Article

Not peer-reviewed version

Eco-Innovation in Construction: Forecasting Natural Fibre-Reinforced Concrete Strength Using Machine Learning

Hussein Hamel Zghair * , Iman Kattoof Harith , Tholfekar habeeb Hussain

Posted Date: 16 March 2026

doi: 10.20944/preprints202603.1162.v1

Keywords: sustainable concrete; natural fibre; recycled coarse aggregate; machine learning; prediction model

Article

Eco-Innovation in Construction: Forecasting Natural Fibre-Reinforced Concrete Strength Using Machine Learning

Hussein Hamel Zghair [1,*], Iman Kattoof Harith [2] and Tholfekar Habeeb Hussain [3]

[1] Civil Engineering Depart., University of Technology, Iraq
[2] Civil Engineering Department, College of Engineering, Al-Qasim Green University, Babylon, Iraq
[3] Water Resources Management Engineering Department, College of Engineering, Al-Qasim Green University, Babylon
* Correspondence: 11539@uotechnology.edu.iq

Abstract

Conventional concrete faces limitations such as brittleness, low energy absorption, and a significant environmental impact due to its reliance on natural resources. Integrating natural fibers (NF) and recycled coarse aggregate (RCA) into concrete presents a promising avenue for enhancing both performance and sustainability. However, accurately predicting the strength of these innovative concrete mixtures remains challenging. This study investigates the predictive capabilities of two machine learning (ML) models: Classification and Regression Trees (CART) and Stepwise Polynomial Regression (SPR), in forecasting the compressive and splitting tensile strength of NF-reinforced concrete incorporating RCA. Results unequivocally demonstrate the superior predictive accuracy of the CART model. CART exhibited significantly higher R-squared values and lower error metrics (RMSE, MAD, MAPE, MSE) for both compressive and splitting tensile strength. For compressive strength CART achieved $R^2 = 0.91$, RMSE = 5.5686, MSE = 31.0098, MAD = 4.1076, and MAPE = 0.1055, while for splitting tensile strength, it achieved $R^2 = 0.89$, RMSE = 0.3954, MSE = 0.1563, MAD = 0.2996, and MAPE = 0.0939. These findings underscore the significant potential of ML, particularly CART, in optimizing the design of sustainable concrete structures by enabling more precise and efficient strength predictions, ultimately contributing to more sustainable and resilient infrastructure.

Keywords: sustainable concrete; natural fibre; recycled coarse aggregate; machine learning; prediction model

1. Introduction

The construction industry is under increasing pressure to adopt sustainable practices to mitigate the environmental impact of traditional building materials, which contribute significantly to resource depletion and climate change. This urgency has driven the exploration of eco-friendly alternatives such as natural fibers (e.g., coir, bamboo, sisal) and recycled concrete aggregates (RCA). Natural fibers, derived from renewable resources (Figure 1), offer advantages like high specific strength, low density, and biodegradability, enhancing concrete's tensile strength, impact resistance, and energy absorption while reducing reliance on petroleum-based materials [1–3]. Concurrently, RCA repurposed from demolished structures addresses construction waste challenges, conserving natural resources and reducing landfill burdens [4]. However, integrating these materials into concrete introduces complexities. RCA's adhered mortar increases porosity, weakening the interfacial transition zone and compromising mechanical properties [5–10], while natural fibers require precise dosing to balance performance and workability [11].

a. Coir Fiber b) Bamboo Fiber c) Ramie fiber

d) Jute Fiber e) Kenaf Fiber f) Sisal fiber

Figure 1. Natural fibres a) coir fibre [29], b) bamboo fibre [30], c) ramie fibre [31], d) jute fibre [32], e) kenaf fibre [33] and f) sisal fibre [34].

Machine learning (ML) has emerged as a critical tool for optimizing these sustainable composites. Recent studies highlight its potential but reveal critical limitations. For example, gradient boosting achieved high accuracy ($R^2 = 0.91$) in predicting compressive strength of recycled concrete powder (RCP)-modified concrete but focused solely on cement replacement [12]. Artificial neural networks (ANNs) demonstrated strong training performance (R^2: 0.93–0.99) for RCA concrete but suffered from overfitting, yielding unreliable testing results (R^2: 0.67–0.78) [13]. Random forest (RF) models excelled in dual-target predictions (compressive and tensile strength) but lacked interpretability for engineering applications [14]. Linear approaches like stepwise polynomial regression (SPR) struggled with multicollinearity among predictors, inflating variable significance [15–17], while hybrid models (e.g., GA-ANN) and machine learning architectures (e.g., ANNs) faced computational or data impracticality [18,19]. Numerous recent studies have incorporated machine learning (ML) into civil engineering applications [20–22].

While the incorporation of sustainable materials in concrete has been widely investigated, achieving an optimal balance between performance and practical applicability remains a key challenge. Predictive modeling approaches for RCA concrete have progressed from early empirical formulations [23–28]to sophisticated machine learning techniques.

As outlined in Table 1, previous studies exhibit significant limitations. To address these gaps, this study employs CART and SPR machine learning approaches. CART provides transparent decision rules (e.g., RCA ≤ 40% thresholds) and ranks predictors (e.g., water-to-binder ratio dominance), while SPR refines nonlinear terms via backward elimination. Dual-target predictions of compressive and splitting tensile strength capture material interdependencies, validated rigorously through 10-fold cross-validation. This framework balances accuracy, interpretability, and robustness, advancing the design of sustainable concrete systems.

 doi:10.20944/preprints202603.1162.v1

Table 1. Key Findings and Limitations of Previous Studies.

Study	Methodology	Key Findings	Limitations
Manan et al. [12]	Gradient Boosting	Achieved $R^2 = 0.91$ for RCP-modified concrete	Ignored fiber-reinforced systems
Manan et al. [13]	Artificial Neural Networks (ANNs)	High training accuracy (R^2: 0.93–0.99)	Overfitting led to poor testing performance (R^2: 0.67–0.78)
Manan et al. [14]	Random Forest (RF)	Accurate dual-target predictions	Lack of interpretability for engineering use
Kattoof et al. [15]	Stepwise Polynomial Regression (SPR)	Managed nonlinear interactions	Struggled with multicollinearity among predictors
González-Taboada et al. [23]	Genetic Programming	Predicted mechanical properties of RCA concrete	Limited to single-property predictions
Behnood et al. [27]	M5P Model Tree	Predicted modulus of elasticity for RCA concrete	Narrow focus on RCA, excluded NF synergy
Current Study	Natural fiber (coir, bamboo, sisal) + RCA concrete	CART + SPR	Dual-target predictions ($R^2 = 0.91$ compressive, 0.89 tensile); transparent variable importance via CART; resolved multicollinearity via SPR.

2. Research Methodology

The methodology utilized 406 experimental datasets to evaluate compressive and splitting tensile strength of natural fiber-reinforced concrete with recycled aggregates. Nine variables were analyzed via statistical measures (mean, skewness, kurtosis) and histograms, supported by a Pearson correlation matrix identifying key relationships. Two machine learning models were developed: Stepwise Polynomial Regression (SPR) with backward elimination ($p<0.05$) for significant second-order terms, and Classification and Regression Tree (CART) pruned ($\alpha=0.01$) with node size ≥3. Generalizability was tested via 10-fold cross-validation, using R2, RMSE, MAD, MAPE, and MSE. The research methodology flowchart is presented in Figure 2.

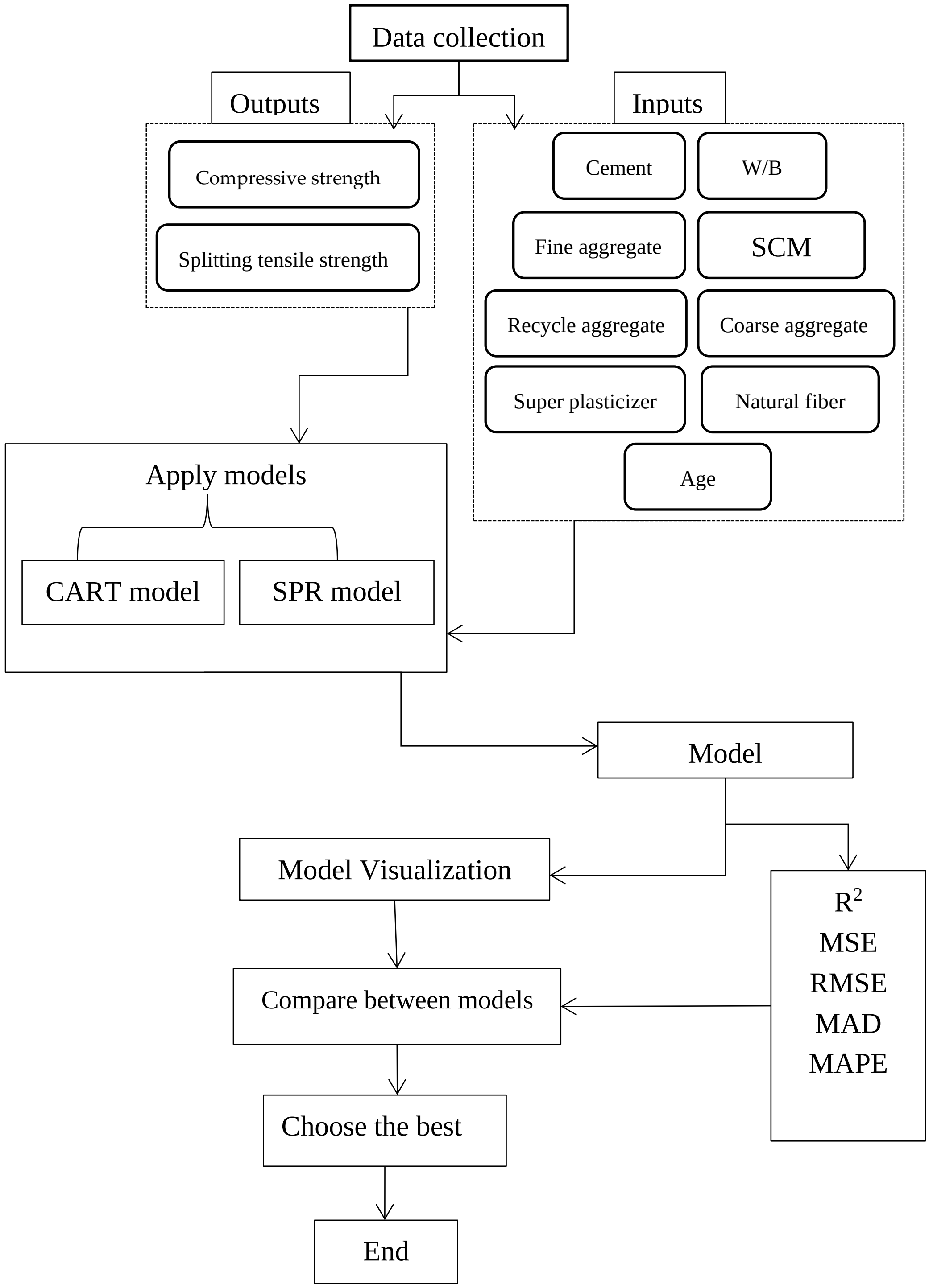

Figure 2. Research methodology.

2.1. Data Preparation and Distribution

To derive a robust model, a wide-ranging datasets was collected by leveraging the findings of previous research, including studies by [4,35–44]. This datasets includes 406 experimental results, each covering measurements of cement (C), water to binder ratio (W/B), fine aggregate (FA), coarse aggregate (CA), recycle coarse aggregate (RCA), supplementary cementitious materials (SCM), super plasticizer (SP), natural fibre (NF), age, and the corresponding compressive strength (f_{cu}) and splitting tensile strength (f_t) of sustainable concrete.

The natural fibre includes bamboo fibre, sisal Fibre, coir Fibre, jute fibre, ramie fibres, kenaf fibre. Table 2 presents basic statistics for these key variables. It includes information on the range of values (minimum and maximum), measures of central tendency (mean), and measures of variability

doi:10.20944/preprints202603.1162.v1

(coefficient of variation). Additionally, Table 1 provides insights into the shape of the distributions (kurtosis and skewness), which characterize the data distribution beyond simple measures of central tendency and variability. Notably, the raw dataset (n = 406) was used without pre-processing (e.g., normalization, outlier removal), ensuring the models reflected real-world variability.

Table 2. Basic statistics of variables for scour depth prediction.

Variable	Abbr.	Mean	CoefVar	Minimum	Maximum	Skewness	Kurtosis
Cement	C	445.025	13.25	243	513.7	-1.27	1.42
W/B Ratio	W/B	0.417827	20.63	0.277228	0.55	-0.10	-1.09
Fine Aggregate	FA	597.192	16.69	470.3	784	0.30	-1.49
Coarse Aggregate	CA	1126.48	11.19	898	1420	0.09	-0.56
RCA	RCA	39.6429	119.50	0	100	0.44	-1.75
SCM	SCM	13.7744	180.76	0	153.9	2.15	6.69
Superplasticizer	SP	4.17523	159.79	0	20.52	1.07	-0.78
Natural fibre	NF	0.332512	194.84	0	3	2.13	4.07
Age	Age	30.1675	108.59	1	180	1.99	4.88
Cube- compressive strength	f_{cu}	42.3838	44.06	9.74726	96.7	0.94	0.31
Splitting tensile strength	f_t	3.48685	35.14	1.16813	6.2	0.25	-0.74

The histograms in Figure 2 reveal a diverse range of distribution shapes. Some variables exhibit approximately normal distributions, characterized by bell-shaped curves, while others display skewed distributions, either positively or negatively. The mean values, indicated by vertical red lines, provide an estimate of the central tendency for each variable within each case. The shapes of the distributions themselves offer valuable insights into the underlying data. For example, a skewed distribution might suggest the presence of outliers or indicate that the data is not evenly distributed.

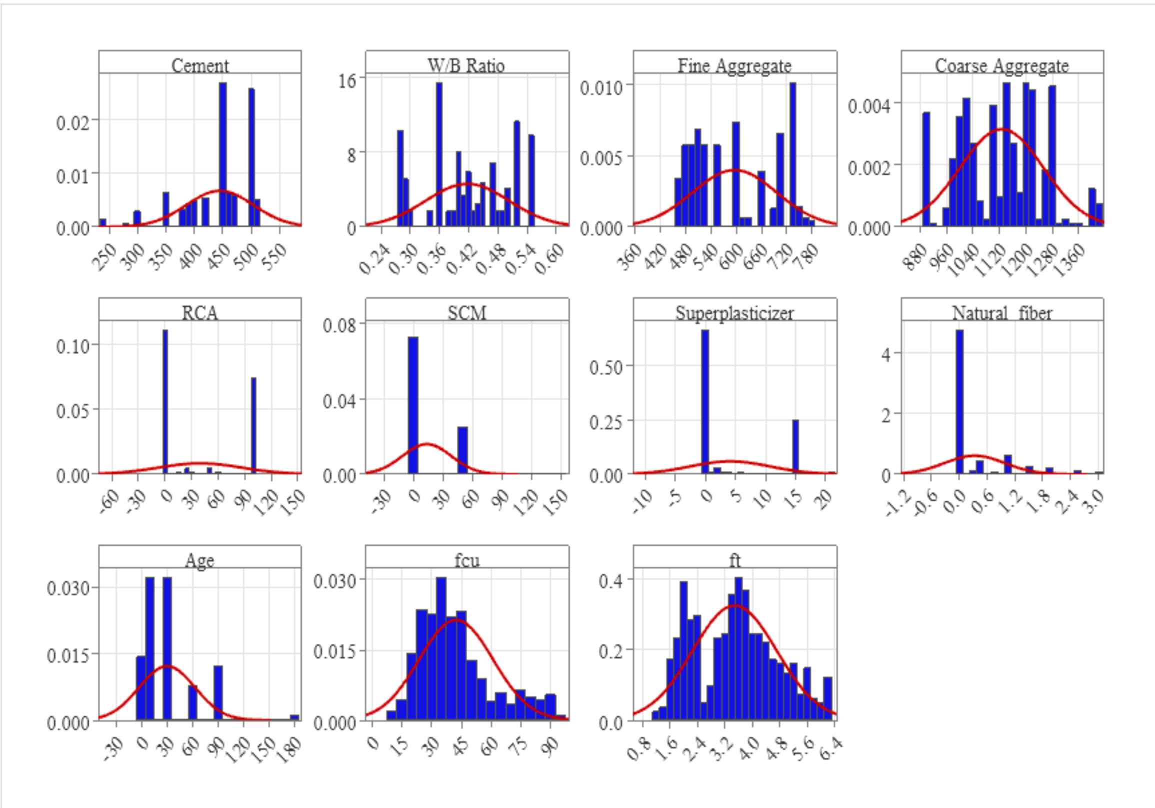

Figure 2. Histograms of input and output variables for sustainable concrete.

2.2. Correlation Matrix

Correlation matrices are statistical tools that visualize relationships between multiple variables in a dataset. These square, symmetrical matrices display correlation coefficients (often Pearson's R) for each variable pair. Pearson's R measures the strength and direction of linear relationships, ranging from -1 (perfect negative) to +1 (perfect positive). A value of 0 indicates no linear correlation. Importantly, Pearson's R may not accurately reflect non-linear relationships. Correlation matrices provide a concise overview of interdependencies within a dataset, helping researchers understand which variables are related and how strongly [45].

Figure 3 presents the Pearson correlation matrix between the input variables and the compressive strength (f_{cu}). Significant linear relationships are identified, including strong negative correlations for water-to-binder ratio (W/B, R=−0.76) and fine aggregate content (FA, R=−0.53), and positive correlations for supplementary cementitious materials (SCM, R=+0.54) and superplasticizer content (SP, R=+0.69). The notably high correlation for SP likely reflects its functional role in reducing water demand and enhancing workability, which indirectly improves compaction and strength development. Conversely, the moderate correlation for SCM may arise from dosage-dependent effects or synergistic interactions with other binder components, which will be further explored through nonlinear modeling.

These robust correlations raise concerns about multicollinearity among predictors, potentially inflating the perceived influence of individual variables in linear analyses. To address this, predictive methodologies for estimating f_{cu} and tensile strength (f_t) will employ machine learning techniques (e.g., regularization, sensitivity analysis) to disentangle complex relationships and quantify variable contributions more robustly. While f_t exhibits weaker linear associations with the input variables, the comparative analysis of compressive and tensile behavior will clarify how material factors differentially govern these mechanical properties. This dual approach ensures a balanced interpretation of variable impacts, particularly for SP and SCM, beyond correlation-based conclusions.

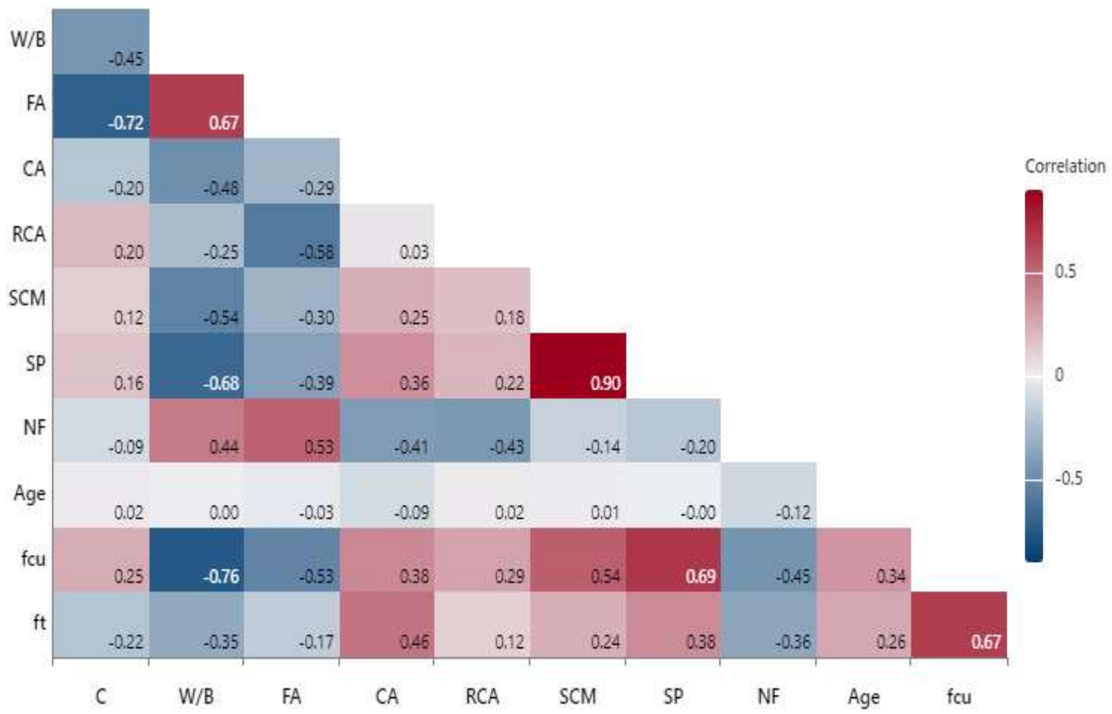

Figure 3. Correlation matrix of input variables and output variables.

 doi:10.20944/preprints202603.1162.v1

2.3. Regression & ML Models Assumptions and Limitations

2.3.1. Stepwise Polynomial Regression Model

Stepwise Polynomial Regression (SPR) is a statistical technique for building fatal models. It iteratively refines the model by strategically add or remove predictor variable based on their statistical augury. Starting with a basic model, SPR gradually improves its accuracy by identifying the most effective factors. This process can involve forward selection (adding variables), backward elimination (removing variables), or a combination of both. potential limitation of these approaches is their limited ability to detect and incorporate interactions between different predictor variables [46]. In the SPR process, a second order polynomial model is employed, incorporating nine independent variables and one dependent variable (y).

$$y = b0 + \sum bi\ xi + \sum \alpha ij\ xi\ xj + \sum \alpha ii\ x_i^2 \tag{1}$$

Where y denotes expected value of the response variable b_0 denotes the constant term b_i stands for the coefficient of the independent variable x.

Numerous studies have discovered the application of SPR in predicting concrete performance as evidence by [15,47]. The primary strength of SPR compare to other techniques lies in its automated feature selection and model building process.

The SPR model incorporated:

- Polynomial Order: Restricted to second-degree terms to avoid overfitting while capturing nonlinear relationships.
- Term Selection: A p-value threshold of 0.05 was used to retain statistically significant interactions and quadratic terms.
- Validation: 10-fold cross-validation confirmed generalizability.

2.3.2. Classification and Regression Tree (CART) Model

Classification and Regression Trees (CART) is a powerful machine learning method for constructed predictive models. This approach utilizes decision trees to classify categorical outcomes and predict continuous values. CART models are built by initially growing a large tree and then pruning it back to improve simplicity and prevent overfitting. For categorical targets the Gini index is used to measured impurity within each node of tree. For continuous targets the least squares deviation method is employed to determine optimal splits [48,49].

Numerous studies have examine the application of CART in predicted concrete performance as evidence by [50–52]. A key advantage of CART lies in its ability to automatically identify the most influential predictors and determine the optimal threshold values for each predictor to effectively classifying target variable.

The CART model was tuned using the following hyperparameters:

- Pruning (α): Cost-complexity pruning with α = 0.01 was applied to balance tree depth and overfitting.
- Terminal Node Size: A minimum of 3 observations per node was enforced to ensure meaningful splits.
- Splitting Criterion: The least squares deviation method minimized variance in terminal nodes for regression tasks.
- Validation: The optimal subtree was selected via 10-fold cross-validation, which minimized RMSE while preserving predictive power.

2.4. K-Fold Cross-Validation Technique

Cross validation (CV) is a crucial technique for evaluated machine learning (ML) model performance and ensure their ability to generalize to unseen data. It helps mitigate bias and overfitting by dividing the dataset into subsets training the model on a portion and then evaluating its performance on the remaining data [53].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1162.v1

K-fold CV is a widely used method where the dataset divides into 'k' equal sized folds. The model is trained on (k-1) folds and evaluate on the remaining fold. This process is repeated (k) times with each fold serving as the validation set once. Common performance metrics such as Root Mean Squared Error (RMSE) and R-squared are used to assess model performance [54]. 10-fold CV is often considered a good balance between reliable variance estimation and computational efficiency [55].

In this research 10 fold CV was employed on a 406 instance dataset to assess model performance. The model with the lowest average RMSE across the 10 iterations was selected as the best performing model, Figure 4.

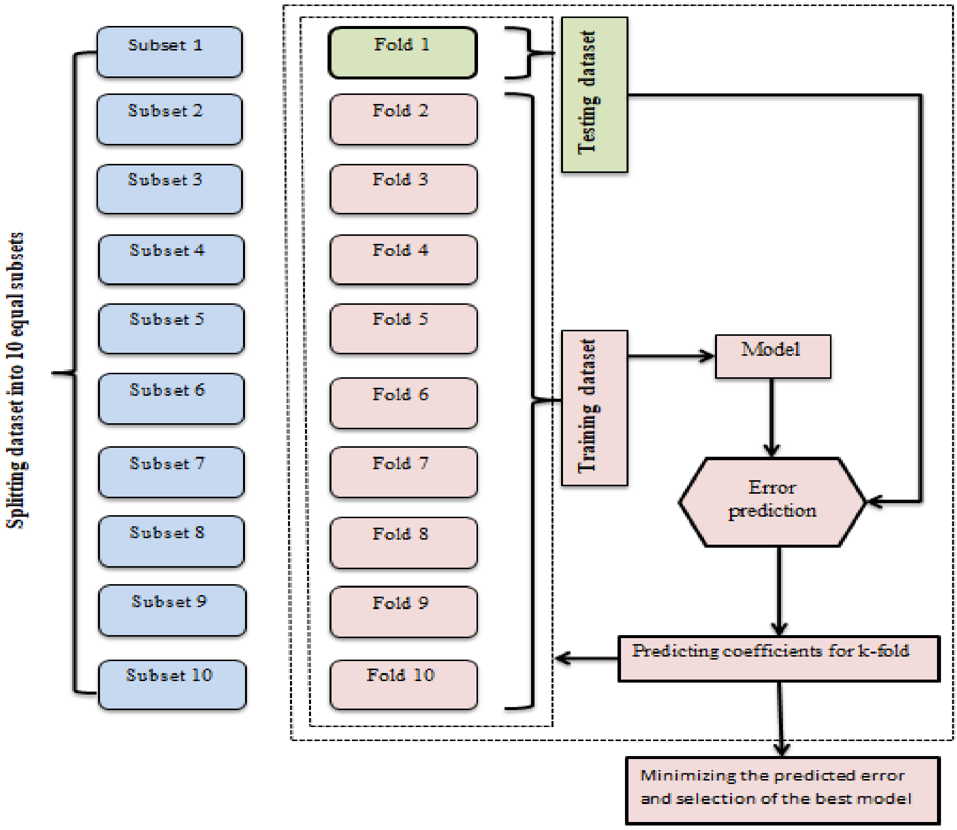

Figure 4. K-fold cross validation technique procedures [55].

2.5. Criteria for Evaluating Models

To evaluate the predictive capabilities of current models it is crucial to employ a diverse set of metrics [16,47]. These metrics for example mean square error (MSE) root mean square error (RMSE) mean absolute deviation (MAD) and mean absolute percentage error (MAPE) assess the discrepancy between model predictions and actual observations. Lower values in these metrics indicate better model performance. Other metrics including R squared (R^2) adjusted R squared and predicted R squared evaluate the model's ability to capture the trends observed in the real world data. By utilizing Equations (2)–(6), a comprehensive assessment of the predictive accuracy of existing models can be achieved.

$$R^2 = \left(\frac{\sum_{i=1}^{n} \left(x_i - x_{i(m)}\right)\left(y_i - y_{i(m)}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_i - x_{i(m)}\right)^2 \sum_{i=1}^{n} \left(y_i - y_{i(m)}\right)^2}} \right)^2 \quad (2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4}$$

$$MAD = (\sum \ |y_i - x_i \ |) \ /n \tag{5}$$

$$MAPE = 1/n \sum((y_i - x_i)/y_i) * 100 \tag{6}$$

n: data points, x_i: foresee values, y_i: Actual values, m: Average data points rats.

3. Findings and Interpretation

3.1. Stepwise Polynomial Regression Model

3.1.1. Compressive Strength

A Stepwise Polynomial Regression (SPR) model was developed to predict concrete strength, beginning with a full second-order polynomial model (Equation (1)) to account for potential nonlinear relationships while avoiding overcomplexity. The model was iteratively refined using backward elimination, where terms with statistically insignificant contributions (*p-value > 0.05*) were systematically removed until only significant predictors (*p-value < 0.05*) remained. This approach ensured a balance between interpretability and predictive power: restricting the polynomial to second-degree terms prevented overfitting, while the p-value threshold streamlined the model by excluding redundant interactions. The final parsimonious model retained nine independent variables (Equation (7)), achieving strong generalizability as validated by 10-fold cross-validation ($R^2 = 0.86$–0.89). This optimized structure not only reduced computational complexity but also preserved critical relationships between input parameters, such as the synergistic effects of superplasticizer and water-to-binder ratio, which significantly influenced compressive strength ($p < 0.001$). By rigorously adhering to statistical and domain-driven criteria, the SPR model effectively captured the mechanistic behavior of sustainable concrete while maintaining transparency for engineering applications.

$$f_{cu} = -3.5 + 0.2212\, C - 52.86\, W/B + 2.696\, SP + 26.09\, NF + 0.698\, Age - 0.000218\, C * C + 1.651\, NF * NF - 0.002384\, Age * Age - 0.0356\, C * NF - 0.000498\, C * Age - 4.39\, W/B * SP - 33.1\, W/B * NF - 1.188\, SP * NF + 0.00619\, SP * Age - 0.0928\, NF * Age \tag{7}$$

Equation (7) creates a nonlinear formula between the compressive strength and the examined variables. The formula framework showed in Equation (7) demonstrates that a majority of the examined variables significantly influence the compressive strength.

Table 3 reveals that the linear terms of cement, water-to-binder ratio (W/B), superplasticizer, natural fiber, and age significantly influence the response variable, with P-values below 0.05%. Conversely, the quadratic terms of most variables exhibit insignificant effects on the response variable, as indicated by P-values exceeding 0.05%, leading to their removal from the model. Notably, the quadratic terms for cement, natural fiber, and age demonstrate a statistically significant effect on the response variable, suggesting these variables independently exert substantial effect on the response. Furthermore, interactions between most variables exhibit statistical significance for the response variable, with P-values below 0.05% [56,57].

Table 3. ANOVA results of SPR model.

Term	Coef	P-Value
Constant	-3.5	0.806
C	0.2212	0.002
W/B	-52.86	0.000

doi:10.20944/preprints202603.1162.v1

SP	2.696	0.000
NF	26.09	0.007
Age	0.698	0.000
C*C	-0.000218	0.019
NF*NF	1.651	0.040
Age*Age	-0.002384	0.000
C*NF	-0.0356	0.007
C*Age	-0.000498	**0.103**
W/B*SP	-4.39	0.000
W/B*NF	-33.1	0.029
SP*NF	-1.188	0.000
SP*Age	0.00619	0.000
NF*Age	-0.0928	0.003
Statistical metrics errors		
R^2	87.34%	
Adj. R^2	86.85%	
Pred. R^2	86.45%	
10-fold R^2	86.05%	

To assess the model's ability to generalize to unseen data, the coefficient of determination (R^2) was calculated. The R^2 values of 87.34%, with adjusted R^2 of 86.85% and predicted R^2 of 86.45%, indicate strong predictive performance. 10-fold R^2 was 86.05%, which indicates good predictive performance.

Figure 5 displays the normal probability plot for residuals of the compressive strength model. While the central tendency of residuals aligns closely with the reference line (indicating approximate normality), deviations in the upper and lower tails suggest mild non-normality. These deviations reflect inherent variability in the material system, particularly in extreme strength values influenced by stochastic factors such as localized microcracking or heterogeneous binder distribution.

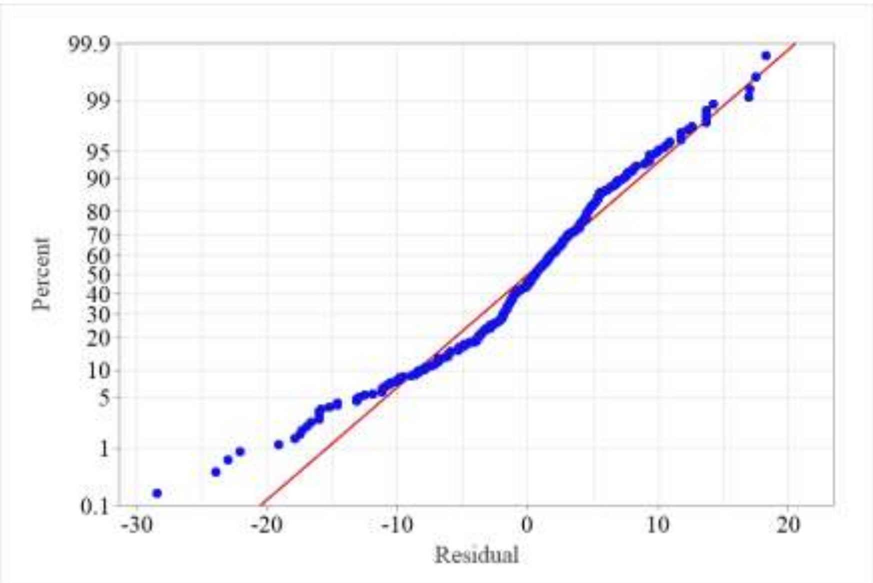

Figure 5. Normal Probability Plot for compressive strength of sustainable concrete.

The deterministic regression framework assumes normality, which may underestimate uncertainty in tail regions. In contrast, stochastic analyses (e.g., probabilistic modeling) inherently capture this variability, leading to broader confidence intervals and more conservative strength predictions. To reconcile these differences, future iterations of the model will employ robust regression techniques or data transformations (e.g., Box-Cox) to better accommodate non-normal residuals. This dual perspective ensures that both deterministic predictions and stochastic variability are rigorously addressed, enhancing the model's reliability for practical applications.

Figure 6 shows a "Versus Fits" plot, which visualizes the residuals (errors) of a model. The model predicts the f_{cu} of sustainable concrete. The horizontal axis displays the fitted values, or the values predicted by the model for each data point. The vertical axis signifies the residuals, which are the difference between the actual detected f_{cu} and the corresponding fitted values. The position of a dot indicates the relationship between its fitted value and its residual. Ideally, the residuals should be randomly scattered around the horizontal line at zero. This would suggest that the model is a suitable for the dataset, with no discernible patterns or trends in the errors. In this plot, the residuals appear to be randomly distributed around the zero line. There is no clear evidence of systematic patterns or trends. Based on this visual inspection, the "Versus Fits" plot suggests that the model used to predict f_{cu} is likely a good fit for the data. The random distribution of residuals indicates that the model's predictions are reasonably accurate.

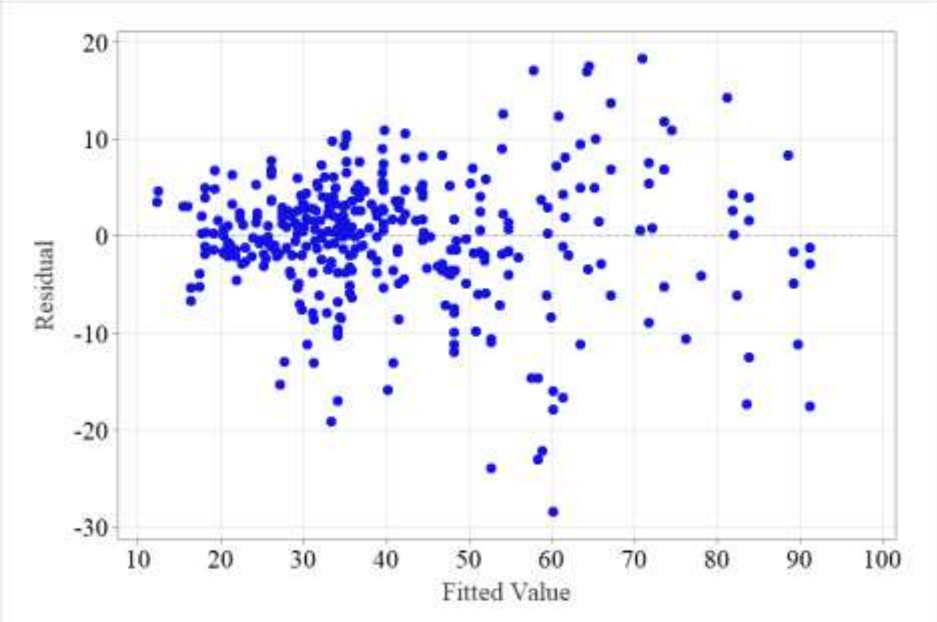

Figure 6. Versus fits plot on compressive strength of sustainable concrete.

3.1.2. Splitting Tensile Strength

A stepwise polynomial regression approach was used to develop the model, starting with a second-order polynomial (Equation (1)). Terms with insignificant effects ($p > 0.05$) were removed iteratively, resulting in a parsimonious model with nine variables (Equation (8)). This model effectively represents the relationships between the variables.

$$
\begin{aligned}
f_t = {} & 113.2 - 0.1655\,C + 204.7\,W/B - 0.0816\,FA - 0.1939\,CA \\
& + 0.2185\,RCA - 4.332\,SCM + 15.24\,SP - 0.23\,NF \\
& + 0.0269\,Age - 44.2\,W/B * W/B + 0.000072\,FA * FA \\
& + 0.000048\,CA * CA - 0.000235\,RCA * RCA - 0.1726\,NF * NF \\
& - 0.000145\,Age * Age - 0.1787\,C * W/B + 0.000128\,C * FA \\
& + 0.000156\,C * CA - 0.000073\,C * Age - 0.1835\,W/B * FA \\
& + 0.0355\,W/B * CA - 0.2811\,W/B * RCA + 4.27\,W/B * SCM \\
& - 14.12\,W/B * SP - 7.09\,W/B * NF + 0.000023\,FA * CA \\
& - 0.000922\,FA * SCM + 0.00571\,FA * NF - 0.000062\,CA * RCA \\
& + 0.003071\,CA * SCM - 0.00928\,CA * SP + 0.000030\,CA * Age \\
& - 0.00377\,RCA * SCM + 0.00894\,RCA * SP + 0.000048\,RCA \\
& * Age
\end{aligned}
\tag{8}
$$

Equation (8) creates a nonlinear formula between the splitting tensile strength and the examined variables. This formula framework, as showed in Equation (8), demonstrates that all of the examind variables significantly influence the response variable. As presented in Table 4, the linear terms of (cement, W/B ratio, coarse aggregate, recycle coarse aggregate, supplementary cementitious materials, super plasticizer, and age) variables significantly affect the f_t, with P-values less than 0.05%.

Table 4. ANOVA results of SPR model.

Term	Coef	P-Value	
Constant	113.2	0.000	Significant
C	-0.1655	0.000	Significant
W/B	204.7	0.000	Significant
FA	-0.0816	**0.058**	**insignificant**
CA	-0.1939	0.000	Significant
RCA	0.2185	0.000	Significant
SCM	-4.332	0.000	Significant
SP	15.24	0.000	Significant
NF	-0.23	**0.832**	**insignificant**
Age	0.0269	0.050	Significant
W/B*W/B	-44.2	0.006	Significant
FA*FA	0.000072	0.000	Significant
CA*CA	0.000048	0.000	Significant
RCA*RCA	-0.000235	0.000	Significant
NF*NF	-0.1726	0.006	Significant
Age*Age	-0.000145	0.000	Significant
C*W/B	-0.1787	0.000	Significant
C*FA	0.000128	0.001	Significant
C*CA	0.000156	0.000	Significant
C*Age	-0.000073	0.000	Significant
W/B*FA	-0.1835	0.000	Significant
W/B*CA	0.0355	0.008	Significant
W/B*RCA	-0.2811	0.000	Significant
W/B*SCM	4.27	0.000	Significant
W/B*SP	-14.12	0.000	Significant
W/B*NF	-7.09	0.000	Significant
FA*CA	0.000023	**0.095**	**insignificant**
FA*SCM	-0.000922	0.000	Significant
FA*NF	0.00571	0.002	Significant

CA*RCA	-0.000062	0.014	Significant
CA*SCM	0.003071	0.000	Significant
CA*SP	-0.00928	0.000	Significant
CA*Age	0.000030	0.000	Significant
RCA*SCM	-0.00377	0.001	Significant
RCA*SP	0.00894	0.015	Significant
RCA*Age	0.000048	0.002	Significant
Statistical metrics errors			
R^2	88.94%		
Adj. R^2	87.89%		
Pred. R^2	86.66%		
10-fold R^2	86.33%		

The quadratic terms of most variables (W/B, FA, CA, RCA, NF, and age) significantly influence the response variable, as indicated by P-values less than 0.05%. This suggests that these variables alone have a substantial impact on the response. Furthermore, interactions between most of these variables are also statistically significant for the response variable, with P-values less than 0.05% [58,59].

To evaluate the model's capability to generalize to unseen data, the coefficient of determination (R^2) was calculated. The R^2 values of 88.94%, with adjusted R^2 of 87.89% and predicted R^2 of 86.66%, indicate strong predictive performance. 10-fold R^2 was 86.33%, which indicates good predictive performance.

Figure 7 visually assesses whether data follows a normal distribution, a key assumption in many statistical analyses. The normality plot in the in Figure 7 suggests that the residuals of f_t model are approximately normally distributed, with some minor deviations.

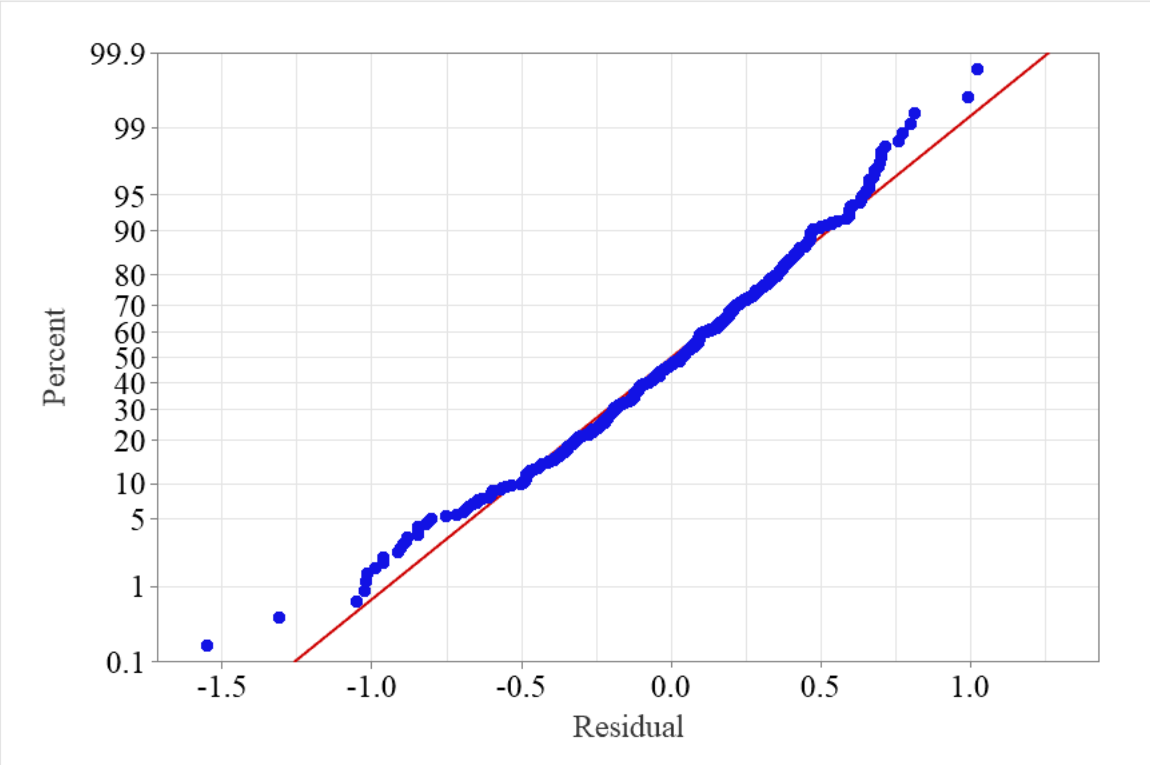

Figure 7. Normal Probability Plot for splitting tensile strength of sustainable concrete.

Figure 8 shows a "Versus Fits" plot, which visualizes the residuals (errors) of a model. The model predicts the f_t of sustainable concrete. In Figure 8, the residuals are scattered randomly around the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1162.v1

zero line. There are no discernible patterns, such as funnels, curves, or clusters. This random distribution shows that the model's predictions are unbiased and consistent across the range of fitted values. A random scatter of residuals suggests that the model is a suitable for the data. This implies that the model's predictions are accurate and reliable, and that the chosen predictors effectively capture the underlying relations in the dataset. Residual analysis reveals a strong correlation between observation points, prediction points, and residual fluctuations, which fall within a range of -1 to 1. The prediction results suggest that the SPR model effectively captured the patterns within the training data. However, to assess its predictive performance on unseen data, further testing with data beyond the training set is crucial.

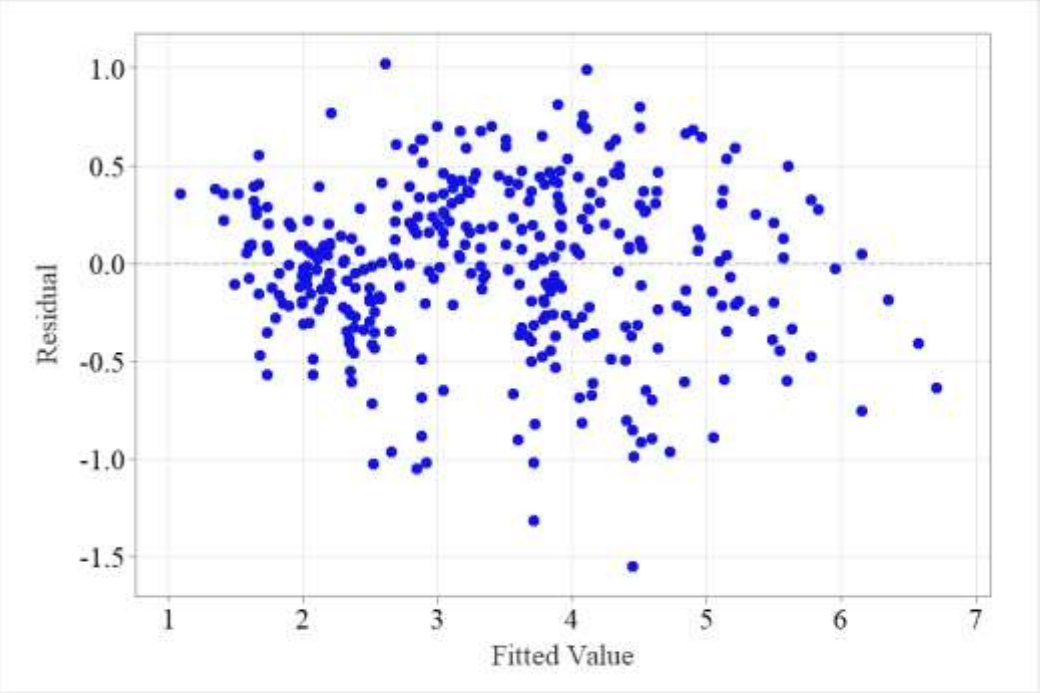

Figure 8. Versus fits plot of for splitting tensile strength of sustainable concrete.

3.2. CART Regression Model

3.2.1. Compressive Strength

The Classification and Regression Tree (CART) model incorporated all 9 predictors (Table 5), each identified as statistically significant through Gini impurity reduction. The final optimized tree structure comprised 39 terminal nodes, with a minimum node size of 3 observations to prevent overfitting by avoiding overly granular splits. Node splitting was governed by the 1-standard-error (1-SE) rule, selecting the simplest tree within one standard error of the maximum R^2 value to balance complexity and generalizability. Cost-complexity pruning (α = 0.01) was applied to trim branches that contributed minimally to predictive accuracy, yielding a parsimonious structure while retaining critical relationships (e.g., water-to-binder ratio dominance). Model performance was rigorously validated via 10-fold cross-validation, achieving a testing RMSE of 5.57 MPa for compressive strength, underscoring robust generalization. This combination of constraints (α = 0.01, node size = 3) ensured the model captured nonlinear interactions without overfitting noise, as evidenced by the tight alignment between training (R^2 = 0.95) and testing (R^2 = 0.91) performance.

Table 5. Compressive strength model summary.

Total predictors	9
Important predictors	9

 doi:10.20944/preprints202603.1162.v1

Number of terminal nodes	39
Minimum terminal node size	3
Node splitting	Least squared error
Optimal tree	Within 1 standard error of maximum R-squared
Model validation	10-fold cross-validation

The CART model demonstrates excellent performance on the training data, as evidenced by the high R^2 value (95.16%), low RMSE (4.1052), MSE (16.8529), MAD (2.985), and MAPE (0.0757) as presented in Table 5. Figure 9's scatter plot visually confirms the model's accuracy on the training data, with points clustering closely around the diagonal line. While the model performs well on the training data, its generalization ability on unseen data remains uncertain and requires further testing.

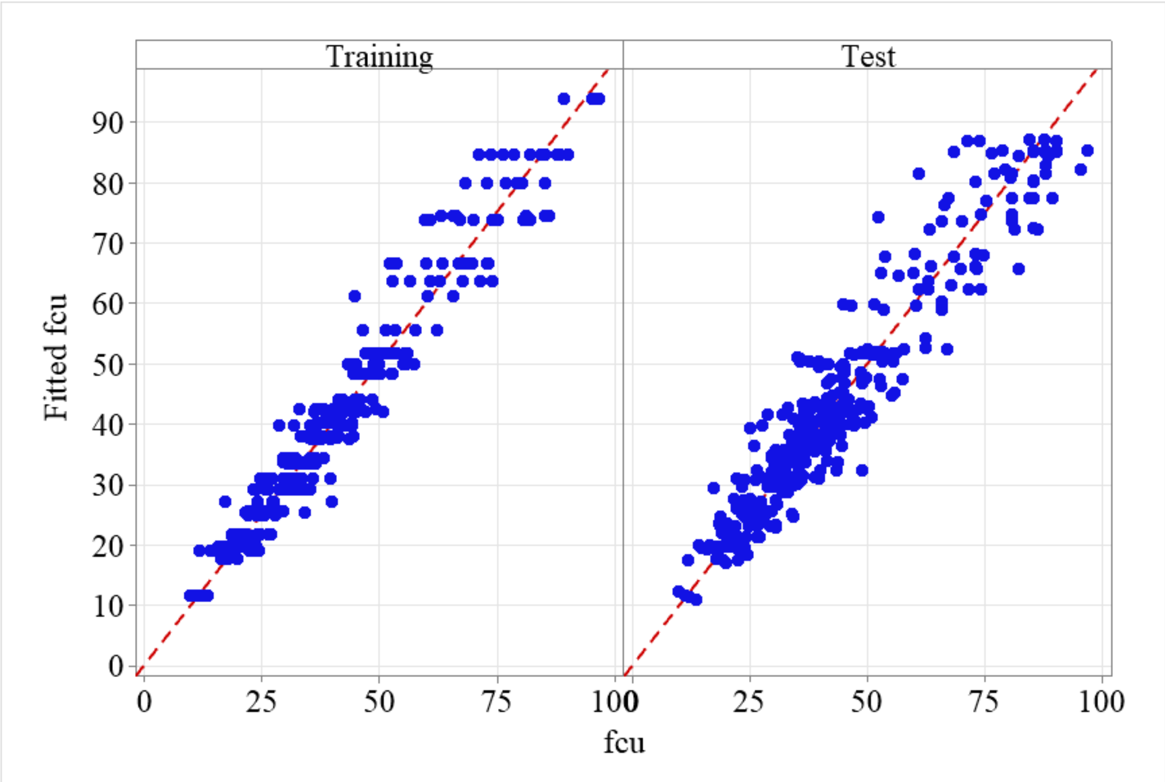

Figure 9. Scatter plots for training and testing sets for compressive strength of sustainable concrete.

The scatter plot in Figure 9 reveals that while the training set exhibits a tighter fit, the testing set still demonstrates a strong overall trend. Although there's a slight tendency to underestimate f_{cu} values at higher levels within the testing set, this doesn't necessarily signify poorer performance. The model's effectiveness (Table 6) is further supported by its high R-squared value (91.09%), low RMSE (5.5686), MSE (31.0098), MAD (4.1076), and MAPE (0.1055). Figure 9 visually emphasizes the importance of assessing the overall spread and deviation from the diagonal line, rather than solely focusing on the direction of bias. Despite some variability in the testing set, the model effectively generalizes to unseen data, indicating its potential for reliable predictions.

Table 6. Statistical metrics error.

Statistics	Training	Test
R-squared	95.16%	91.09%
Root mean squared error (RMSE)	4.1052	5.5686
Mean squared error (MSE)	16.8529	31.0098
Mean absolute deviation (MAD)	2.9850	4.1076

 doi:10.20944/preprints202603.1162.v1

Mean absolute percent error (MAPE)	0.0757	0.1055

Boxplots provide a visual representation of data distribution, showing median, quartiles, and outliers. In Figure 10, both training and testing sets exhibit roughly symmetric distributions of residuals centered around zero, suggesting unbiased errors. The comparable spread of residuals across sets indicates consistent prediction accuracy. This suggests good model performance and generalization to unseen data. Residuals reveal model errors and how well the model fits the dataset. Outliers in residuals might indicate poorly modeled data points, while changes in residual spread with predicted values could signal heteroscedasticity.

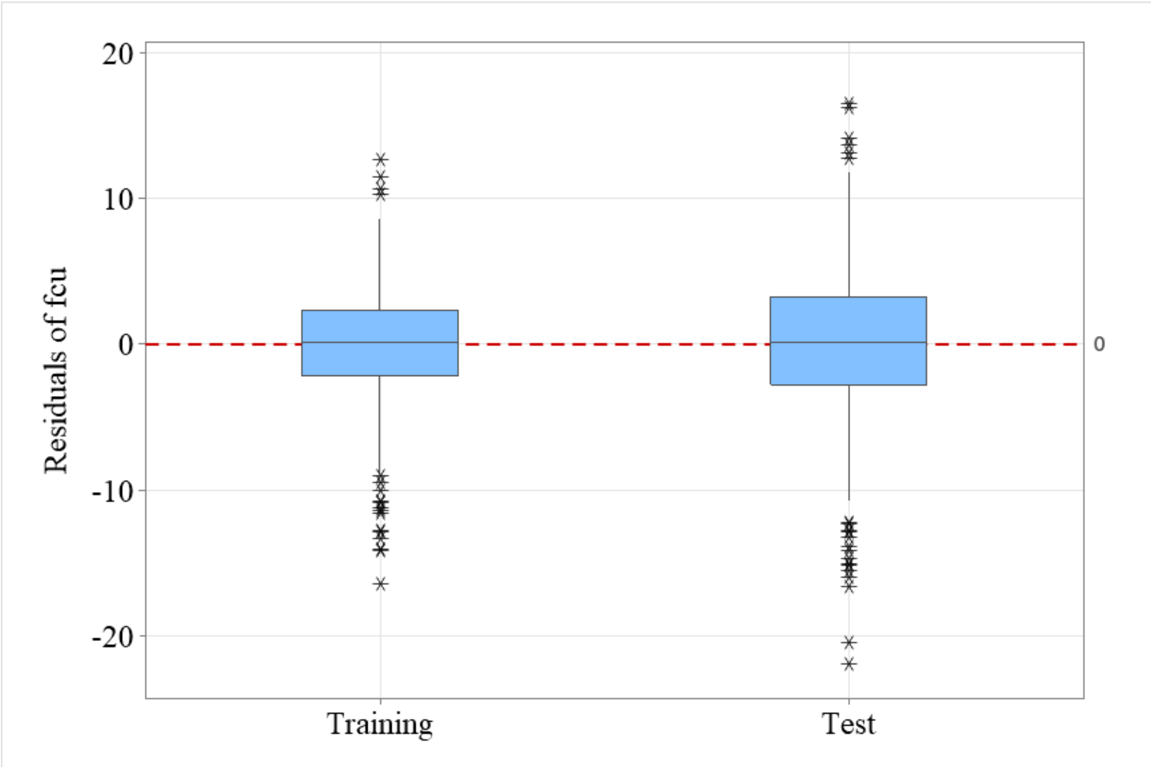

Figure 10. Residuals of compressive strength of sustainable concrete.

3.2.2. Splitting Tensile Strength

The CART model utilizes 9 predictors, all of which were deemed important as presented in Table 7. The decision tree has 28 terminal nodes, with a minimum size of 3 observations per node. Node splitting is determined by the criterion of being within 1 standard error of the maximum R^2 value. The optimal tree is chosen based on minimizing the least squared error. Model validation is performed using 10-fold cross-validation.

Table 7. Splitting tensile strength model summary.

Total predictors	9
Important predictors	9
Number of terminal nodes	28
Minimum terminal node size	3
Node splitting	Least squared error
Optimal tree	Within 1 standard error of maximum R-squared
Model validation	10-fold cross-validation

The CART model demonstrates excellent performance on the training data, as evidenced by the high R^2 value (94.55%), low RMSE (0.2858), MSE (0.0817), MAD (0.2301), and MAPE (0.0741) as

 doi:10.20944/preprints202603.1162.v1

presented in Table 8. Figure 11's scatter plot visually confirms the points are more tightly clustered around the diagonal line, suggesting a good fit on the training data. This is expected, as the model is trained specifically on this data.

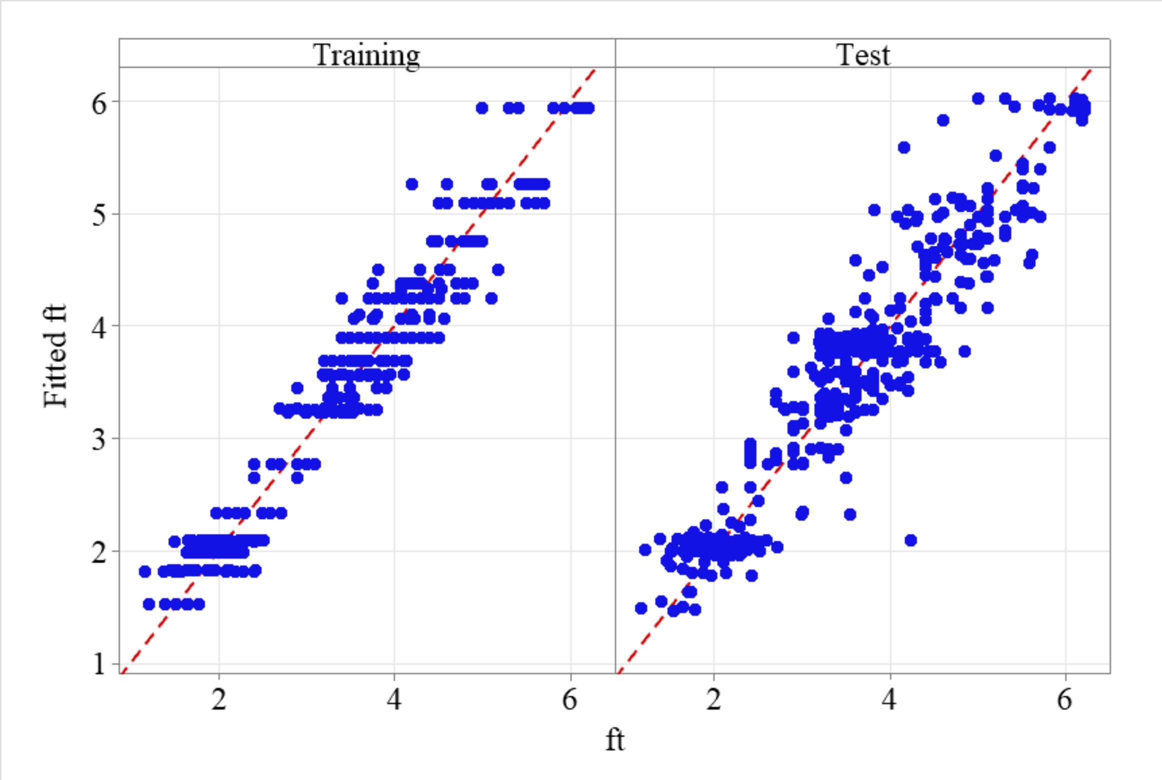

Figure 11. Scatter plots for training and testing sets for splitting tensile strength of sustainable concrete.

Table 8. Statistical metrics error.

Statistics	Training	Test
R^2	94.55%	89.56%
RMSE	0.2858	0.3954
MSE	0.0817	0.1563
MAD	0.2301	0.2996
MAPE	0.0741	0.0939

The scatter plot in Figure 11 reveals that while the training set exhibits a tighter fit, the model still captures the overall trend but with more variability. There might be a tendency to underestimate the f_t for higher values, which is a different behavior compared to the training set. This is due to the distribution of data points in the training and testing sets might be different, which can affect the model's performance. The model might be too complex for the available data, leading to overfitting.

The model's effectiveness is further supported by its high R-squared value (89.56%), low RMSE (0.3954), MSE (0.1563), MAD (0.2996), and MAPE (0.0939). Figure 11 visually emphasizes the importance of assessing the overall spread and deviation from the diagonal line, rather than solely focusing on the direction of bias. Despite some variability in the testing set, the model effectively generalizes to unseen data, indicating its potential for reliable predictions.

Figure 12 shows that both training and testing sets have symmetrically distributed residuals centered around zero, suggesting unbiased errors. The comparable spread of residuals suggests consistent prediction accuracy across both sets. This indicates good model performance and generalization to unseen data. Residuals reveal model errors, while the boxplot visualizes their distribution.

 doi:10.20944/preprints202603.1162.v1

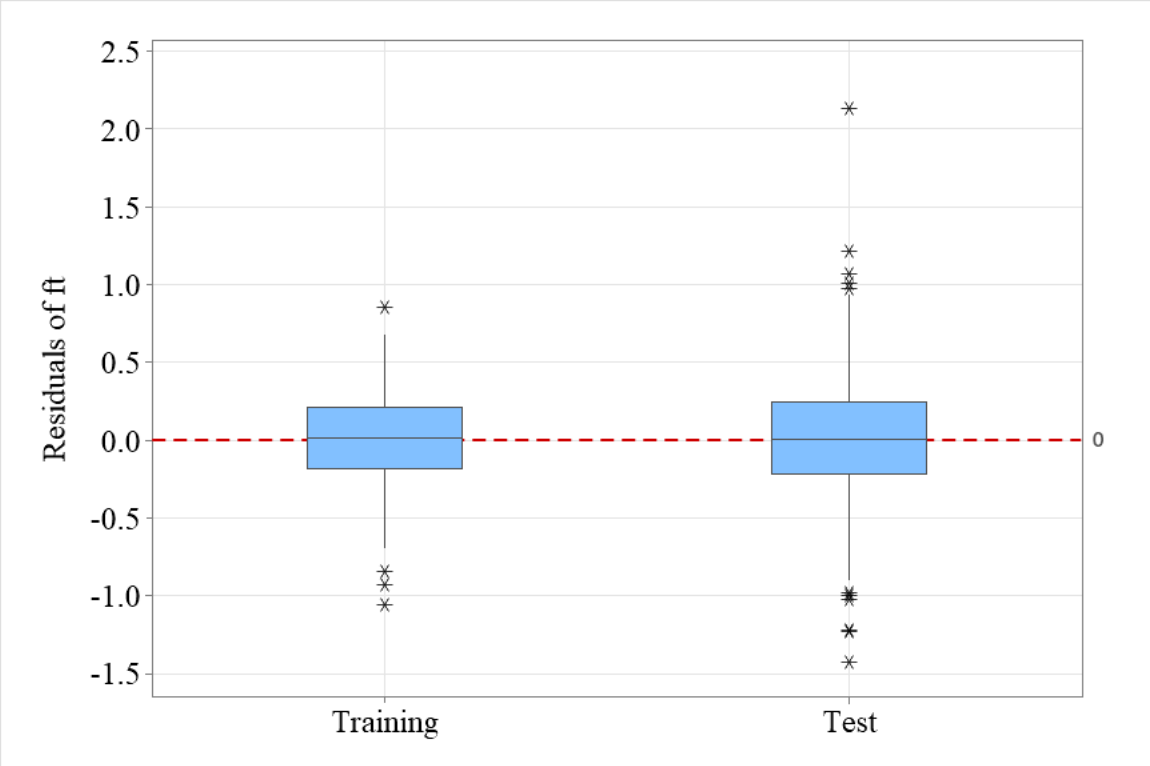

Figure 12. Boxplot of residuals for splitting tensile strength for sustainable concrete.

3.2.3. Variable Importance

Variable importance was derived using Gini impurity reduction for compressive and splitting tensile strength. Figures 13 and 14 illustrate variable influences on compressive and splitting tensile strength.

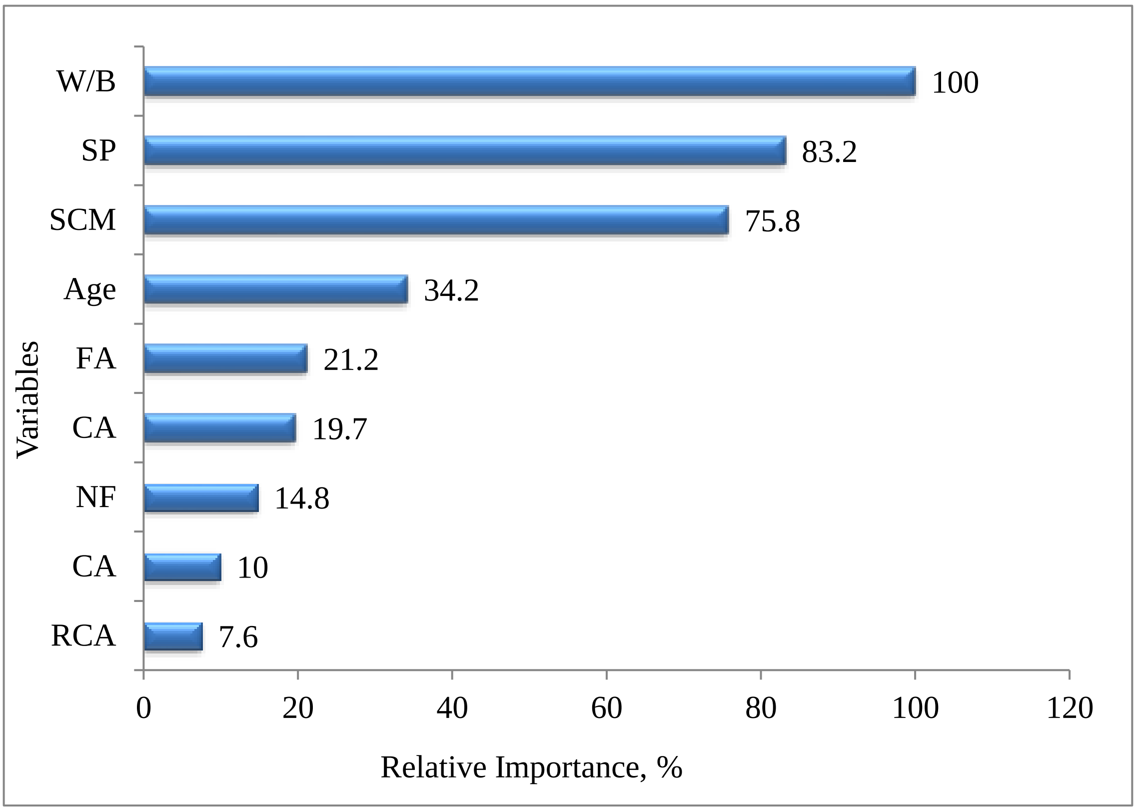

Figure 13. Variable importance of variables on f_{cu} of sustainable concrete.

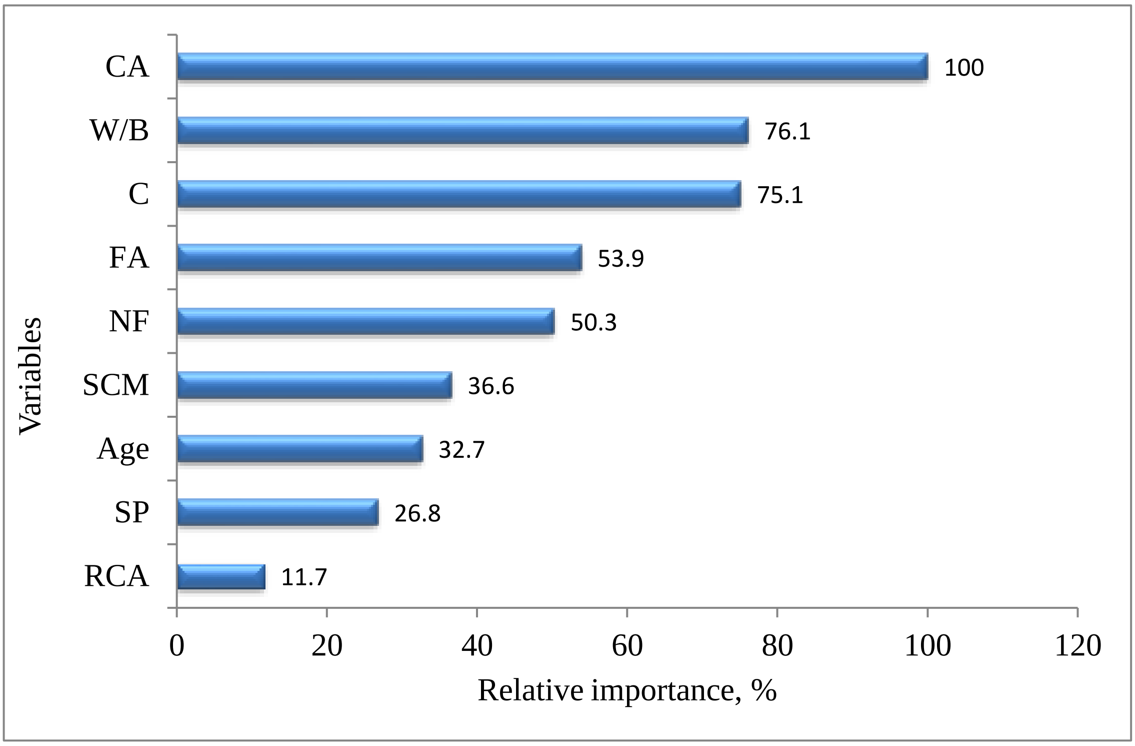

Figure 14. Variable importance of variables on f_t of sustainable concrete.

The water-to-binder ratio (W/B, 100% importance) dominates compressive strength, as lower ratios enhance matrix density by reducing porosity, while higher ratios weaken the binder. Superplasticizer (SP, 83.2%) optimizes particle packing, and supplementary cementitious materials (SCMs, 75.8%) refine pores via pozzolanic reactions. Age (34.2%) reflects curing benefits, while fine aggregate (FA, 21.2%) and coarse aggregate (CA, 19.7%) play secondary roles as fillers. Natural fibers (NF, 14.8%) balance crack bridging with void risks, and cement (10%) shows diminished impact due to synergies with SCMs. Recycled coarse aggregate (RCA, 7.6%) has minimal influence, suggesting controlled quality or low replacement levels.

Coarse aggregate (CA, 100% importance) is paramount for tensile strength, with angularity and interlock enhancing load transfer. W/B (76.1%) and cement (75.1%) govern matrix cohesion and porosity, while FA (53.9%) improves packing, and natural fibers (50.3%) delay crack propagation. SCMs (36.6%) and age (32.7%) have limited roles, focusing on durability and gradual hydration. SP (26.8%) aids fiber distribution indirectly, and RCA (11.7%) shows minor impact due to potential quality controls. Collectively, W/B, SP, and SCMs drive compressive strength (~83% variability), while CA, W/B, and cement dominate tensile behavior (~75%), emphasizing aggregate-matrix synergy and water management in sustainable design.

The divergence in feature importance underscores property-specific optimization strategies:

- For compressive strength: Prioritize strict W/B control (e.g., ≤0.43 via SP optimization) and SCM integration to refine pore structure.
- For tensile strength: Optimize aggregate gradation (angular, well-graded CA) and balance cement content to enhance matrix cohesion, supplemented by natural fibers (≤0.5% by volume) for crack resistance.
- General guidelines: RCA quality (e.g., pre-treatment to reduce adhered mortar) and curing duration are critical for minimizing variability. CART-derived thresholds (e.g., W/B ≤0.43, NF ≤0.5%) offer actionable rules to reduce trial-and-error in mix design.

 doi:10.20944/preprints202603.1162.v1

3.3. Models Validation and K-Fold Cross-Validation

To assess model reliability and generalizability, 10-fold cross-validation was implemented. This robust technique divides the dataset into ten subsets, iteratively training and evaluating the model on different combinations. This approach provides a more reliable performance estimate than a single train-test split.

Employing 10-fold cross-validation, the CART model demonstrated superior performance to the SPR model. For the SPR model, the 10-fold cross-validated R^2 values for SPR-compressive strength and SPR-splitting tensile strength were 86.05% and 86.33%, respectively. In contrast, the CART model achieved higher 10-fold cross-validated R^2 values for CART-compressive strength and CART-splitting tensile strength of 91.09% and 89.56%, respectively.

These 10-fold cross-validation results, combined with other statistical metrics (RMSE, MSE, MAD, MAPE), strongly indicate the high generalization capacity and accuracy of the developed models. This suggests that the models are likely to exhibit reliable performance when applied to new, unseen data, making them valuable tools for predicting the strength of sustainable concrete in real-world engineering applications.

3.4. Comparison Between Applied Machine Learning Models

Model performance was assessed using a combined violin and box plot analysis (Figure 15). This visualization effectively compared data distributions, with the box plot representing the interquartile range (IQR) and the violin plot depicting the probability density. The analysis revealed that the CART model outperformed the SPR model in capturing the full range of potential outcomes. This was evident in the violin plots, where the CART model accurately predicted both the minimum and maximum values of both strength parameters.

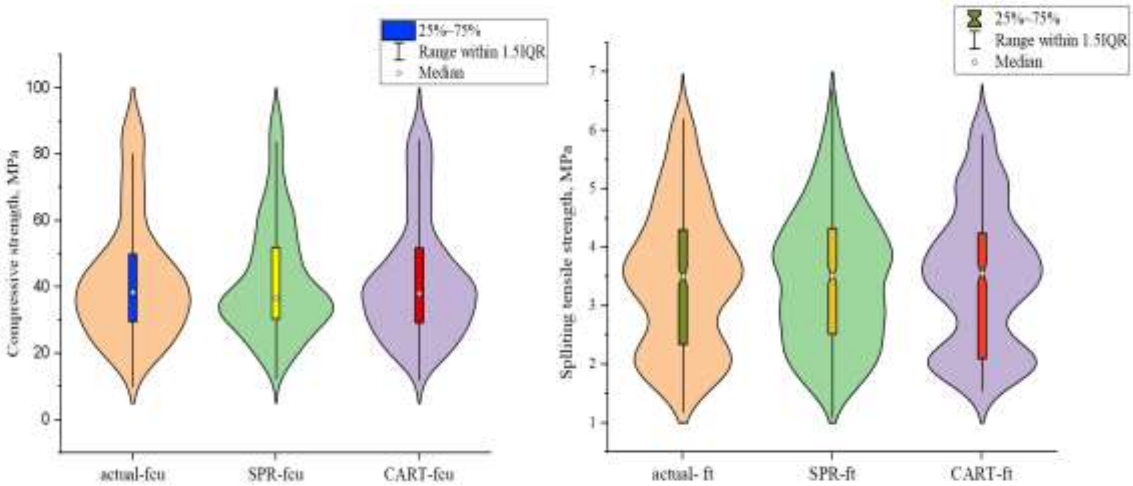

Figure 15. Villon with box plot for a) f_{cu} and b) f_t for sustainable concrete.

Taylor diagrams provide a comprehensive framework for evaluating model performance by comparing predictions to experimental data. They assess three key aspects: correlation coefficient, standard deviation, and root mean square error. Models positioned closer to the experimental point (represented by a red point) exhibit higher correlation, closer agreement in variability, and lower overall error, indicating superior performance [60]. Points along the correlation axis suggest models with good variability but weaker linear relationships. Conversely, points along the standard deviation axis indicate models with good correlation but different levels of variability compared to the experimental data [61,62].

Figure 16 presents Taylor diagrams for (f_{cu} and f_t) prediction. The results clearly demonstrate the superior predictive accuracy of the CART model over the SPR model, as evidenced by higher correlation coefficients and lower RMSE values.

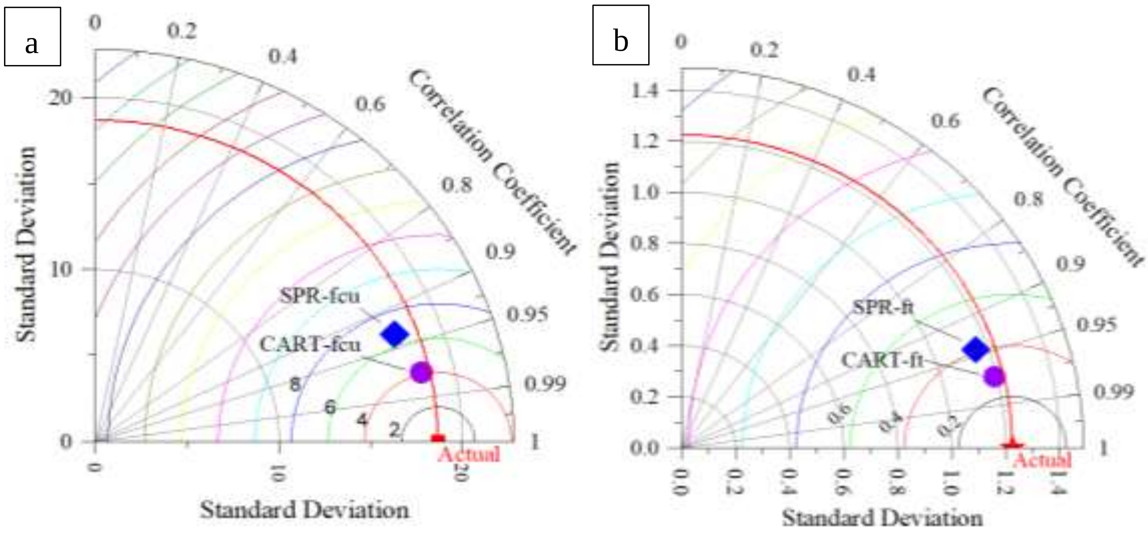

Figure 16. Taylor diagram for a) f_{cu} and b) f_t for sustainable concrete.

CART's superiority stems from its ability to:

1. **Handle Nonlinearity and Multicollinearity:** Unlike SPR, which relies on predefined polynomial terms (e.g., Equations (9) and (10)), CART inherently captures nonlinear interactions (e.g., threshold effects like *W/B ≤ 0.43* for compressive strength) without inflating variable significance due to multicollinearity.
2. **Provide Actionable Interpretability:** CART's hierarchical splits (e.g., *NF ≤ 0.5%* for optimal fiber dosage) offer engineers direct, rule-based guidelines for mix design, bypassing SPR's complex equations.
3. **Generalize Robustly:** Cross-validation confirmed CART's stability (testing R^2 = 0.91 vs. SPR's 0.86), as SPR's backward elimination often discarded critical interactions (e.g., *SP-SCM* synergy).

4. Conclusions

This study establishes a machine learning framework to predict the strength of natural fiber-reinforced recycled aggregate concrete, addressing critical gaps in interpretability and multicollinearity inherent to prior models. By applying Classification and Regression Trees (CART) and Stepwise Polynomial Regression (SPR) as distinct methodologies, the trade-off between accuracy and interpretability is resolved. This approach achieves dual-target predictions (compressive strength: R^2 = 0.91; tensile strength: R^2 = 0.89) while independently elucidating material interactions. CART's hierarchical splits revealed mechanistic insights: the water-to-binder ratio (W/B ≤ 0.43) governs compressive strength by densifying the matrix, countering RCA's porosity; coarse aggregate angularity dominates tensile strength through interlock-driven crack resistance; and superplasticizer (SP)-fiber synergies optimize workability without compromising strength, validated through SPR interaction terms ($p < 0.001$). These findings bridge ML predictions with materials science principles, offering engineers actionable mix design guidelines (e.g., RCA ≤ 40%, NF ≤ 0.5%) that reduce trial-and-error and enable scalable adoption of sustainable concrete in structural systems such as beams, pavements, or foundations.

The framework's generalizability (10-fold R^2 = 0.86–0.91) ensures robustness across variable RCA quality and fiber types, making it adaptable to real-world construction scenarios where material consistency is challenging. By embedding material science into predictive analytics, this work empowers engineers to optimize eco-friendly concrete without sacrificing structural resilience—directly supporting circular economies through reduced reliance on virgin aggregates and landfill waste. Future integration of durability metrics (e.g., chloride ingress) and hybrid ML approaches (e.g., XGBoost) will extend its utility to harsh environments like marine infrastructure, aligning with global goals for sustainable and resilient construction. Thus, the CART-SPR framework not only

 doi:10.20944/preprints202603.1162.v1

advances predictive accuracy but also redefines practical design paradigms, transforming theoretical insights into codifiable rules for next-generation green infrastructure.

Author Contributions: I. K., K. K., wrote the main manuscript text; A. Y. reviewed the manuscript.

Funding: This study did not receive any specific grant from funding agencies in the public commercial sectors.

Ethics and Consent to Participate declarations: not applicable.

Informed Consent Statement: This study does not involve human participants, and no informed consent was required.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Research Involving Human Participants and/or Animals: This study did not involve research on human participants or animals.

Clinical trial number: not applicable.

Consent to Publish declaration: not applicable**.**

Conflicts of Interest: The authors declare no conflicts of interest related to this study.

References

1. M. Sivaraja and M. S. Pillai, "Study on durability of natural fibre concrete composites using mechanical strength and microstructural properties," vol. 33, no. 6, pp. 719–729, 2010.
2. K. M. M. Rao and K. M. Rao, "Extraction and tensile properties of natural fibers : Vakka , date and bamboo," vol. 77, pp. 288–295, 2007, doi: 10.1016/j.compstruct.2005.07.023.
3. K. Ghavami, "Bamboo as reinforcement in structural concrete elements," vol. 27, pp. 637–649, 2005, doi: 10.1016/j.cemconcomp.2004.06.002.
4. A. Ajdukiewicz and A. Kliszczewicz, "Influence of recycled aggregates on mechanical properties of HS / HPC," vol. 24, pp. 269–279, 2002.
5. R. V Silva, J. De Brito, and R. K. Dhir, "The in fl uence of the use of recycled aggregates on the compressive strength of concrete : a review," 2014, doi: 10.1080/19648189.2014.974831.
6. R. Vasco, D. Brito, and R. Kumar, "Establishing a relationship between modulus of elasticity and compressive strength of recycled aggregate concrete," *J. Clean. Prod.*, 2016, doi: 10.1016/j.jclepro.2015.10.064.
7. Silva, R.V. & Brito, Jorge & Dhir, Ravindra. Tensile strength behaviour of recycled aggregate concrete. Construction and Building Materials. 83. 108-118., 2015. 10.1016/j.conbuildmat.2015.03.034.
8. SILVA, R. V. Use of recycled aggregates from construction and demolition waste in the production of structural concrete. Res. Net, 2015.
9. N. Kisku, H. Joshi, M. Ansari, S. K. Panda, S. Nayak, and S. C. Dutta, "A critical review and assessment for usage of recycled aggregate as sustainable construction material," *Constr. Build. Mater.*, 2016, doi: 10.1016/j.conbuildmat.2016.11.029.
10. U. Johnson and M. Rahman, "USE OF RECYCLED CONCRETE AGGREGATE IN CONCRETE : A REVIEW," *J. Civ. Eng. Manag.*, vol. 19, no. 6, pp. 796–810, 2013, doi: 10.3846/13923730.2013.799093.
11. Y. Guo, J. Zhang, G. Chen, and Z. Xie, "Compressive behaviour of concrete structures incorporating recycled concrete aggregates , rubber crumb and reinforced with steel fi bre , subjected to elevated temperatures," *J. Clean. Prod.*, pp. 1–11, 2014, doi: 10.1016/j.jclepro.2014.02.036.
12. A. Manan et al., "Case Studies in Construction Materials Machine learning prediction of recycled concrete powder with experimental validation and life cycle assessment study," *Case Stud. Constr. Mater.*, vol. 21, no. September, p. e04053, 2024, doi: 10.1016/j.cscm.2024.e04053.
13. A. Manan, Z. Pu, A. Majdi, W. Alattyih, S. K. Elagan, and J. Ahmad, "Sustainable optimization of concrete strength properties using artificial neural networks : a focus on mechanical performance Sustainable optimization of concrete strength properties using arti fi cial neural networks : a focus on mechanical performance".

 doi:10.20944/preprints202603.1162.v1

14. A. Manan, Z. Pu, J. Ahmad, and M. Umar, "Multi-targeted strength properties of recycled aggregate concrete through," 2024, doi: 10.1108/EC-07-2024-0635.
15. I. Kattoof, H. Ahmed, M. Abdulhadi, and M. L. Hussien, "Prediction of concrete strength by hierarchical stepwise regression using ultrasonic pulse velocity and Schmidt rebound hammer," *Innov. Infrastruct. Solut.*, pp. 1–11, 2024, doi: 10.1007/s41062-024-01766-1.
16. I. Kattoof, H. Zainab, H. Abbas, M. Kareem, and H. Mohammed, "Comparison of artificial neural network and hierarchical regression in prediction compressive strength of self—compacting concrete with fly ash," *Innov. Infrastruct. Solut.*, 2024, doi: 10.1007/s41062-024-01367-y.
17. I. Kattoof, A. Mousa, and M. L. Hussien, "Harnessing machine learning for accurate estimation of compressive strength of high-performance self-compacting concrete from non-destructive tests : A comparative study," *Constr. Build. Mater.*, vol. 451, no. October, p. 138779, 2024, doi: 10.1016/j.conbuildmat.2024.138779.
18. A. Manan, P. Zhang, S. Ahmad, and J. Ahmad, "Prediction of fl exural strength in FRP bar reinforced concrete beams through a machine learning approach," no. April, 2024, doi: 10.1108/ACMM-12-2023-2935.
19. A. Manan, P. Zhang, S. Ahmad, M. Umar, and A. Raza, "Machine learning prediction model integrating experimental study for compressive strength of carbon-nanotubes composites," *J. Eng. Res.*, no. June, 2024, doi: 10.1016/j.jer.2024.08.007.
20. M. Wani, S.R., Suthar, "Utilizing machine learning approaches within concrete technology offers an intelligent perspective towards sustainability in the construction industry: a comprehensive review.," *Multiscale Multidiscip. Model. Exp. Des.*, vol. 8, no. 1, 2025, doi: https://doi.org/10.1007/s41939-024-00601-5.
21. M. Wani, S.R., Suthar, "Using soft computing to forecast the strength of concrete utilized with sustainable natural fiber reinforced polymer composites.," *Asian J Civ Eng*, vol. 25, pp. 5847–5863, 2024, doi: https://doi.org/10.1007/s42107-024-01150-5.
22. M. A. Wani, S.R., Suthar, "Comparative Analysis of the Predictive Performance of Tree-Based and Artificial Neural Network Approaches for Compressive Strength of Concrete Utilising Waste.," *Int. J. Pavement Res. Technol.*, 2024, doi: https://doi.org/10.1007/s42947-024-00454-8.
23. I. González-taboada, B. González-fonteboa, F. Martínez-abella, and J. L. Pérez-ordóñez, "Prediction of the mechanical properties of structural recycled concrete using multivariable regression and genetic programming," vol. 106, pp. 480–499, 2016, doi: 10.1016/j.conbuildmat.2015.12.136.
24. H. Naderpour, A. Hossein, and P. Fakharian, "Compressive strength prediction of environmentally friendly concrete using arti fi cial neural networks," vol. 16, no. October 2017, pp. 213–219, 2018, doi: 10.1016/j.jobe.2018.01.007.
25. Z. Duan, S. Kou, and C. Poon, "Prediction of compressive strength of recycled aggregate concrete using artificial neural networks," vol. 40, pp. 1200–1206, 2013.
26. A. Trocoli, A. Dantas, M. B. Leite, and K. D. J. Nagahama, "Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks," *Constr. Build. Mater.*, vol. 38, pp. 717–722, 2013, doi: 10.1016/j.conbuildmat.2012.09.026.
27. A. Behnood, J. Olek, and M. A. Glinicki, "Predicting modulus elasticity of recycled aggregate concrete using M5 0 model tree algorithm," *Constr. Build. Mater.*, vol. 94, pp. 137–147, 2015, doi: 10.1016/j.conbuildmat.2015.06.055.
28. A. Gholampour, A. H. Gandomi, and T. Ozbakkaloglu, "New formulations for mechanical properties of recycled aggregate concrete using gene expression programming," *Constr. Build. Mater.*, vol. 130, pp. 122–145, 2017, doi: 10.1016/j.conbuildmat.2016.10.114.
29. Fertilefibre Admin. How We Make Our Peat-Free Coir Composts. 2008. Available online: https://www.fertilefibre.com/blog/peat-free-environment/coir-composts
30. Textile Learner. Abaca Fiber (Manila Hemp) Uses/Application of Abaca Fiber. 2013. Available online: https://textilelearner.blogspot.com/2013/04/abaca-fiber-manila-hemp-usesapplication.html
31. Tanmay Halaye. Ramie Fiber Market Competitive Research and Precise Outlook 2019 to 2025. 2019. Available online: https://themarketresearchnews.com/2019/04/02/ramie-fiber-market-competitive-researchand-precise-outlook-2019-to-2025/

32. Jute Geotextiles by Deyute, Geotextile Jute Fiber 732 gr/m2 122 cm. 2019. Available online: https: //www.deyute.com/product/geotextiles-natural-fibers/91 https:%0A//www.deyute.com/product/geotextiles-natural-fibers/91
33. Sunstrands. The Basics of Kenaf Fiber and Hurd. 2019. Available online: https://www.sunstrands.com/2019/ the-uses-of-kenaf-fiber
34. BabaMu, Sisal Fibers. 2019. Available online: https://pixabay.com/photos/sisal-sisal-fibers-sisal-palm-4319997/
35. J. Thomas, N. N. Thaickavil, and P. M. Wilson, "Strength and durability of concrete containing recycled concrete aggregates," *J. Build. Eng.*, vol. 19, no. March, pp. 349–365, 2018, doi: 10.1016/j.jobe.2018.05.007.
36. Gómez-Soberón, José. Shrinkage of concrete with replacement of aggregate with recycled concrete aggregate. ACI Special Publications., 2014, 209.
37. K. Habeeb and A. Rawi, "EFFECT OF ADDING SISAL FIBER AND IRAQI BAUXITE ON SOME PROPERTIES OF CONCRETE ", 2009.
38. T. F. Lam and J. M. Yatim, "Mechanical properties of kenaf fiber reinforced concrete with different fiber content and fiber length," vol. 1, no. 1, pp. 11–21, 2015.
39. K. Younis, F. F. Jirjees, G. Medhat, and B. Ali, "EXPERIMENTAL STUDY ON PERFORMANCE OF RECYCLED AGGREGATE CONCRETE : EFFECT OF REACTIVE MINERAL," no. March, 2019.
40. E. B. Ogunbode, E. I. Egba, O. A. Olaiju, A. S. Elnafaty, and S. A. Kawuwa, "Microstructure and Mechanical Properties of Green Concrete Composites Containing Coir Fibre," vol. 61, pp. 1879–1884, 2017, doi: 10.3303/CET1761311.
41. M. S. Islam and S. Ju, "Influence of jute fiber on concrete properties," vol. 189, pp. 768–776, 2018.
42. A. J. Babafemi, J. T. Kolawole, and O. B. Olalusi, "MECHANICAL AND DURABILITY PROPERTIES OF COIR FIBRE REINFORCED CONCRETE," vol. 14, no. 3, pp. 1482–1496, 2019.
43. A. A. Okeola, S. O. Abuodha, and J. Mwero, "Experimental Investigation of the Physical and Mechanical Properties of Sisal Fiber-Reinforced," 2018, doi: 10.3390/fib6030053.
44. Q. Jin et al., "Strength Properties of Bamboo Fiber Reinforced Concrete Strength Properties of Bamboo Fiber Reinforced Concrete", doi: 10.1088/1757-899X/981/3/032063.
45. K. S. et al. Bas, Y.J., Kakrasul, J.I., Ismail, "Advanced predictive techniques for estimating compressive strength in recycled aggregate concrete: exploring interaction, quadratic models, ANN, and M5P across strength classes," *Multiscale Multidiscip. Model. Exp. Des.*, vol. 8, no. 122, doi: https://doi.org/10.1007/s41939-024-00705-y.
46. S. Tohidi and Y. Sharifi, "Empirical modeling of distortional buckling strength of half-through bridge girders via stepwise regression method," *Adv. Struct. Eng.*, vol. 18, no. 9, pp. 1383–1398, 2015, doi: 10.1260/1369-4332.18.9.1383.
47. I. Kattoof, H. Wissam, N. Mustafa, and S. S. Mohammed, "Prediction of high-performance concrete strength using machine learning with hierarchical regression," *Multiscale Multidiscip. Model. Exp. Des.*, no. 2009, 2024, doi: 10.1007/s41939-024-00467-7.
48. LOH, Wei-Yin. Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery, 2011, 1.1: 14-23.
49. BREIMAN, Leo, et al. Classification and regression trees. Routledge, 2017
50. G. W. Cha, W. H. Hong, and K. G. Lee, "A Study on Estimation of Demolition Waste Generation Index by using CART (Classification and Regression Tree) Analysis," no. Icsmme, pp. 90–93, 2015.
51. J. Chou, C. Tsai, A. Pham, and Y. Lu, "Machine learning in concrete strength simulations : Multi-nation data analytics," vol. 73, pp. 771–780, 2014.
52. S. Alam and U. Gazder, "Classification and regression tree (CART) modelling for analysis of shear strength of FRP-RC members," *Arab J. Basic Appl. Sci.*, vol. 28, no. 1, pp. 397–405, 2021, doi: 10.1080/25765299.2021.1984033.
53. S. Radiation, S. Duration, and E. Models, "Sustainable Energy Technologies and Assessments , Volume 40 , August 2020 , 100768 doi : https://doi.org/10.1016/j.seta.2020.100768 Sustainable Energy Technologies and Assessments , Volume 40 , August 2020 , 100768 doi : https://doi.org/10.1016/j.seta.2020.100768," vol. 40, no. August, 2020.

54. KOHAVI, Ron, et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai. 1995. p. 1137-1145
55. M. A. Khan et al., "Geopolymer Concrete Compressive Strength via Arti fi cial Neural Network , Adaptive Neuro Fuzzy Interface System , and Gene Expression Programming With K-Fold Cross Validation," vol. 8, no. May, pp. 1–19, 2021, doi: 10.3389/fmats.2021.621163.
56. I. K. Harith, M. S. Hassan, and S. S. Hasan, "Liquid nitrogen effect on the fresh concrete properties in hot weathering concrete," *Innov. Infrastruct. Solut.*, vol. 7, no. 1, 2022, doi: 10.1007/s41062-021-00731-6.
57. I. Kattoof, H. Muna, M. A. A. Mousa, and A. Mohammed, "Harnessing machine learning for accurate estimation of concrete strength using non-destructive tests : a comparative study," *Multiscale Multidiscip. Model. Exp. Des.*, vol. 123, 2025, doi: 10.1007/s41939-024-00605-1.
58. I. K. Harith, M. S. Hassan, S. S. Hasan, and A. Majdi, "Optimization of liquid nitrogen dosage to cool concrete made with hybrid blends of nanosilica and fly ash using response surface method," *Innov. Infrastruct. Solut.*, vol. 8, no. 5, pp. 1–15, 2023, doi: 10.1007/s41062-023-01107-8.
59. W. Nadir, I. K. Harith, and A. Y. Ali, "Optimization of ultra-high-performance concrete properties cured with ponding water," *Int. J. Sustain. Build. Technol. Urban Dev.*, vol. 13, no. 4, pp. 454–471, 2022, doi: 10.22712/susb.20220033.
60. K. E. Taylor, "in a single diagram," vol. 106, pp. 7183–7192, 2001.
61. S. S. Band et al., "Mechanics Groundwater level prediction in arid areas using wavelet analysis and Gaussian process regression," 2021, doi: 10.1080/19942060.2021.1944913.
62. W. Science, N. K. Alomari, P. Sihag, and A. M. S. Al-janabi, "Modeling of scour depth and length of a diversion channel flow system with soft computing techniques," no. March, 2023, doi: 10.2166/ws.2023.026.