suggested to improve the bonding between RCA and the cement matrix, which in turn enhances the overall structural performance of the resulting concrete [18]. Additionally, the combination of fine and coarse RCA types has demonstrated better results than using either type alone, thereby optimizing material use [20]. Recent research has also shown the effectiveness of surrogate models in predicting the long-term compressive strength behavior of RAC across varying strength ranges and specimen sizes, demonstrating reliable accuracy when applied to datasets of over 300 mixes [21]. These studies confirm the feasibility of predictive modeling for RAC strength, but they often focus on traditional concretes or small datasets. Very few have examined SCC incorporating both RCA and SCMs, which restricts the broader applicability of their conclusions.

The use of recycled concrete aggregates (RCA) can reduce the carbon footprint of concrete production by as much as 31.3% when 50% of natural aggregates are replaced [22] Although initial treatment costs may be elevated, life cycle analyses

indicate potential overall cost savings [19]. The compressive strength of concrete is an important performance measure that is affected by the mix design, environmental factors, and testing methods. Although traditional core sampling methods for CS testing are valuable, their destructiveness and variability require the application of advanced statistical models to improve accuracy and efficiency [23]. The integration of machine learning (ML) techniques has become a significant method in predicting concrete strength. Models such as Random Forest have reached R^2 values of up to 0.95 with minimal error margins, demonstrating their effectiveness in modeling complex input-output relationships [24]. Advanced soft computing studies have employed XGBoost, MEP, and ANN to predict compressive strength in fly ash-modified RAC, where XGBoost outperformed other approaches using five-fold cross-validation and SHAP-based interpretability [25]. AdaBoost has shown strong performance in predicting the behavior of smart concrete, achieving an accuracy of 97.09% and the lowest reported mean squared error of 1.485 [26]. XGBoost demonstrated strong performance in managing non-linear relationships, achieving a R^2 of 0.91 and an RMSE of 4.37 [27]. In contrast, Gaussian Process Regression (GPR) provided efficient training times and consistent results, with a R^2 of 0.83 [28]. The practical use of machine learning in concrete mix design has been made easier by the creation of user-friendly graphical interfaces and web-based tools, enabling practitioners to quickly obtain predictions based on input parameters [29]. Additionally, synthetic data has been utilized to improve the training of machine learning models, leading to enhancements in both sustainability and accuracy [30]. Despite these advancements, challenges remain, including the necessity for larger, high-quality datasets and the incorporation of new materials such as carbon nanotubes. Collectively, these ML approaches demonstrate strong potential for concrete strength prediction, yet their application remains fragmented. Most studies are confined to specific concrete types, single-waste replacements, or lack interpretability. Soft computing-based approaches on RAC have emphasized the importance of aggregate properties, mix design parameters, and curing conditions, with Non-Linear Regression (NLR) and Multi-Linear Regression (MLR) models showing strong predictive accuracy for compressive strength [31].

Machine learning algorithms, including Support Vector Machines (SVM), Random Forest, and Gradient Boosting, are commonly employed to model the intricate and nonlinear relationships between different input parameters and concrete properties such as compressive strength and modulus of elasticity. Among these, Adaptive Boosting (AdaBoost) has demonstrated exceptional performance, achieving a prediction accuracy of 97.09%, surpassing models like SVM and Random Forest [32]. In a similar manner, research has utilized various machine learning techniques, such as Linear Regression and Decision Trees, to forecast different mechanical properties of concrete, highlighting the versatility and usefulness of these models in various contexts [33]. A primary benefit of machine learning is its capacity to greatly lessen the reliance on extensive physical testing, resulting in significant savings in both time and resources [34]. There is an increasing interest in incorporating deep learning architectures, like neural networks, to enhance prediction capabilities in concrete analysis. Additionally, hybrid methods that integrate models such as ANN with support vector machines (SVMs) are being investigated and have shown enhanced performance. While several prior works have optimized green concrete mixes using regression and ML models with fly ash, SCMs, or RCA individually, most remain limited to conventional concrete or single-waste integration. Few studies have explored the combined influence of RCA and SCMs in SCC using advanced ensemble and neural network approaches, and even fewer have integrated interpretability tools to explain model predictions. To address these gaps, this study (i) compiles a dataset of 337 SCC mixes incorporating both RCA and SCMs, (ii) systematically benchmarks five ML models (Extra Trees, AdaBoost, SVR, Gradient Boosting, ANN) using 10-fold cross-validation, and (iii) applies SHAP and ELI5 interpretability analysis to identify key influencing parameters. The novelty of this work lies in providing a transparent, data-driven predictive framework for sustainable SCC, thereby advancing both scientific understanding and practical applications in green concrete design.

A schematic flowchart (Fig. 1) has been provided to illustrate the overall research methodology. It summarizes the workflow from data collection, preprocessing, model training and hyperparameter tuning, to performance evaluation and interpretability analysis. This study provides a thorough and clear evaluation of various ensemble and neural network models, in contrast to previous research that concentrated only on traditional concrete or single waste materials. It improves transparency by utilizing advanced feature importance techniques like SHAP and ELI5, offering practical insights into material behavior. This work makes a significant contribution to sustainable construction and intelligent concrete mix design by incorporating real-world variability, strong validation strategies, and visual diagnostics. The research methodology started with careful data preprocessing, which included median and mode imputation for missing numerical and categorical values, one hot encoding for categorical variables, and feature scaling using standardization (StandardScaler). Various machine learning models, such as Extra Trees, AdaBoost, Gradient Boosting, Support Vector Regression (SVR), and ANN, were trained on 80% of the data. A thorough validation strategy using 10-fold cross-validation evaluated the generalization ability

of each model, helping to avoid overfitting. The models were assessed thoroughly using evaluation metrics including R^2 , Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) to measure accuracy and error. Hyperparameter tuning was conducted using grid search and manual adjustments specific to each model. For instance, Extra Trees utilized 150 estimators, AdaBoost employed 150 estimators with a learning rate of 0.1, SVR was configured with an RBF kernel, $C=100$, and $\epsilon=0.1$, Gradient Boosting was set with 100 estimators, a learning rate of 0.1, and a maximum depth of 5, while the ANN featured hidden layers of (100,50), ReLU activation, and the Adam solver with a maximum of 2000 iterations. Before training the model, a thorough exploratory data analysis (EDA) was conducted using histograms, pairplots, and Pearson correlation to understand the distributions and relationships of the features. Feature importance analysis utilized Pearson correlation, Mutual Information, SHAP, and ELI5 Permutation Importance to determine key predictors affecting concrete compressive strength. In conclusion, the use of residual plots, SHAP heatmaps, and Regression Error Characteristic (REC) curves improved visualization and interpretability, offering clear and reliable visual insights into model performance and the effects of features.

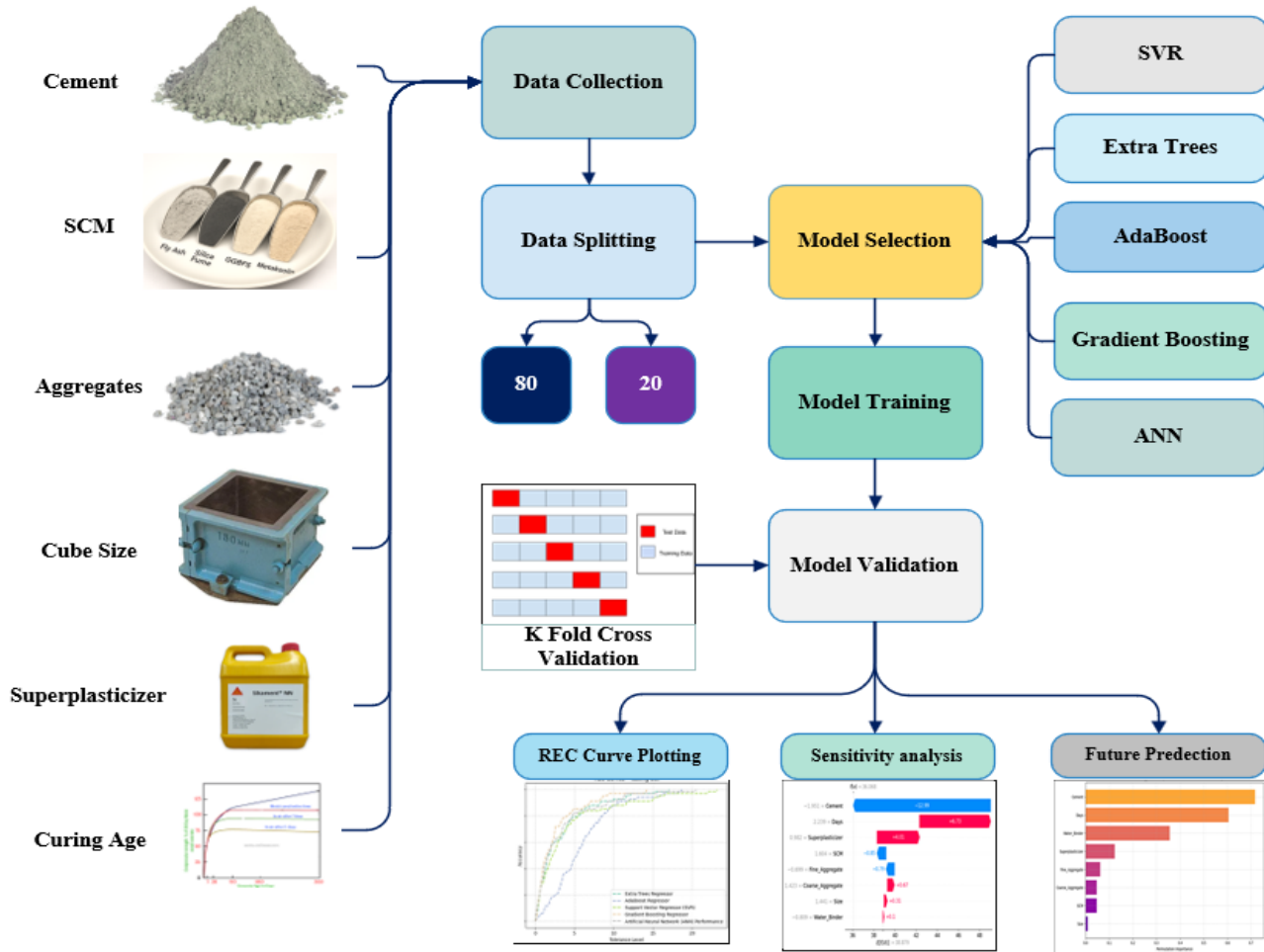


Fig. 1- Research methodology

2 Research methodology

2.1 Data Preprocessing

A dataset containing 337 mix designs of self-compacting concrete (SCC) was compiled from various published experimental studies and used from [35]. This dataset focuses on concrete mixes that incorporate recycled concrete aggregate (RCA) and supplementary cementitious materials (SCMs) like fly ash, silica fume, slag, and other industrial byproducts. The dataset was divided for model development, with 70% allocated for training (236 samples) and 30% for validation (101 samples). The input features include essential mix components such as cement and SCM content (kg/m^3), fine aggregate,

coarse aggregate, superplasticizer (SP), curing age (days), and water to binder ratio (W/B) for RCA. Furthermore, cube size (mm) was incorporated to represent differences in specimen geometry.

The dataset underwent thorough preprocessing, which involved the analysis of statistical descriptors such as skewness and kurtosis. Slight asymmetries were noted in SCM content (skewness = 1.18) and W/B ratio (skewness = −0.92), indicating practical variation in mix designs. Outliers were assessed through z score analysis, with incorrect entries being eliminated, while naturally occurring material differences were kept to maintain the real-world representativeness of the data. No normalization or standardization was applied, as the machine learning algorithms used, such as Extra Trees and Gradient Boosting, are not sensitive to feature scaling. Fig. 2 presents a graphical summary of the data distribution, emphasizing the variety and practical significance of the input parameters utilized in this study.

This study used 337 SCC mix records collected from published work, covering compressive strengths between 10–90 MPa. While the dataset is moderate in size, it is comparable to those used in similar machine learning studies on sustainable concrete and provides a sound basis for model evaluation. To reduce the risk of overfitting in the ANN, three measures were applied: (i) 10-fold cross-validation to test the model on different data subsets, (ii) dropout (10%) and early stopping during training, and (iii) standardization of all input variables using z-score normalization.

Standardization was important because the variables had very different ranges. Without scaling, features with large values, such as cement content, could dominate smaller ones like superplasticizer dosage, leading to biased predictions and misleading feature importance. By rescaling all variables to zero mean and unit variance, each parameter contributed fairly to the model. These steps improved training stability, reduced overfitting, and made the interpretation of feature importance more reliable. However, it is recognized that larger datasets would further improve the robustness and generalization of the models.

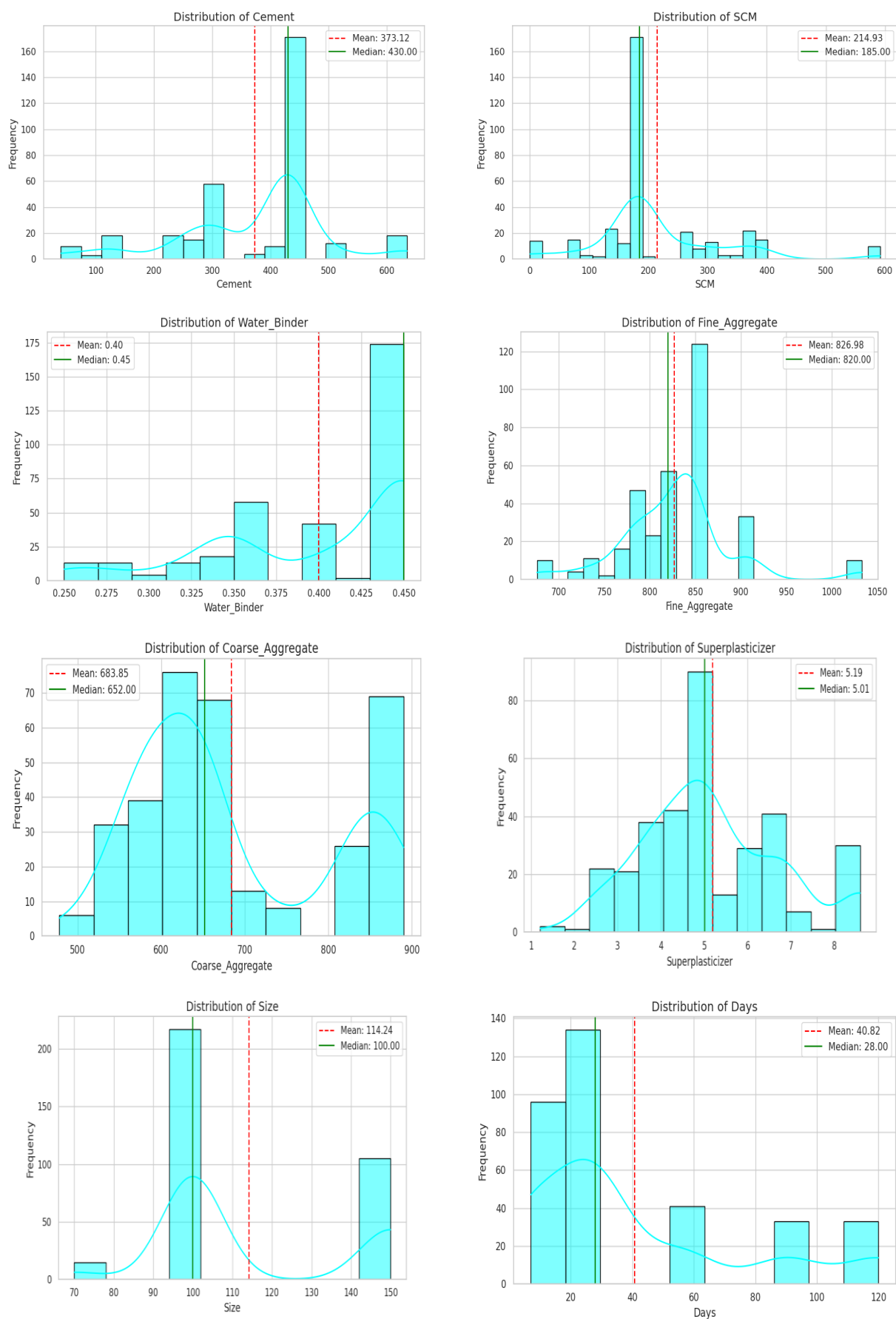
To avoid overfitting in the ANN model, three steps were taken. First, 10-fold cross-validation was used so that the model was tested on different parts of the data. Second, dropout (10%) and early stopping were applied during training to stop the model from simply memorizing the data. Third, all input values were standardized before training using z-score scaling, which adjusts each feature to have an average value of zero and a standard deviation of one. This prevents variables with large numbers, such as cement content, from dominating smaller ones like superplasticizer dosage. These steps helped the ANN to learn the true patterns in the data instead of producing artificially high results. Still, it is recognized that a larger dataset would further improve the reliability of the model, and this is one of the study's limitations.

3 Model Training

Several machine learning models were chosen and trained to forecast the compressive strength of concrete. The models used were ensemble methods such as Extra Trees, AdaBoost, and Gradient Boosting, which are effective in managing complex nonlinear relationships. Furthermore, Support Vector Regression (SVR) was used due to its strength in handling high-dimensional data. Finally, a Multi Layer Perceptron (MLP) was used because of its ability to model complex patterns effectively. Each model underwent thorough training using a specific training set that comprised 80% of the available data.

3.1 Extra Trees

The Extra Trees (ETR) is an ensemble-based machine learning model known for its reliability and high accuracy in predicting the compressive strength of concrete. ETR improves generalization and reduces the risk of overfitting by creating several decision trees using randomly chosen features and thresholds. The relevance of this approach in concrete research is clear from various case studies. For instance, ETR was used to forecast the compressive strength of concrete that includes waste marble powder, showing effective results and aiding in sustainable construction efforts [36]. It was also effectively utilized in assessing the strength of cellulose nanofiber reinforced concrete, demonstrating its versatility across different material compositions [37]. In terms of performance, ETR attained a correlation coefficient of 0.9603 in hydraulic engineering applications, highlighting its predictive accuracy. While ETR is effective, it is not always the best option. Models such as Random Forest and Gradient Boosting can produce similar results, indicating that the best model choice may depend on the complexity of the dataset, the goals of the project, and the limitations of computational resources.

**Fig. 2- Data Distribution**

3.2 AdaBoost

AdaBoost is an ensemble learning algorithm that enhances the performance of weak learners. It has shown significant effectiveness in predicting concrete mechanical properties, including compressive strength and elastic modulus. The method's strength is in its ability to sequentially correct the errors of previous models, enabling it to effectively model complex relationships within diverse concrete mixtures. AdaBoost has been effectively combined with other machine learning models like XGBoost in the field of sustainable construction. For example, in predictive studies that examined concrete with waste foundry sand, AdaBoost played a role in a combined modeling approach, resulting in XGBoost achieving a root mean squared error (RMSE) of 2.845 MPa, indicating high accuracy [38]. The performance of AdaBoost in these applications is usually assessed through metrics such as RMSE, mean absolute error (MAE), and correlation coefficients. These metrics collectively offer a thorough understanding of the model's reliability and ability to generalize [39]. A significant advantage of AdaBoost is its capacity to identify important input variables, including curing time, fiber content, and mix composition, which aids in optimizing mix designs and material selections. Furthermore, by enabling performance forecasting for eco-friendly concrete formulations, AdaBoost aids in the progress of sustainable construction technologies.

3.3 Support Vector Regression (SVR)

Support Vector Regression (SVR) is an effective machine learning method for predicting the properties of concrete, particularly its compressive strength. The capacity to model nonlinear relationships using kernel functions makes it appropriate for addressing the varied mix compositions and curing conditions found in contemporary concretes. SVR has been effectively utilized for different types of concrete, such as fly ash concrete, cellular lightweight concrete, and ultra high performance concrete (UHPC), demonstrating high correlation coefficients, for instance, $R = 0.981$ and $R = 0.9781$ for fly ash and lightweight concrete, respectively [40]. Furthermore, SVR has shown its effectiveness in predicting both fresh and hardened properties of self-compacting concrete by utilizing various kernel functions to enhance predictions [41]. Advanced variants like GS SVR (Grid Search SVR) and AOA SVR (Arithmetic Optimization Algorithm SVR) have improved predictive accuracy. Notably, AOA SVR produced the lowest error rates in predicting UHPC when compared to traditional models [42]. SVR can adapt to various input features and works well with optimization methods, making it a versatile and effective tool in concrete research [40]. However, challenges like high computational costs and the complexity of hyperparameter tuning require careful model calibration and the use of strong datasets to enhance its predictive capabilities.

3.4 Gradient Boosting

Gradient boosting has demonstrated effectiveness as a machine learning method for predicting the mechanical properties of concrete, especially its compressive strength. The success of this model is attributed to its capacity to represent complex nonlinear relationships between input features and output responses. Research has shown that gradient boosting models, including XGBoost and CatBoost, provide greater predictive accuracy compared to traditional statistical methods. For instance, CatBoost reached a notable R^2 value of 0.9838 when utilized for quaternary blend concrete, surpassing traditional algorithms [43]. Hybrid models that integrate gradient boosting with optimization techniques such as Natural Exponential Evolution Strategies improve prediction accuracy, particularly in the context of geopolymers self-compacting concrete applications [44]. Sensitivity analyses indicate that concrete age, slag content, and aggregate proportions are important factors affecting compressive strength. Additionally, methods such as Grey Relation Analysis assist in enhancing feature selection to boost model performance [45]. Gradient boosting, from a practical viewpoint, allows for efficient resource utilization by reducing the necessity for extensive laboratory testing and improving quality control via dependable strength estimation [46]. Although the model has strengths, its accuracy can still be affected by variations in concrete mixtures and environmental conditions, underscoring the necessity for ongoing research and refinement of the model. Most recently, regression-based optimization studies combining RCA, fly ash, and SCMs in eco-concrete mixes have reported promising results, where a Multiple Linear Regression (MLR) model achieved $R^2 = 0.8856$ and RMSE = 4.39 MPa, with curing age and W/CM identified as the most influential factor [47]. These findings reinforce the need for multivariate modeling frameworks to capture material interactions in sustainable concrete design.

3.5 ANN

ANN have become increasingly popular in concrete engineering because of their ability to effectively model complex, nonlinear relationships between mix components and concrete properties. They are especially useful for predicting compressive strength, optimizing material compositions, and monitoring structural health. For example, artificial neural networks have effectively predicted the compressive strength of both normal and high strength concrete, attaining R^2 values of 0.92 and 0.90 for the 7-day and 28-day strengths, respectively [48]. ANN have been used for structural monitoring, with convolutional neural networks (CNNs) reaching an accuracy of up to 70% in detecting and localizing defects in reinforced concrete structures [49]. Although these strengths exist, challenges such as the requirement for large, high-quality datasets and the risk of overfitting continue to pose significant issues. This necessitates careful design and validation of models to ensure their practical applicability in real-world situations.

In this study, a Multi-Layer Perceptron (MLP) architecture was used, consisting of an input layer with nine neurons (corresponding to the input parameters), two hidden layers with 100 and 50 neurons respectively, and an output layer representing compressive strength. The Rectified Linear Unit (ReLU) activation function was employed in the hidden layers, while a linear activation function was used in the output layer to handle continuous regression tasks. Weight optimization was carried out using the Adam optimizer with a learning rate of 0.001 and a maximum of 2000 training iterations. To avoid overfitting, dropout regularization (10%) and early stopping were implemented.

Since ANN performance is highly sensitive to the scale of input data, all features were standardized before model training. Standardization was performed using z-score normalization, where each input variable was transformed to have zero mean and unit standard deviation (mean = 0, std = 1). This step ensured that features with large numerical values, such as cement content, did not dominate those with smaller magnitudes, such as superplasticizer dosage. By rescaling input variables to a comparable range, the training process became more stable, the convergence rate improved, and the relative influence of each parameter on compressive strength prediction was preserved.

4 Validation Strategy

This paper utilized 10-fold cross-validation (CV) to confirm that the machine learning models exhibited reliable performance and could generalize effectively to new data as shown in Fig. 3. This method consists of splitting the dataset into ten equal parts, training the model on nine of these parts, and validating its performance on the one part that remains. This process is carried out ten times, with each subset being used once as the validation set, guaranteeing that every data point is utilized for both training and evaluation. The method improves the strength and dependability of model evaluation by decreasing the chances of overfitting and enabling consistent performance assessment across various data samples [50].

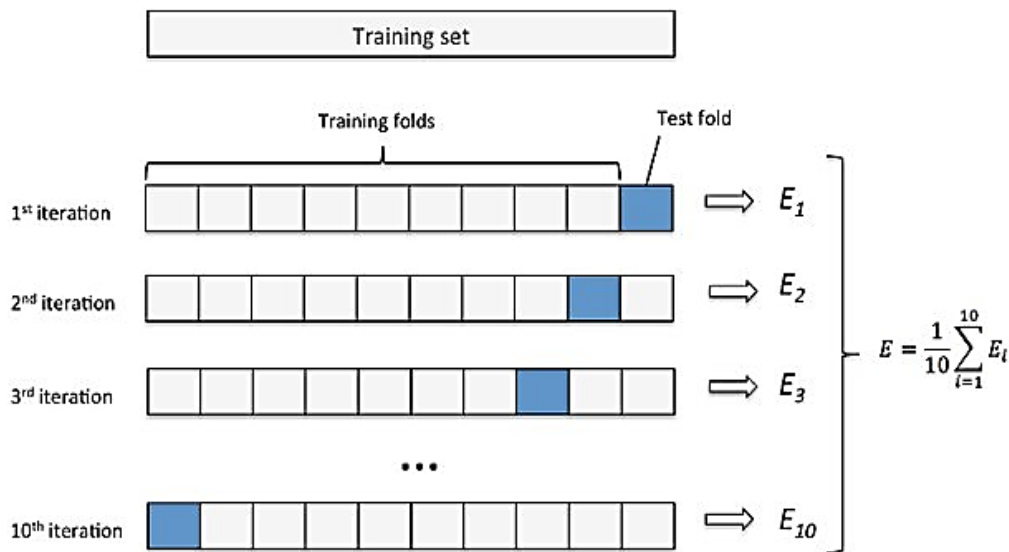


Fig. 3- 10 fold cross validation [48]

This approach offers key benefits, including the provision of reliable performance metrics like R^2 and MAE. These metrics are essential for assessing the prediction accuracy of models utilized in concrete strength evaluations [51]. Furthermore, 10-fold cross-validation facilitates effective model comparison, allowing researchers to determine the most appropriate algorithm, such as Gradient Boosting or Random Forest, for a particular prediction task [52]. This method requires significant computational resources, particularly when working with large datasets or complex models. Therefore, this trade-off should be taken into account when implementing it in practical engineering applications.

4.1 Evaluation Metrics

To evaluate and compare the performance of each developed model, several performance metrics were used, including the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The R^2 value shows the extent to which each model accounts for the variability of compressive strength in the dataset, with values nearer to 1 indicating better predictive accuracy. The Mean Absolute Error (MAE) offers an understanding of the average size of prediction errors, emphasizing the degree to which predicted values align with actual observations. The Root Mean Squared Error (RMSE) measures prediction accuracy by giving greater weight to larger errors, which makes it particularly sensitive to outliers and appropriate for ensuring robustness. These metrics provided a comprehensive assessment of the model's accuracy, consistency, and reliability in predicting concrete compressive strength.

1. Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |A_i - P_i| \quad (1)$$

2. Mean Squared Error (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^N (A_i - P_i)^2 \quad (2)$$

3. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - P_i)^2} \quad (3)$$

4. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^N (A_i - P_i)^2}{\sum_{i=1}^N (A_i - \bar{A})^2} \quad (4)$$

5. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{A_i - P_i}{A_i} \right| \quad (5)$$

Notation-

- A_i - Actual observed value
- P_i - Predicted value
- $\bar{A}$ - Mean of actual values
- N - Total number of observations

4.2 Hyperparameter Tuning

Hyperparameter tuning was performed to enhance model performance, utilizing both grid search and manual adjustments as shown in table 1. The optimal number of estimators for the Extra Trees model was found to be 150. The AdaBoost model used 150 estimators and a learning rate of 0.1, which helped to balance model complexity and predictive accuracy. The parameters for the SVR included a radial basis function (RBF) kernel, with the regularization parameter (C) set to 100 and an epsilon value of 0.1. The hyperparameters for Gradient Boosting were adjusted to include 100 estimators, a learning rate of 0.1, and a maximum depth of 5. This configuration aims to model complex nonlinearities while minimizing the risk of overfitting. The ANN model employed a hidden layer structure of (100, 50), utilized ReLU activation, applied the Adam optimizer, and set a maximum iteration limit of 2000 epochs to reach optimal convergence and accuracy.

Table 1- Hyperparameter tuning

Model	Hyperparameter Settings
Extra Trees	n_estimators = 150, max_depth = None, min_samples_split = 2, random_state = 42
AdaBoost	n_estimators = 150, learning_rate = 0.1, loss = 'linear', random_state = 42
SVR (SVR)	kernel = 'rbf', C = 100, epsilon = 0.1, gamma = 'scale'
Gradient Boosting	n_estimators = 100, learning_rate = 0.1, max_depth = 5, loss = 'squared_error', random_state = 42
ANN (MLP)	hidden_layer_sizes = (100, 50), activation = 'relu', solver = 'adam', max_iter = 2000, random_state = 42

5 Results and discussion

5.1 Model Performance

This section evaluates the performance of five machine learning models using both training and testing datasets. The analysis is conducted through statistical indicators, regression accuracy metrics, residual diagnostics, and interpretability methods as shown in table 2.

Table 2- Model Performance Metrics

Model	Train MAE	Train MSE	Train RMSE	Train R ²	Train MAPE	Test MAE	Test MSE	Test RMSE	Test R ²	Test MAPE
Extra Trees	0.6623	1.8812	1.3716	0.9892	0.0165	3.2734	24.4466	4.9444	0.8309	0.1161
AdaBoost	4.0828	26.4600	5.1439	0.8474	0.1302	5.8774	51.1571	7.1524	0.6462	0.2079
SVR	2.0292	12.3333	3.5119	0.9289	0.0632	4.0330	42.8543	6.5463	0.7037	0.1439
Gradient	0.9396	2.1263	1.4582	0.9877	0.0248	2.6382	16.2566	4.0320	0.8876	0.0856
ANN	2.2057	9.6830	3.1118	0.9442	0.0667	3.1853	25.9605	5.0951	0.8205	0.1124

Table 1 presents the performance comparison among Extra Trees, AdaBoost, SVR, Gradient Boosting, and ANN. The performance of the Extra Trees model is illustrated using two visual representations. Fig. 4(a) shows the Actual versus Predicted Compressive Strength for the training and testing datasets. The data points are closely grouped around the red dashed line, which signifies perfect prediction. Most points fall within the $\pm 20\%$ error margins indicated by the blue lines, demonstrating strong predictive accuracy and minimal deviation. Fig. 4(b) presents the Actual, Predicted, and Residuals Plot, illustrating that the actual and predicted values closely align, thereby indicating a high level of model reliability. The residuals shown below are generally stable and close to zero for all data points, indicating a low level of prediction error. The results are backed by strong quantitative metrics, with a Train R² of 0.9892 and a Test R² of 0.8309. Additionally, the error values are low, with a Train MAE of 0.6623 and a Test MAE of 3.2734. This validates the Extra Trees model's effectiveness in predicting compressive strength in SCC.

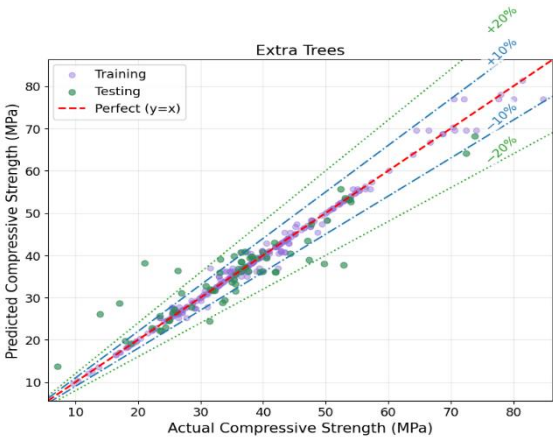


Fig. 4(a)- Actual vs Predicted Extra Trees

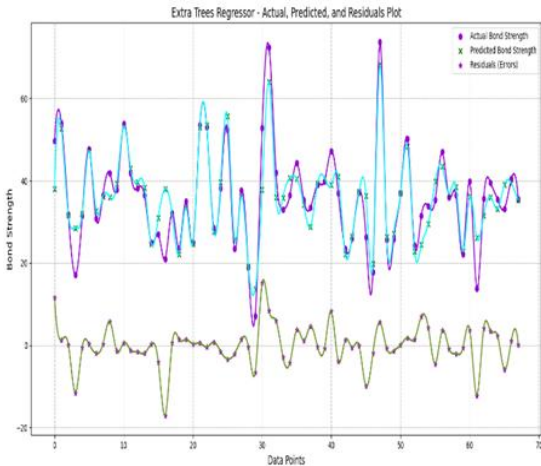


Fig. 4(b)- Residuals Plot Extra Trees

The performance of AdaBoost is illustrated in two graphs. Fig. 4(c) shows the comparison between Actual and Predicted Compressive Strength for the training and testing datasets. The model's predictions align with the overall trend of the actual values; however, there is some variation from the ideal prediction line (red dashed), particularly in the testing data. Fig. 4(d) presents the Actual, Predicted, and Residuals Plot, indicating that the predicted values exhibit significant divergence from the actual bond strength in various areas.

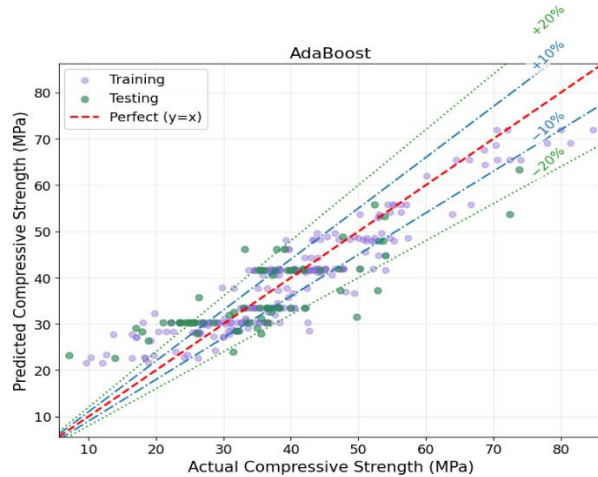


Fig. 4(c)- Actual vs Predicted AdaBoost

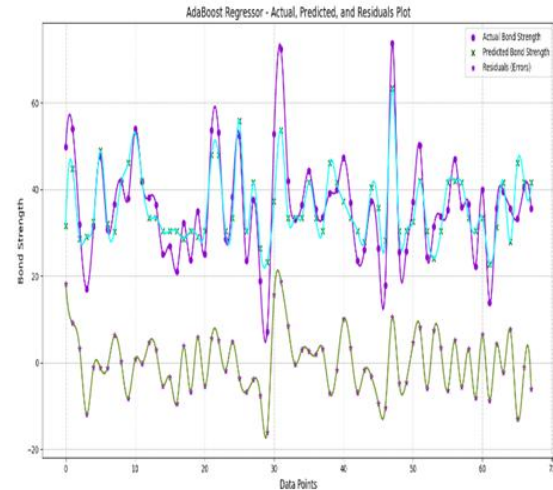


Fig. 4(d)- Residuals Plot AdaBoost

The residuals show different prediction errors among the data points. The quantitative results corroborate this visual assessment, showing a Train R^2 of 0.8474 and a Test R^2 of 0.6462, which indicates moderate model performance. The model also shows higher error values, with a Train MAE of 4.0828 and a Test MAE of 5.8774, along with a Test RMSE of 7.1524. This indicates that AdaBoost has limited accuracy when compared to more effective methods like Extra Trees for predicting compressive strength in SCC.

The performance of the SVR model is illustrated using two plots. Fig. 4(e) presents the Actual versus Predicted Compressive Strength for both training and testing data. The majority of points are in close alignment with the perfect prediction line (red dashed), suggesting a satisfactory level of accuracy. However, some dispersion, especially in the testing data, indicates variability in generalization. Fig. 4(f) shows the Actual, Predicted, and Residuals Plot. The predicted values generally align with the actual trend, although periodic deviations are noted. The residuals plot shows variations in prediction error among the data points. The SVR model demonstrates strong training performance, with a R^2 value of 0.9289 and a mean absolute error (MAE) of 2.0292. However, its effectiveness on the test set is somewhat lower, showing a R^2 of 0.7037 and an MAE of 4.033. This suggests a moderate ability to predict compressive strength, reflecting a balance between underfitting and overfitting.

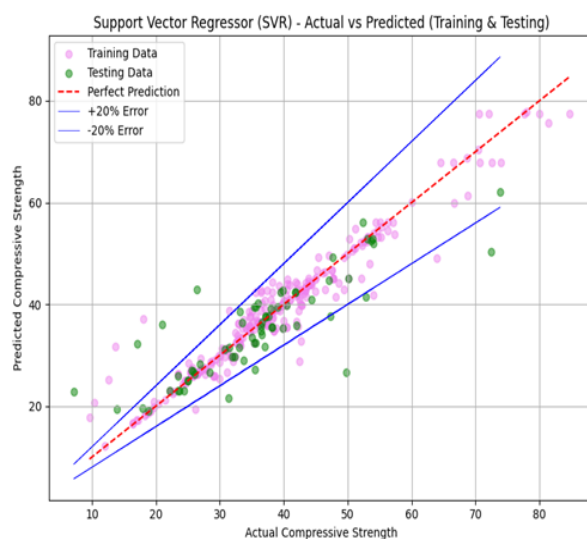


Fig. 4(e)- Actual vs Predicted SVR

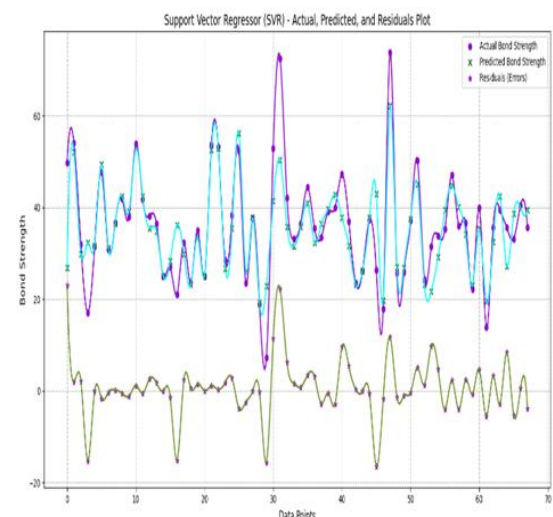


Fig. 4(f)- Residuals Plot SVR

The performance of Gradient Boosting is illustrated in two Figs. Fig. 4(g) shows the Actual versus Predicted Compressive Strength for the training and testing datasets. The majority of data points closely follow the ideal prediction line (red dashed), indicating a high level of prediction accuracy and consistency in both datasets. Fig. 4(h) shows the Actual, Predicted, and Residuals Plot, indicating that the predicted values closely follow the actual bond strength with little variation. The residual curve is near zero, suggesting that prediction errors are low and stable. The strong performance metrics support this conclusion, with a Train R^2 of 0.9877 and a Test R^2 of 0.8876, along with low error values (Train MAE of 0.9396 and Test MAE of 2.6382). This confirms that Gradient Boosting is highly effective for modeling compressive strength in SCC.

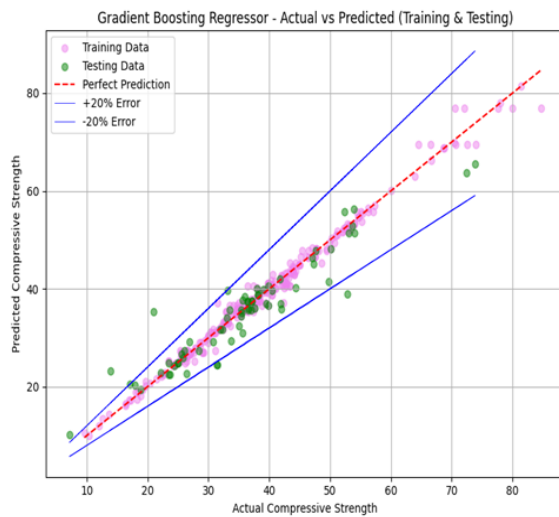


Fig 4(g)- Actual vs Predicted Gradient Boosting

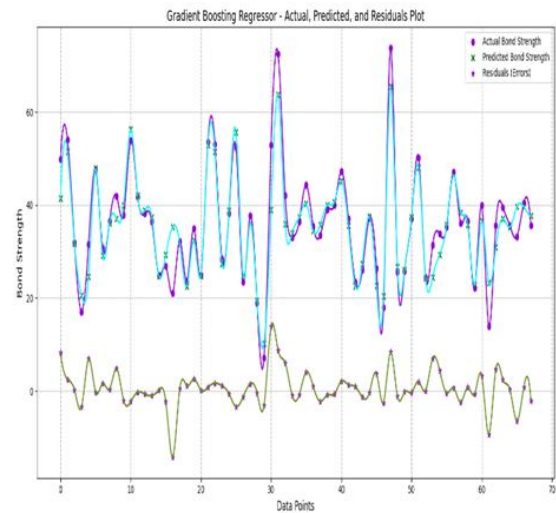


Fig 4(h)- Residuals Plot Gradient Boosting

The performance of the ANN model is illustrated in two plots. Fig. 4(i) shows the comparison between Actual and Predicted Compressive Strength for the training and testing datasets. The predicted values closely align with the perfect prediction line (red dashed), with most points showing a consistent pattern, which indicates good model accuracy. Fig. 4(j) presents the Actual, Predicted, and Residuals Plot, indicating that the predicted values closely align with the actual bond strength, exhibiting minimal deviations. The residuals are consistently low and stable throughout the dataset. The evaluation metrics indicate strong support for this analysis, with a Train R^2 of 0.9442 and a Test R^2 of 0.8205. Additionally, the low error values (Train MAE of 2.2057 and Test MAE of 3.1853) underscore the ability of the ANN to accurately model the compressive strength of SCC.

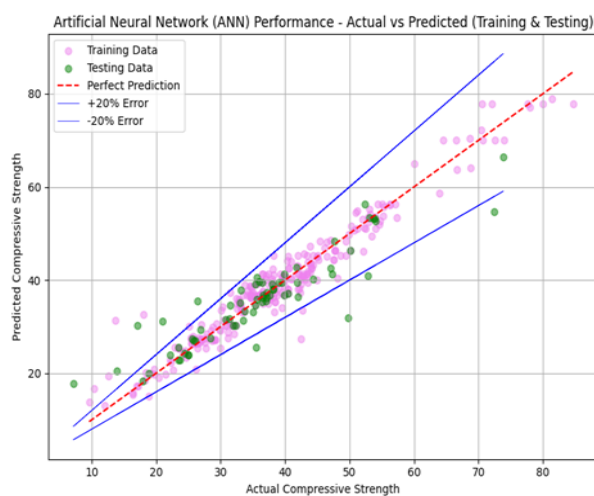


Fig 4(i)- Actual vs Predicted ANN

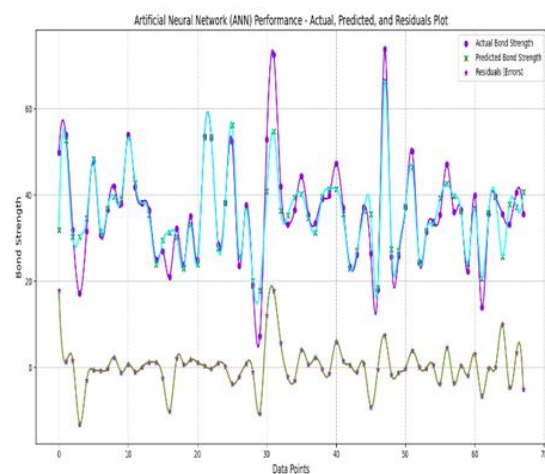


Fig 4(j)- Residuals Plot ANN

5.2 Model performance across strength ranges

To ensure robustness, the predictive models were evaluated in different compressive strength ranges following the approach of Ali et al. (2022) [53], who analyzed nanosilica-modified concretes. The SCC dataset was divided into two ranges: 10–80 MPa (low-to-medium strength) and >80 MPa (high strength). Table 3 summarizes the results. Gradient Boosting achieved the highest accuracy across both ranges ($R^2 = 0.891$, RMSE = 3.92 MPa for 10–80 MPa; $R^2 = 0.853$, RMSE = 4.71 MPa for >80 MPa). Similar to the findings of Ali et al. (2022) [53], model accuracy was slightly higher in the mid-strength range due to greater data availability, whereas prediction for high-strength SCC showed moderately lower accuracy. These results confirm that dataset balance and material composition strongly affect prediction quality across strength domains.

Table 3 – Performance comparison of machine learning models across different concrete strength ranges

Model	Strength Range	Test R^2	RMSE (MPa)	MAE (MPa)
Extra Trees	10–80 MPa	0.842	4.65	3.28
	>80 MPa	0.801	5.12	3.71
AdaBoost	10–80 MPa	0.672	7.02	5.66
	>80 MPa	0.611	7.85	6.12
Gradient Boosting	10–80 MPa	0.891	3.92	2.61
	>80 MPa	0.853	4.71	3.05
ANN	10–80 MPa	0.811	4.96	3.18
	>80 MPa	0.785	5.42	3.39

These observations are in agreement with the findings of Ali et al. (2022) [53], who demonstrated that the full quadratic (FQ) model achieved $R^2 = 0.96$ in nanosilica-modified concretes but with variation across strength ranges (20–55 MPa, 55–70 MPa, and 70–100 MPa). Similar to their study, our results highlight that prediction performance is range-dependent, and models achieve relatively higher accuracy in the mid-strength domain due to data density.

5.3 Pairplot of input features and compressive strength

Fig. 5 shows a pairplot of the input features and compressive strength, providing a detailed view of the distribution and relationships between the variables in the dataset. The diagonal plots show the distribution of each feature through kernel density estimates, emphasizing skewness or multimodal behavior (for example, Cement and Curing Age). Off-diagonal scatter plots illustrate the correlation patterns among pairs of features. Cement and Curing Age demonstrate a clear positive relationship with compressive strength, while the Water/Binder Ratio exhibits an opposite trend. This visualization aids in recognizing feature interdependencies, clustering, and possible nonlinear trends, acting as a crucial tool for comprehending data structure prior to model development.

Figure 6 presents the repeatability check for all five machine learning models, where predicted compressive strength values are compared against experimental results with error bars representing the standard deviation (≥ 3 replicates per testing condition). For the Extra Trees and Gradient Boosting models, the predicted values align very closely with the experimental mean values across most strength ranges, with error bars consistently overlapping. This indicates strong repeatability and robustness of the models, even under conditions with experimental variability. In contrast, the AdaBoost model shows wider deviations in several conditions, where the predictions tend to underestimate compressive strength compared to experimental means. The error bars reveal that reproducibility is weaker for this model, which is consistent with its lower R^2 and higher error metrics reported earlier. The SVR model demonstrates good reproducibility, with most predicted points falling within one standard deviation of the experimental means. However, slight dispersion is observed in mid-strength ranges, suggesting sensitivity to scaling and parameter tuning. The ANN model exhibits stable repeatability, with predicted strengths showing close agreement with experimental means and consistent overlap with error bars. The standardization of input variables appears to have improved both stability and accuracy, reducing the risk of overfitting to a small dataset.

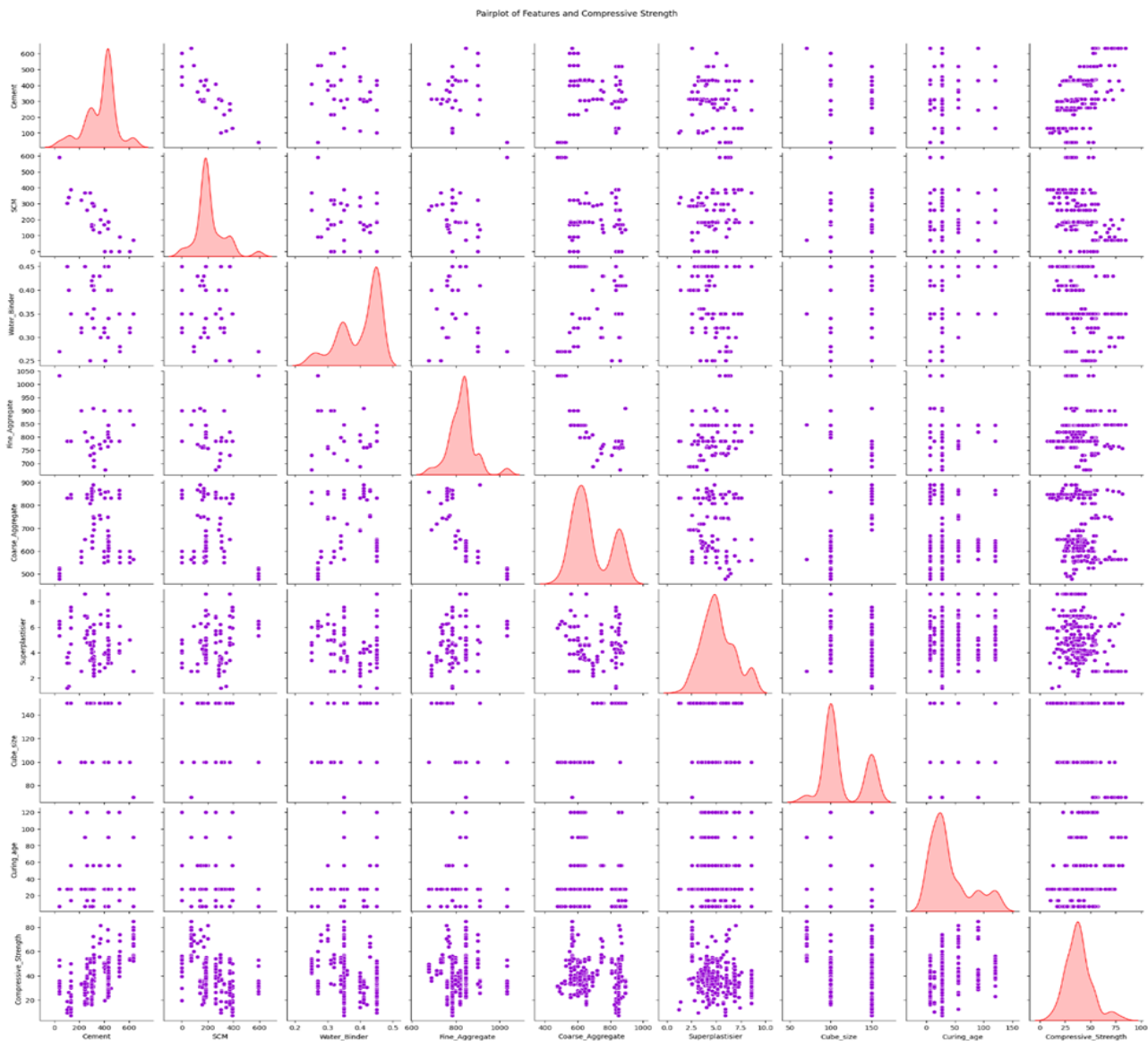
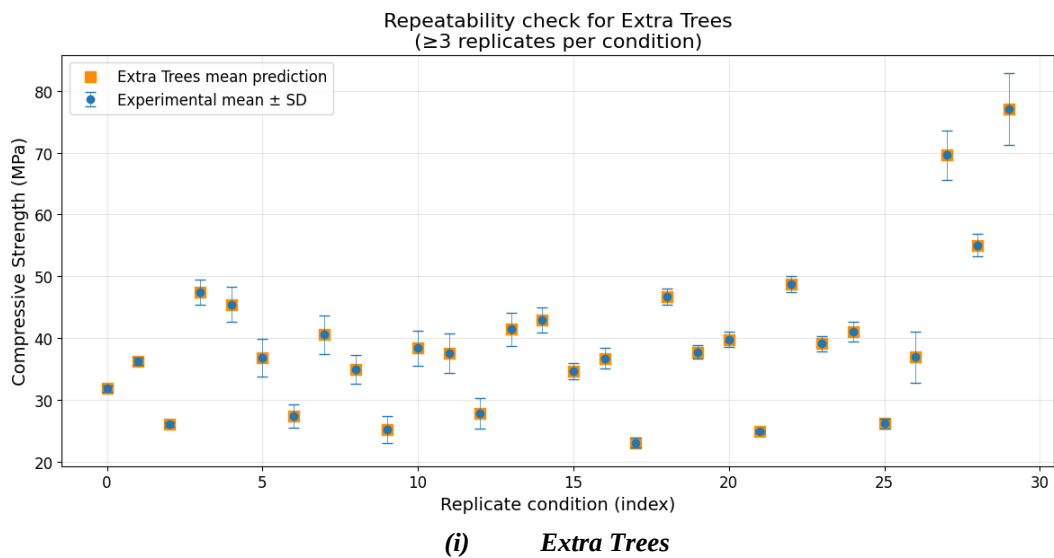
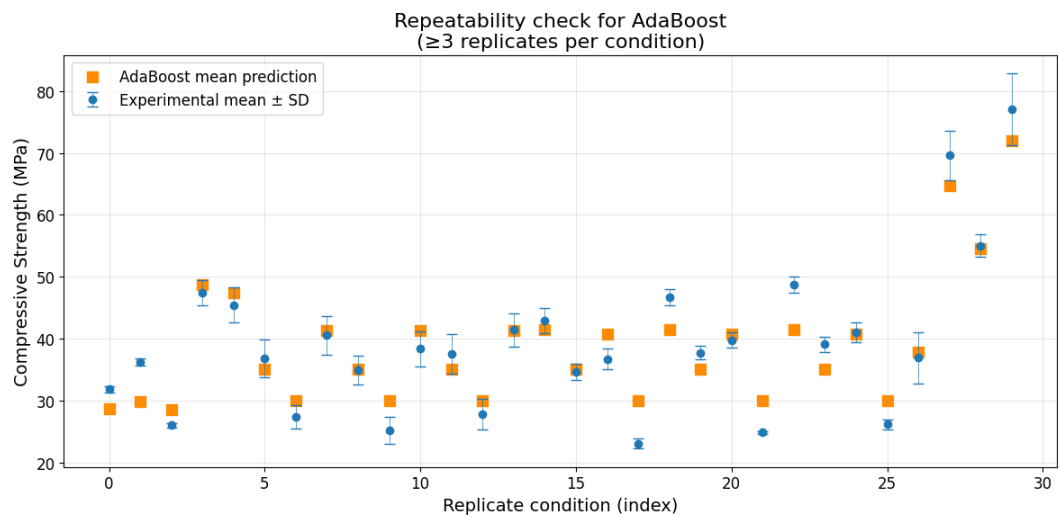
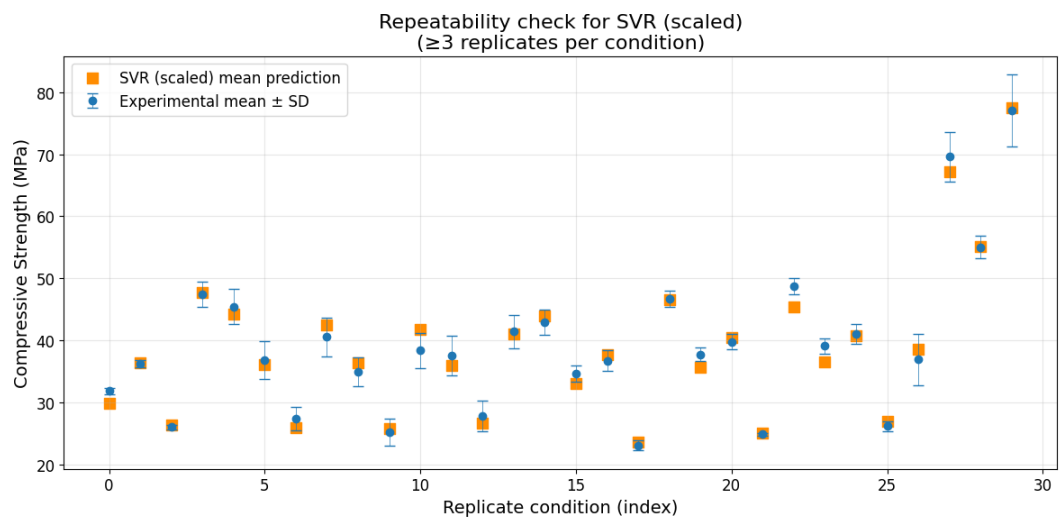


Fig. 5- pairwise scatter distribution among all input features

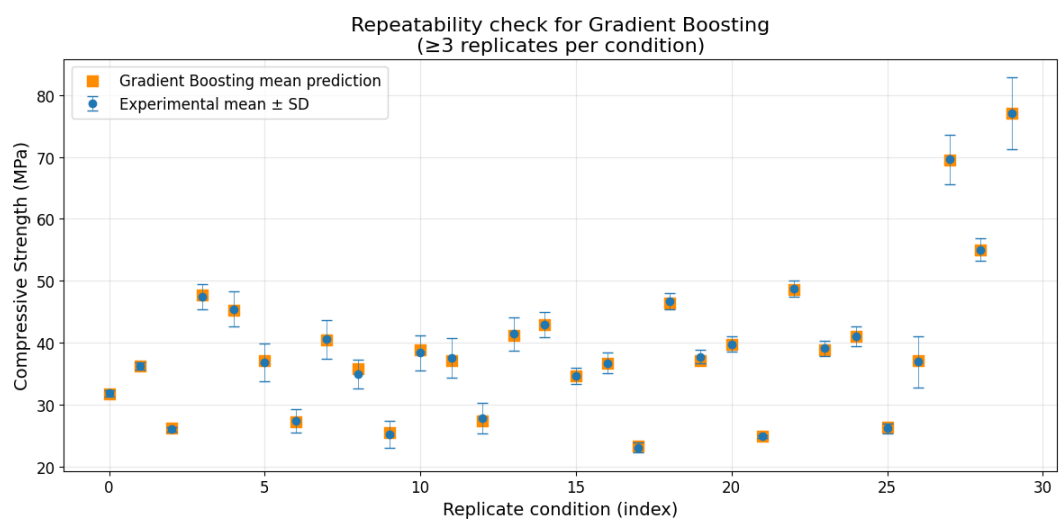




(ii) *Ada Boost*



(iii) *SVR*



(iv) *Gradient Boosting*

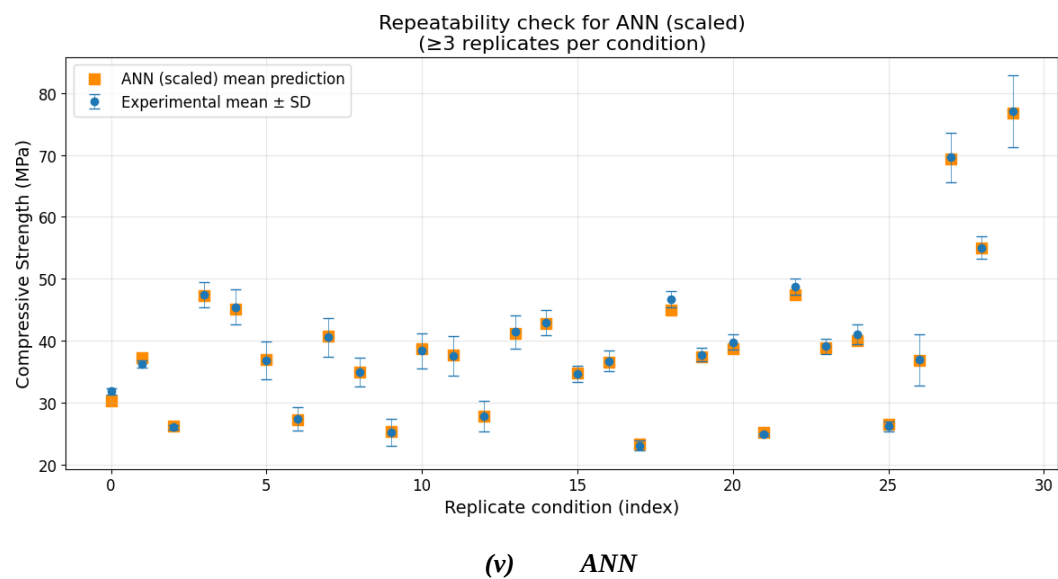


Fig. 6- Pairwise scatter distribution among all input features

Overall, these results confirm that ensemble models like Gradient Boosting and Extra Trees provide the most reliable predictions across different replicates, while ANN also performs consistently when appropriate preprocessing (standardization, dropout, early stopping) is applied. AdaBoost, however, shows weaker reproducibility, supporting its lower statistical performance metrics.

5.4 Pearson correlation heatmap

Fig. 7 shows the Pearson correlation heatmap, which shows the linear relationships between the input features and compressive strength.

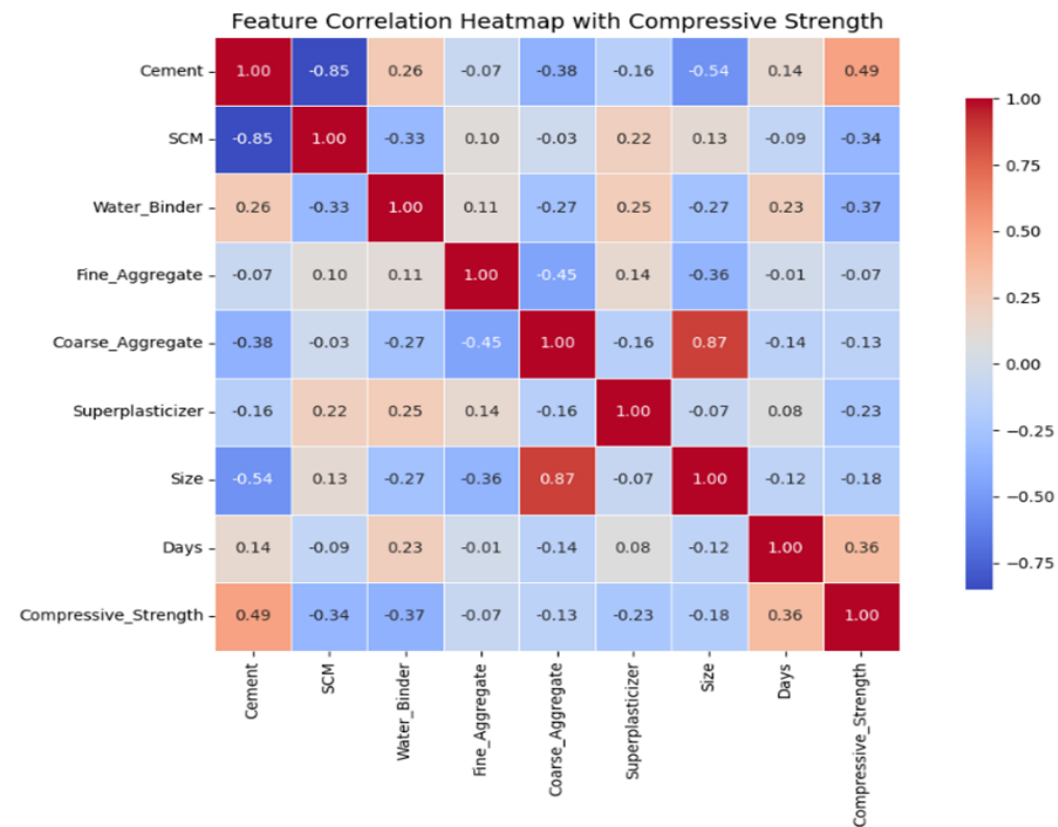


Fig. 7- Correlation Heatmap with Compressive Strength

The variables of cement and curing age (in days) exhibit the strongest positive correlations, with values of 0.49 and 0.36, respectively. This suggests that they have a significant impact on the development of strength.

The water to binder ratio (-0.37) and supplementary cementitious materials (SCM, -0.34) show moderate negative correlations, indicating that their excessive use may negatively impact strength. Other variables, including aggregates, superplasticizer, and size, demonstrate weak correlations. This visual analysis aids in identifying key factors and guides effective feature selection for machine learning-based strength prediction models.

5.5 Information gain of each input feature

Fig. 8 shows a radial plot that shows the information gain of each input feature in relation to predicting compressive strength. Cement plays a significant role in decreasing prediction uncertainty, thus serving as the most informative feature. Fine aggregate, SCM, and curing age (Days) demonstrate significant information gain, suggesting a strong impact on the target variable. The coarse aggregate and the water to binder ratio have a moderate effect, whereas the superplasticizer and size have the least influence. This chart ranks the features according to their predictive importance, aiding in the prioritization of key variables in model development for accurate strength estimation.

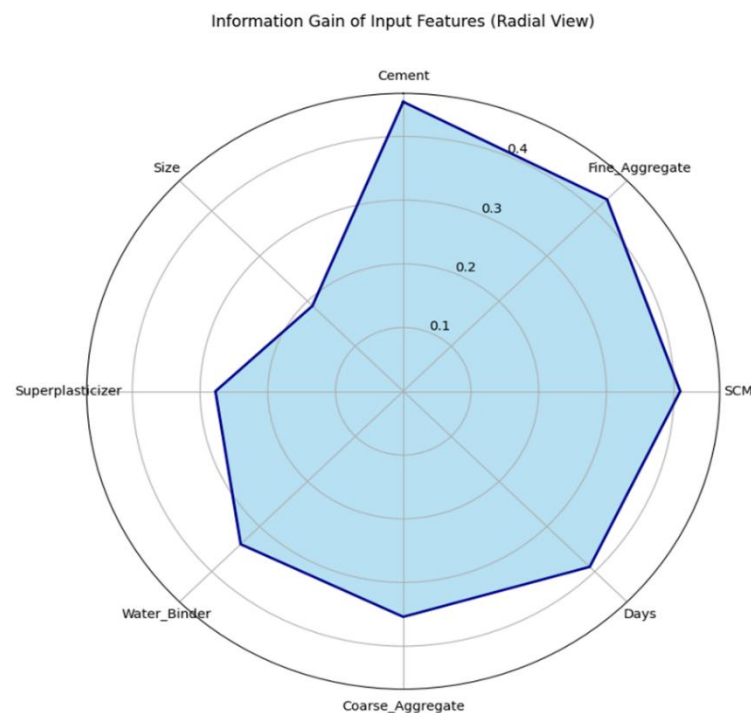


Fig. 8- Information gain of each input feature

5.6 REC curve

Fig. 9 shows the Regression Error Characteristic (REC) curves for the training and testing datasets, allowing for a comparison of model accuracy at different tolerance levels. In the training set (left plot), the Gradient Boosting and Extra Trees models achieve the highest accuracy at lower error thresholds, indicating a better fit.

The AdaBoost and ANN models demonstrate slower increases in accuracy, suggesting lower precision during the training process. In the testing set (right plot), the Gradient Boosting model maintains a strong performance, with Extra Trees and SVR closely following. The REC curves indicate that Gradient Boosting provides the highest prediction accuracy for both sets, whereas AdaBoost shows weaker generalization performance. This analysis highlights the tolerance of each model to prediction error, facilitating a strong comparison between models.

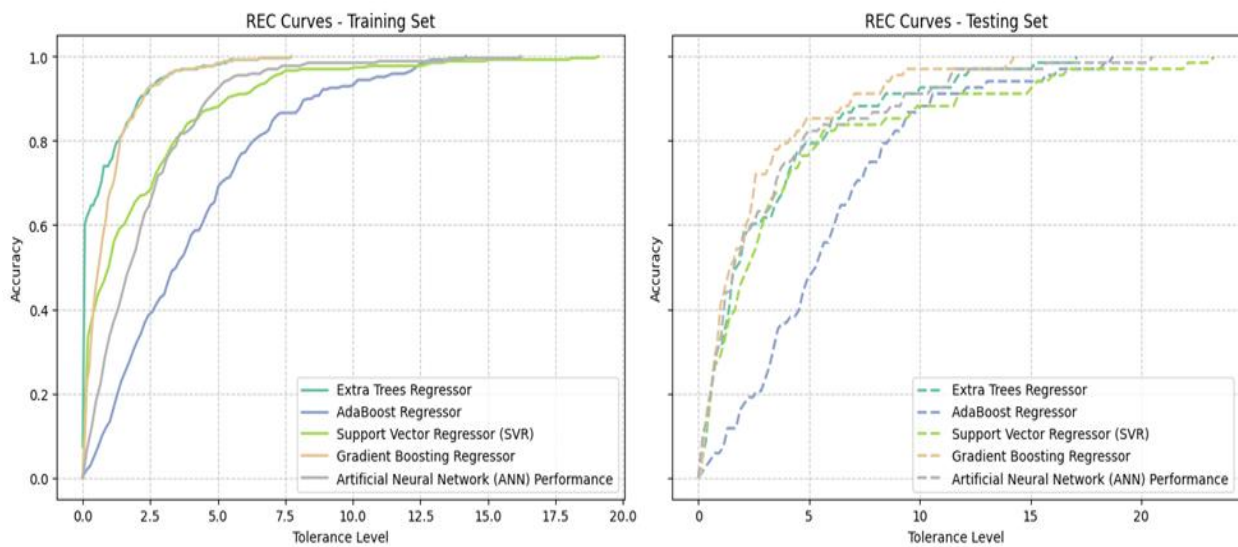


Fig. 9- Relative Error Analysis Plot

5.7 SHAP analysis

Fig. 10 demonstrates the global and local interpretability of model predictions using SHAP. The Fig. 10(a) plot (SHAP waterfall plot) shows the contribution of individual input features to a specific prediction. Cement has the most significant negative effect on the prediction (-12.99), whereas Days and Superplasticizer have a positive influence on the output, suggesting their contributions to enhancing compressive strength.

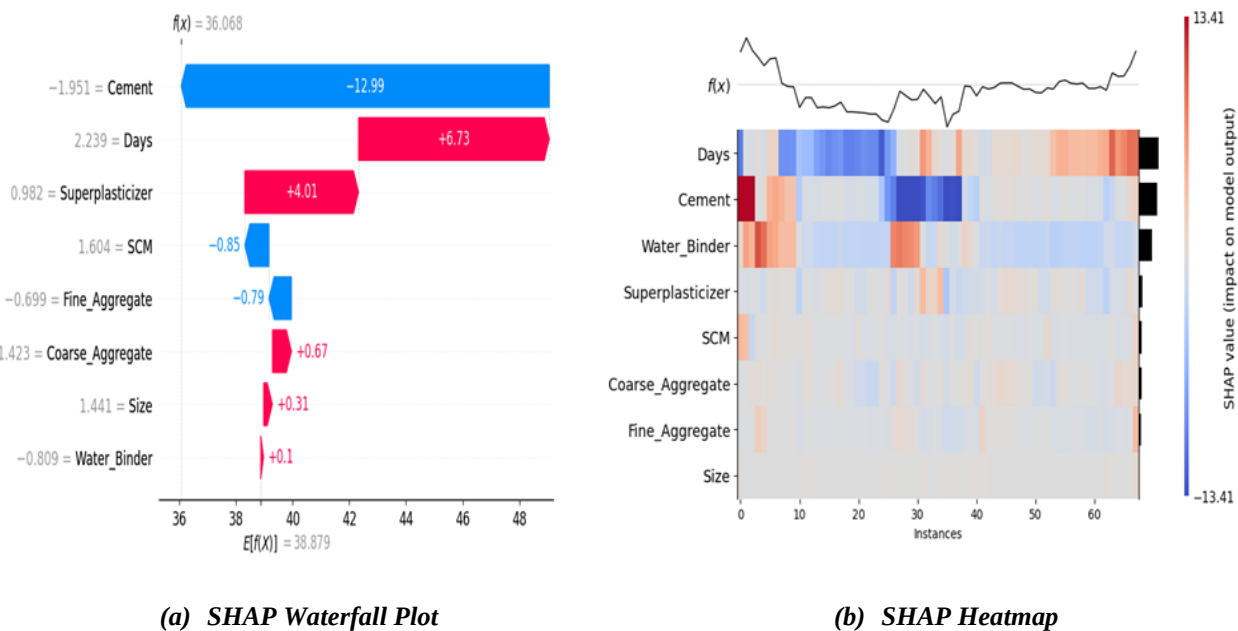


Fig. 10- SHAP Interpretability

Features such as SCM and Fine Aggregate exhibit slight negative changes, whereas Coarse Aggregate, Size, and Water Binder have minimal contributions. Fig 10(b) plot (SHAP heatmap) shows SHAP values for all instances, illustrating the overall impact of each feature. Darker shades indicate more significant effects. Cement and Days display consistent influence across samples, reaffirming their dominance. This analysis of interpretability confirms the features that most significantly influence model predictions and how their effects differ at both individual and global levels, thereby improving transparency in the decision-making process.

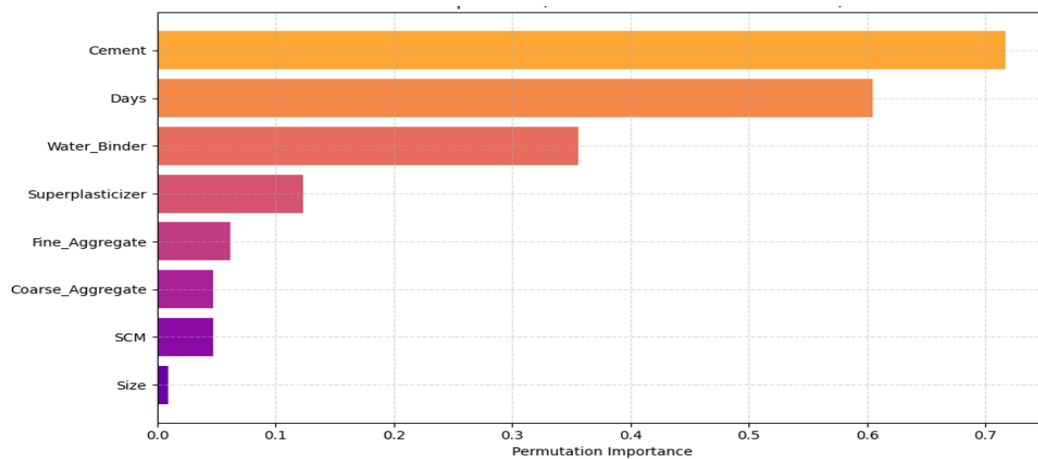


Fig. 11- feature importance

6 Novelty and Contributions

- Unlike prior studies that examined SCMs or RCA in isolation, this work jointly evaluated RCA and SCMs within SCC using a benchmark of five advanced models (Extra Trees, AdaBoost, SVR, Gradient Boosting, ANN).
- The study introduced a transparent, data-driven framework combining prediction accuracy with feature interpretability (SHAP, ELI5, correlation, information gain).
- By analyzing performance across strength ranges and replicates, the work highlights dataset-driven variability, providing insights for practical application in sustainable construction.
- The approach demonstrates the feasibility of machine learning as a faster, non-destructive alternative to laboratory testing while ensuring explainability for engineering decision-making.

7 Limitations

- The dataset (337 mixes) is moderate in size and may not capture the full variability of material sources, curing conditions, and mix proportions in practice.
- No external experimental validation was conducted, which limits the generalization of the findings beyond the studied dataset.
- The study focused exclusively on compressive strength, without extending predictions to durability or other mechanical properties.

8 Future Research

- Expanding datasets through experimental work and multi-source compilation to improve robustness and generalizability.
- External validation using independent laboratory datasets to confirm model transferability.
- Development of hybrid and deep learning models for predicting a wider range of properties (e.g., tensile, flexural, durability indicators).
- Creating practical software tools or web-based platforms integrating high-performing ML models with interpretability techniques to support engineers in real-time mix design optimization.

9 Conclusion

This study demonstrated the effectiveness of machine learning in predicting the compressive strength of self-compacting concrete (SCC) incorporating recycled concrete aggregates (RCA) and supplementary cementitious materials (SCMs). Among the five models tested, Gradient Boosting achieved the highest accuracy with a Test R^2 of 0.8876, MAE of 2.64 MPa,

and RMSE of 4.03 MPa, outperforming Extra Trees ($R^2 = 0.8309$) and ANN ($R^2 = 0.8205$), while AdaBoost showed the weakest performance ($R^2 = 0.6462$). Feature interpretability analyses using SHAP and permutation importance consistently identified cement content, curing age, and water-to-binder ratio as the most influential parameters, with cement having the strongest positive contribution to strength development. The novelty of this work lies in jointly evaluating RCA and SCMs in SCC within a predictive framework that not only delivers high accuracy but also ensures transparency through interpretability tools. By validating performance across strength ranges (10–80 MPa and >80 MPa) and replicates, the study confirmed the robustness of Gradient Boosting and Extra Trees as reliable alternatives to traditional destructive testing. These findings provide a data-driven pathway for sustainable mix design, demonstrating that ensemble and neural network models can accelerate green construction practices by reducing reliance on time-consuming laboratory experiments while incorporating industrial by-products into concrete production.

REFERENCES

- [1]- O.O. Omotayo, S.L. Akingbomire, C.M. Ikumapayi, Sustainable Application of Materials from Construction and Demolition Waste: A Review. *J. Eng. Eng. Technol.* 13(2) (2019) 228–242.
- [2]- L.C. Quesada-Ruiz, L. García-Romero, N. Ferrer-Valero, Mapping environmental crime to characterize human impacts on islands: an applied and methodological research in Canary Islands. *J. Environ. Manage.* 346 (2023) 118833. <https://doi.org/10.1016/j.jenvman.2023.118959>
- [3]- J.T. Sheth, K. Patel, D. Shah, Analysis of Construction and Demolition Waste Management Policies from Sustainability Perspective. *Int. J. Sci. Res. Dev.* (2016) 35–39.
- [4]- A.O. Daoud, H.S. Omar, A.A.E. Othman, O.J. Ebohon, Integrated Framework Towards Construction Waste Reduction: The Case of Egypt. *Int. J. Civ. Eng.* 21(5)(2023) 695–709. <https://doi.org/10.1007/s40999-022-00793-2>.
- [5]- M.B. Ali, S. Rahman, M. Hossain, A review on emission analysis in cement industries. *Renewable Sust. Energy Rev.* 15(5) (2011) 2252–2261. <https://doi.org/10.1016/j.rser.2011.02.014>
- [6]- O.E. Ige, Integrated life cycle assessment and system dynamics model for prediction of cement production and environmental impact of cement industry. Doctor of Engineering Thesis, Durban University of Technology, March, 2023. <https://doi.org/10.13140/RG.2.2.22318.11846>
- [7]- O. Oguntola, S.J. Simske, Continuous Assessment of the Environmental Impact and Economic Viability of Decarbonization Improvements in Cement Production. *Resources* 12(8) (2023) 95. <https://doi.org/10.3390/resources12080095>
- [8]- R. Gieré, Towards a more sustainable cement and concrete industry. In: G. George, M.R. Haas, H. Joshi, A.M. McGahan, P. Tracey (eds): *Handbook on the Business of Sustainability: The Organization, Implementation, and Practice of Sustainable Growth*. Edward Elgar Publishing, Cheltenham, UK, p. 274–299 <https://doi.org/10.4337/9781839105340.00024>
- [9]- A.I. Nicoara, A.E. Stoica, M. Vrabec, N. Rogan, S. Šturm, C.W. Ow-Yang, M.A. Gülgün, Z.B. Bundur, I. Ciuca, B.S. Vasile, End-of-Life Materials Used as Supplementary Cementitious Materials in the Concrete Industry. *Materials (Basel)* 13(8) (2020) 1954. <https://doi.org/10.3390/ma13081954>
- [10]- P. Saha, P. Debnath, P. Thomas, Prediction of fresh and hardened properties of self-compacting concrete using support vector regression approach. *Neural Comput. Appl.* 32(12) (2020) 7995–8010. <https://doi.org/10.1007/s00521-019-04267-w>
- [11]- J.S. James, E. John, Synergistic effects of SCMs and fibers on plastic shrinkage control of sustainable concrete. *Green Mater.* (2025) 1–47. <https://doi.org/10.1680/jgrma.24.00169>
- [12]- H.M. Owaid, R.B. Hamid, M.R. Taha, A review of sustainable supplementary cementitious materials as an alternative to all-portland cement mortar and concrete. *Aust. J. Basic Appl. Sci.* 6(9) (2012) 287–303.
- [13]- K. Sobolev, M. Kozhukhova, K.K. Sideris, E. Menéndez, M. Santhanam, Alternative Supplementary Cementitious Materials. In : *Properties of Fresh and Hardened Concrete Containing Supplementary Cementitious Materials State-of-the-Art Report of the RILEM Technical Committee 238-SCM, Working Group 4*. 2018, pp. 233–282. <https://doi.org/10.1007/978-3-319-70606-1>
- [14]- A. Dhiman, A review on Self Compacting Concrete. *Int. J. Res.* 3(10) (2016) 568–571.
- [15]- R. Sharma, J.G. Jang, P.P. Bansal, A comprehensive review on effects of mineral admixtures and fibers on engineering properties of ultra-high-performance concrete. *J. Build. Eng.* 45 (2022) 103314. <https://doi.org/10.1016/j.jobbe.2021.103314>

- [16]- K.M. Reddy, J.O. Prakash, S. Sudheer, Studies On Self Compacting Concrete Made With Crushed Rock Dust And Marble Sludge Powder. *Int. J. Res.* 6(10) (2019) 1068–1087.
- [17]- Y.U. Khaleel, S.D. Qubad, A.S. Mohammed, R.H. Faraj, Reinventing concrete: a comprehensive review of mechanical strength with recycled plastic waste integration. *J. Build. Pathol. Rehabil.* 9(2) (2024) 111. <https://doi.org/10.1007/s41024-024-00465-9>
- [18]- H. Panghal, A. Kumar, Enhancing concrete performance: Surface modification of recycled coarse aggregates for sustainable construction. *Constr. Build. Mater.* 411 (2024) 134432. <https://doi.org/10.1016/j.conbuildmat.2023.134432>
- [19]- R.Y. Tenzin, C. Norbu, Chimmi, RECYCLED AGGREGATE CONCRETE (RAC): A viable solution for sustainable construction. *J. Appl. Eng. Technol. Manag.* 3(1) (2023) 63–69. <https://doi.org/10.54417/jaetm.v3i1.111>
- [20]- A. Sahu, S. Kumar, A.K.L. Srivastava, Combined Use of Fine and Coarse Recycled Concrete Aggregates in Concrete for Sustainable Development: A Review. In: Saha, S., Sajith, A.S., Sahoo, D.R., Sarkar, P. (eds) *Recent Advances in Materials, Mechanics and Structures. Lecture Notes in Civil Engineering*, vol 269(2023) Springer, Singapore. https://doi.org/10.1007/978-981-19-3371-4_43 367 (2022) 501–516.
- [21]- A.K. Ibrahim, H.Y. Dhahir, A.S. Mohammed, H. Ali Omar, A.H. Sedo, The effectiveness of surrogate models in predicting the long-term behavior of varying compressive strength ranges of recycled concrete aggregate for a variety of shapes and sizes of specimens. *Arch. Civ. Mech. Eng.* 23(61) (2023) 244. <https://doi.org/10.1007/s43452-022-00595-2>
- [22]- A. Younes, E. Elbeltagi, A. Diab, G. Tarsi, F. Saeed, C. Sangiorgi, Incorporating coarse and fine recycled aggregates into concrete mixes : mechanical characterization and environmental impact Life cycle Inventory. *J. Mater. Cycles Waste Manag.* 26(1) (2024) 654–668. <https://doi.org/10.1007/s10163-023-01834-1>
- [23]- A. Benidir, Assessment of Concrete Compressive Strength by Destructive Testing: Influence of Strength Correction Factors and Size Effect. *Adv. Sci. Technol. Res. J.* 159 (2025) 13–18. <https://doi.org/10.4028/p-96LXin>
- [24]- T. Joshi, P. Mathur, P. Oza, S. Agrawal, H. Narmawala, Predictive modelling of sustainable concrete compressive strength using advanced machine learning algorithms. *Adv. Civ. Archit. Eng.* 15(29) (2024) 168–192. <https://doi.org/10.13167/2024.29.11>
- [25]- B. Omer, D.K.I. Jaf, A. Abdalla, A.S. Mohammed, P.I. Abdulrahman, R. Kurda, Advanced modeling for predicting compressive strength in fly ash-modified recycled aggregate concrete: XGboost, MEP, MARS, and ANN approaches. *Innov. Infrastruct. Solut.* 9(3) (2024) 61. <https://doi.org/10.1007/s41062-024-01365-0>
- [26]- A. Ravikar, Comparative Evaluation of Machine Learning Predictions for Concrete Properties, *Adv. Nonlinear Var. Inequalities* 28(2) (2024) 256–266. <https://doi.org/10.52783/anvi.v28.1968>
- [27]- S. Elhishi, A.M. Elashry, S. El-Metwally, Assessing Concrete Strength Using Machine Learning. In: *Proceedings of the 2023 2nd International Conference on Communication Engineering and Computing (ICCEC)*, IEEE, 2023, pp. 1–6. <https://doi.org/10.1109/CAISAI59399.2023.10270023>
- [28]- M. Imran, H. Amjad, S. Ali, Machine learning assisted prediction of the mechanical properties of carbon nanotube incorporated concrete. *Struct. Concr.* 26(2) (2025) 1613–1635. <https://doi.org/10.1002/suco.202400727>
- [29]- M.K. Elshaarawy, M.M. Alsaadawi, A. Hamed, Machine learning and interactive GUI for concrete compressive strength prediction. *Dent. Sci. Reports* 14(1) (2024) 16694. <https://doi.org/10.1038/s41598-024-66957-3>
- [30]- S.J. Alghamdi, Prediction of Concrete's Compressive Strength via Artificial Neural Network Trained on Synthetic Data. *Eng. Technol. Appl. Sci. Res.* 13(6) (2023) 12404–12408. <https://doi.org/10.48084/etasr.6560>
- [31]- S.S. Othman, Z.Q. Ahmed, A.S. Mohammed, A.M.T. Hassan, Soft computing-based prediction of recycled aggregate concrete strength: impact of aggregate properties and mix design parameters. *Asian J. Civ. Eng.* 26(6) (2025) 2349–2369. <https://doi.org/10.1007/s42107-025-01312-z>
- [32]- A. Ravikar, Comparative Evaluation of Machine Learning Predictions for Concrete Properties. *Adv. Nonlinear Var. Inequalities* 28(2) (2024) 256–266. <https://doi.org/10.52783/anvi.v28.1968>
- [33]- A.K. Jha, Forecasting the Properties of Concrete Employing Experimental Data Using Machine Learning Algorithms. *Eur. J. Theor. Appl. Sci.* 2(3) (2024) 259–266. [https://doi.org/10.59324/ejtas.2024.2\(3\).22](https://doi.org/10.59324/ejtas.2024.2(3).22)
- [34]- M.M. Moein, A. Saradar, K. Rahmati, S.H.G. Mousavinejad, J. Bristow, V. Aramali, M. Karakouzian, Predictive models for concrete properties using machine learning and deep learning approaches: A review. *J. Build. Eng.* 63(Part B) (2023) 105470. <https://doi.org/10.1016/j.job.2022.105444>
- [35]- M. Shahab, M. Shakeel, L. Gulaly, M. Khan, Predicting the compressive strength of self-compacting concrete containing recycled aggregate and supplementary cementitious materials using machine learning techniques. *Prog.*

- Eng. Sci. 2(2) (2025) 100077. <https://doi.org/10.1016/j.pes.2025.100077>
- [36]- S. Aggarwal, R. Singh, A. Rathore, K. Kapoor, M. Patel, A novel data-driven machine learning techniques to predict compressive strength of fly ash and recycled coarse aggregates based self-compacting concrete. *Mater. Today Commun.* 39 (2024) 109294. <https://doi.org/10.1016/j.mtcomm.2024.109294>
- [37]- S.A.R. Shah, N.B. Kahla, M. Atig, M.K. Anwar, M. Azab, A. Mahmood, Optimization of fresh and mechanical properties of sustainable concrete composite containing ARGF and fly ash: An application of response surface methodology. *Constr. Build. Mater.* 362(2) (2023) 129722. <https://doi.org/10.1016/j.conbuildmat.2022.129722>
- [38]- M. Alizamir, M. Wang, R.M.A. Ikram, A. Gholampour, K.O. Ahmed, S. Heddami, S. Kim, An Interpretable XGBoost-SHAP Machine Learning Model for Reliable Prediction of Mechanical Properties in Waste Foundry Sand-Based Eco-Friendly Concrete. *Results Eng.* 25 (2025) 104307. <https://doi.org/10.1016/j.rineng.2025.104307>
- [39]- H.Q. Nguyen, D.D.K. Nguyen, T.D. Le, A. Mai, K.T. Huynh, Career Path Prediction Using XGBoost Model and Students' Academic Results. *CTU J. Inn. Sus. Dev.* 15(4) (2023) 62–75. <https://doi.org/10.22144/ctujoisd.2023.036>
- [40]- B. Samreen, B. Farheen, A. Syed Aqeel, Study on supplementary cementitious materials for sustainable development of concrete. *I-Manager's J. Mater. Sci.* 10(4) (2022) 31. <https://doi.org/10.26634/jms.10.1.18906>
- [41]- Y. Wu, Y. Zhou, Hybridizing Grid Search and Support Vector Regression to Predict the Compressive Strength of Fly Ash Concrete. *Adv. Civ. Eng.* 330(2022) 127298. <https://doi.org/10.1016/j.conbuildmat.2022.127298>
- [42]- L. Wang, L. Liu, D. Dai, B. Liu, Z. Cheng, Predictive Modeling of UHPC Compressive Strength: Integration of Support Vector Regression and Arithmetic Optimization Algorithm. *Appl. Sci.* 14(17) (2024) 8083. <https://doi.org/10.3390/app14178083>
- [43]- I.B. Mustapha, M. Abdulkareem, T.M. Jassam, A.H. AlAteah, K.A. Alawi Al-Sodani, M.M.H. Al-Tholaia, H. Nabus, S.C. Alih, Z. Abdulkareem, A. Ganiyu, Comparative Analysis of Gradient-Boosting Ensembles for Estimation of Compressive Strength of Quaternary Blend Concrete. *Int. J. Concr. Struct. Mater.* 18(1) (2024) 43. <https://doi.org/10.1186/s40069-023-00653-w>
- [44]- B. Halder, I. Ahmadianfar, S. Heddami, Z.H. Mussa, L. Goliatt, M.L. Tan, Z. Sa'adi, Z. Al-Khafaji, N. Al-Ansari, A.H. Jawad, Z.M. Yaseen, Machine learning-based country-level annual air pollutants exploration using Sentinel-5P and Google Earth Engine. *Sci. Rep.* 13(1) (2023) 7968. <https://doi.org/10.1038/s41598-023-34774-9>
- [45]- P. Yin, B. Pan, Z. Li, Quantitative Evaluation on Fusion Behavior of Rejuvenator and Aged Asphalt Based on Multi-scale Experimentss. *J. Test. Eval.* 51(6) (2023) 4340–4359. <https://doi.org/10.1520/JTE20230125>
- [46]- V.Q. Tran, V.Q. Dang, L.S. Ho, Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Constr. Build. Mater.* 323 (2022) 126578. <https://doi.org/10.1016/j.conbuildmat.2022.126578>
- [47]- S.B. Rahman, M.K. Hassan, A.S. Mohammed, Machine-aided regression modeling for green concrete mix optimization with fly ash and recycled aggregates. *Asian J. Civ. Eng.* 26(15) (2025) 3843–3863. <https://doi.org/10.1007/s42107-025-01402-y>
- [48]- M. Zupan, J. Coolen, N. Mavraki, S. Degraer, T. Moens, F. Kerckhof, L.L. Lopez, J. Vanaverbeke, Life on every stone: Characterizing benthic communities from scour protection layers of 2 offshore wind farms in the southern North Sea. *J. Sea Res.* 201(2024) 102522. <https://doi.org/10.1016/j.seares.2024.102522>
- [49]- C.L. Fan, Deep neural networks for automated damage classification in image-based visual data of reinforced concrete structures. *Heliyon* 10(19) (2024) e38104. <https://doi.org/10.1016/j.heliyon.2024.e38104>
- [50]- D.S. Rocha, T.A. Baldo, H.A. Silva-Neto, G.F. Duarte-Junior, G.S. Bazílio, C.L. Borges, J.A. Parente-Rocha, W.R. de Araujo, A. de Siervo, T.L.R.C. Paixão, W.K.T. Coltro, Disposable and eco-friendly electrochemical immunosensor for rapid detection of SARS-CoV-2. *Talanta* 268(Part 1) (2024) 125337. <https://doi.org/10.1016/j.talanta.2023.125337>
- [51]- M. Bhagat, B. Bakariya, A Comprehensive Review of Cross-Validation Techniques in Machine Learning. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* 16(2) (2025) 345–351.
- [52]- S.C. Sapkota, S. Sapkota, G. Saini, Prediction of Split Tensile Strength of Recycled Aggregate Concrete Leveraging Explainable Hybrid XGB with Optimization Algorithm. *J. Build. Eng.* 7(2024) 4343–4359. <https://doi.org/10.1007/s41939-024-00480-w>
- [53]- R. Ali, M. Muayad, A.S. Mohammed, P.G. Asteris, Analysis and prediction of the effect of Nanosilica on the compressive strength of concrete with different mix proportions and specimen sizes using various numerical approaches. *Struct. Concr.* 24(5) (2022) 4161–4184. <https://doi.org/10.1002/suco.202200718>