

Journal of Natural Fibers

ISSN: 1544-0478 (Print) 1544-046X (Online) Journal homepage: www.tandfonline.com/journals/wjnf20

Machine Learning-Based Prediction of Compressive Strength and Sustainability Performance of Cactus Fiber-Reinforced Concrete

Sathvik Sharath Chandra , B. Narendra Kumar , Pshtiwan Shakor , Srushti V. Hosamath , M. VishnuPriyan & George Uwadiegwu Alaneme

To cite this article: Sathvik Sharath Chandra , B. Narendra Kumar , Pshtiwan Shakor , Srushti V. Hosamath , M. VishnuPriyan & George Uwadiegwu Alaneme (2026) Machine Learning-Based Prediction of Compressive Strength and Sustainability Performance of Cactus Fiber-Reinforced Concrete, Journal of Natural Fibers, 23:1, 2655416, DOI: 10.1080/15440478.2026.2655416

To link to this article: https://doi.org/10.1080/15440478.2026.2655416

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.

View supplementary material

Published online: 14 Apr 2026.

Submit your article to this journal

Article views: 1058

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=wjnf20

JOURNAL OF NATURAL FIBERS
2026, VOL. 23, NO. 1, 2655416
https://doi.org/10.1080/15440478.2026.2655416

Taylor & Francis
by informa

OPEN ACCESS Check for updates

Machine Learning-Based Prediction of Compressive Strength and Sustainability Performance of Cactus Fiber-Reinforced Concrete

Sathvik Sharath Chandra[a], B. Narendra Kumar[b], Pshtiwan Shakor[c], Srushti V. Hosamath[a], M. VishnuPriyan[d], and George Uwadiegwu Alaneme[e]

[a]Department of Civil Engineering, Dayananda Sagar College of Engineering, Bengaluru, India; [b]Department of Civil Engineering, VNR VJIET, Hyderabad, India; [c]Department of Civil Engineering, Al-Qalam University, Kirkuk, Iraq; [d]Department of Civil Engineering, School of Engineering, SR University, Warangal, India; [e]Department of Civil Engineering, Kampala International University, Uganda

ABSTRACT

The growing demand for sustainable construction materials has encouraged the adoption of natural fibers as alternatives to synthetic reinforcements in concrete. Cactus fiber (CF) has emerged as a promising reinforcement due to its renewable nature, favorable mechanical properties, and low environmental impact. However, predicting the nonlinear compressive strength behavior of cactus fiber-reinforced concrete (CFRC) remains challenging using conventional empirical approaches. This study applied four advanced machine learning algorithms – Support Vector Regression – Multilayer Perceptron (SVR-MLP) ensemble, Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Extreme Learning Machine (ELM) – to predict CFRC compressive strength. A comprehensive dataset was compiled from published literature incorporating key variables such as cement content, aggregate proportions, fiber dosage, water – binder ratio, and curing age. Model evaluation results showed that XGBoost achieved superior predictive performance with the highest accuracy ($R^2 = 0.9720$) and lowest prediction errors compared to other models. The developed machine learning framework enables reliable strength prediction and optimization of fiber dosage under controlled mix conditions. Findings confirm cactus fiber as a viable low-carbon reinforcement material and demonstrate the effectiveness of artificial intelligence in supporting sustainable, data-driven concrete mix design and enhanced performance reliability.

RESEARCH HIGHLIGHTS

- Advanced machine learning models (SVR–MLP ensemble, LightGBM, XGBoost, and Extreme Learning Machine) were developed to predict the compressive strength of cactus fiber-reinforced concrete (CFRC), addressing its highly nonlinear material behavior.
- XGBoost demonstrated superior predictive performance with a high coefficient of determination (R^2 = 0.9934) and the lowest error metrics, outperforming LightGBM and the ensemble model.
- A comprehensive dataset compiled from published studies incorporated key mix-design and curing parameters, enabling robust data-driven prediction and optimized mix formulation.
- The results confirm cactus fiber as a viable low-carbon and renewable reinforcement capable of enhancing sustainability without compromising mechanical performance.
- Integrating natural fiber-reinforced concrete with machine learning provides a reliable framework for performance forecasting, improved design confidence, and sustainable concrete development.

摘要

隨著對永續建築材料的需求日益增長，天然纖維作為混凝土中合成增強材料的替代品，其應用日益普及。仙人掌纖維 (CF)因其可再生特性、優良的力學性能及低環境影響，已成為一種極具潛力的增強材料。然而，使用傳統經驗方法預測仙人掌纖維增強混凝土 (CFRC)的非線性抗壓強度行為，仍具挑戰性。本研究採用四種先進的機器學習演算法——支援向量迴歸-多層感知器 (SVR-MLP)集合模型、輕量梯度提升機

KEYWORDS

Cactus fiber concrete; compressive strength; XGBoost; sustainable construction; gradient boosting; natural fibers

關鍵詞

仙人掌纖維混凝土; 抗壓強度; XGBoost; 永續建築; 梯度提升法; 天然纖維。

CONTACT George Uwadiegwu Alaneme alanemeg@kiu.ac.ug; tinz2020@gmail.com Department of Civil Engineering, Kampala International University, Uganda

Supplemental data for this article can be accessed online at https://doi.org/10.1080/15440478.2026.2655416

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

(LightGBM)、極限梯度提升機 (XGBoost)及極限學習機 (ELM)——來預測CFRC的抗壓強度。研究團隊從已發表文獻中彙整了包含水泥含量、骨料配比、纖維用量、水灰比及養護天數等關鍵變數的綜合資料集。模型評估結果顯示，相較於其他模型，XGBoost 展現出卓越的預測性能，不僅準確度最高 ($R^2 = 0.9720$)，預測誤差亦最低。所開發的機器學習框架可在受控配比條件下，實現可靠的強度預測及纖維用量的優化。研究結果證實仙人掌纖維是一種可行的低碳增強材料，並展示了人工智慧在支援永續、數據驅動的混凝土配比設計及提升性能方面的有效性。

Introduction

The utilization of concrete on a global scale is the highest of all materials used in the construction sector because it can be produced easily and effectively, has high durability and flexibility (Fathy et al. 2025). Due to rapid urbanization and increased need for additional infrastructure, the requirement for more environmentally friendly construction materials has grown significantly (Maryclara Aguome et al. 2024). The production of traditional concrete utilizes ordinary Portland cement (OPC), which is made from natural resources, and natural aggregate, which generate significant amounts of greenhouse gases and deplete the world's resources (Akeke et al. 2023; Velumani et al. 2023). In recent years, an increasing number of researchers have investigated using alternative materials such as cactus fiber and recycled glass to produce a lower impact version of traditional concrete that will improve the mechanical properties (Kammoun and Trabelsi 2019; Zhang et al. 2024). Traditional concrete has many types of synthetic fibers, such as polypropylene, steel, and carbon fibers added to increase cracking resistance, strength, and durability (Agor, Mbadike, and Alaneme 2023). Although some mechanical benefits may be realized from adding synthetic fibers to traditional concrete, these fibers are non-biodegradable and require considerable amounts of energy to produce them; thus, creating sustainability issues. The increasing attention has been directed toward biodegradable and renewable natural fibers as environmentally responsible reinforcement alternatives (Mohanraj and Krishnasamy 2024; Swathi et al. 2026).

Among natural fibers, cactus fibers have gained popularity due to their renewable nature, biodegradability, and availability in arid and semi-arid climates (Fattah et al. 2025; Fei et al. 2023). Favorable tensile strength, crack-bridging ability, and post-crack energy absorption characteristics have been exhibited by cactus fibers. Experimental studies indicate that optimized fiber dosages can improve tensile and flexural performance, and enhance ductility in CFRC; however, compressive strength behavior in CFRC is characterized by variability, and is influenced by numerous nonlinear interactions, including mix proportion, fiber content, fiber distribution, moisture sensitivity, and curing conditions. Most early studies investigated the basic mechanical properties of CFRC (García et al. 2024; Shanmugavel et al. 2020), but few studies have evaluated CFRC under broader application conditions, and most studies were conducted under specific laboratory conditions and single-mix designs. Empirical models or traditional regression-based models typically fail to accurately predict compressive strength in fiber-reinforced systems due to their multiparametric and nonlinear behavior (Alaneme George and Mbadike Elvis 2019; R. Wang et al. 2024).

ML techniques have recently been employed to estimate mechanical properties of conventional and waste-modified concretes as functions of mixture composition (Alaneme et al. 2024; Medrano-Sanchez et al. 2025). Support vector regression, artificial neural networks, random forests, and gradient boosting algorithms have all shown high predictive capability (Al-Taai et al. 2023; Arreola Sánchez et al. 2015). Recent gradient-boosting methods have demonstrated high accuracy for conventional systems; however, their utilization in bio-based fiber concretes such as CFRC has been limited. Most existing studies utilize primarily statistical performance metrics (e.g., R^2, RMSE) to evaluate model performance, and do not incorporate sustainability-focused indicators into the predictive framework (Alaneme George and Elvis 2019; Fernandez et al. 2024; Hernández, de Cano-Barrita, and Torres-Acosta 2016).

A substantial knowledge gap exists in developing a unified AI-based framework that (i) integrates literature-derived CFRC experimental data into a standardized and reproducible database, (ii) conducts rigorous evaluations of various advanced ML models employing multi-metric validation, and (iii) correlates predicted compressive strength with quantifiable sustainability performance indicators. To the best of the authors' knowledge, no prior study has developed a CFRC-specific integrated AI framework that

simultaneously harmonizes literature-based datasets, benchmarks multiple advanced ML models, and quantitatively links predictive strength outcomes with environmental performance metrics. Developing a framework to address this gap is critical for providing reliable and environmentally-informed design optimization for bio-based fiber concrete systems (el Safhi et al. 2024; N. Kumar et al. 2025).

The primary objective of this study is to develop and validate an integrated ML framework for predicting the compressive strength of CFRC using a harmonized literature-based dataset of 475 different mixes. The framework will compare four advanced ML models (Hybrid Ensemble (SVR – MLP), LightGBM, XGBoost, and ELM) and determine the most robust predictive approach (Mustapha et al. 2024; Tahera et al. 2025). The study will also link predicted compressive strength to environmental performance indicators (embodied carbon assessment, Cement Reduction Efficiency Index (CREI), and durability) in a unified AI-driven decision-making platform (Mortar and González-Calderón 2021). By converting predictive outputs into environmental performance metrics, the proposed framework provides a systematic AI-sustainability workflow to evaluate fiber-dependent performance and its sustainability implications under harmonized base mix compositions while ensuring mechanical reliability.

State-of-the-art and contribution of the present study

Several ML models have been developed and applied to predict the compressive strength of both conventional and modified concrete systems using experimental and/or literature-derived databases. The ML models investigated include Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forest (RF), and gradient boosting techniques, including XGBoost and LightGBM. The majority of the current body of literature focuses on comparing the predictive accuracy of different ML models (i.e., R^2, RMSE). There is limited application of ML to sustainable design applications.

Within the context of natural fiber reinforced concrete, the application of ML to this field is relatively rare. Most prior research has focused on industrial by-products and/or synthetic fibers, whereas bio-based fibers have received relatively little attention in terms of data-driven modeling. More specifically, CFRC does not currently possess a harmonized literature-based database for reproducible ML analyses. While some recent studies have employed ML to assist in fiber dosage optimization under controlled base mix composition, these studies are mostly limited to conventional concrete systems, and few employ a single framework to assess both the sustainability-related aspects of bio-based reinforcement, in addition to the predictive aspects of compressive strength. Therefore, a methodological void exists in providing a systematic connection between the predicted compressive strength and the quantifiable carbon footprint and material efficiency of natural fiber concrete systems.

The study builds upon the conventional benchmarking of ML models by:

- Developing and harmonizing a reusable literature-based CFRC dataset.
- Developing and validating a multi-model learning architecture with a comprehensive multi-metric robustness assessment.
- Quantitatively relating the predicted compressive strength to the estimated embodied carbon content, Cement Reduction Efficiency (CREI), durability-related sustainability metrics, and AI-enabled resource efficiency.
- Proposing an integrated AI-sustainability decision support framework for low-carbon fiber dosage optimization under controlled base mix composition.

Studies focus solely on comparative accuracy of the ML models; the proposed framework will translate the output of the predictive model into quantitative indicators related to environmental and material efficiency, so as to enable sustainable mix design for bio-based fiber concrete systems based on performance.

Methodology

The methodological approach of Figure 1 outlines the steps of an integrated methodology that includes data preparation, ML modeling, evaluation of performance, prediction analysis, and integration of sustainability. Data preparation starts from data collection (and integrity checking), data cleaning, data harmonization,

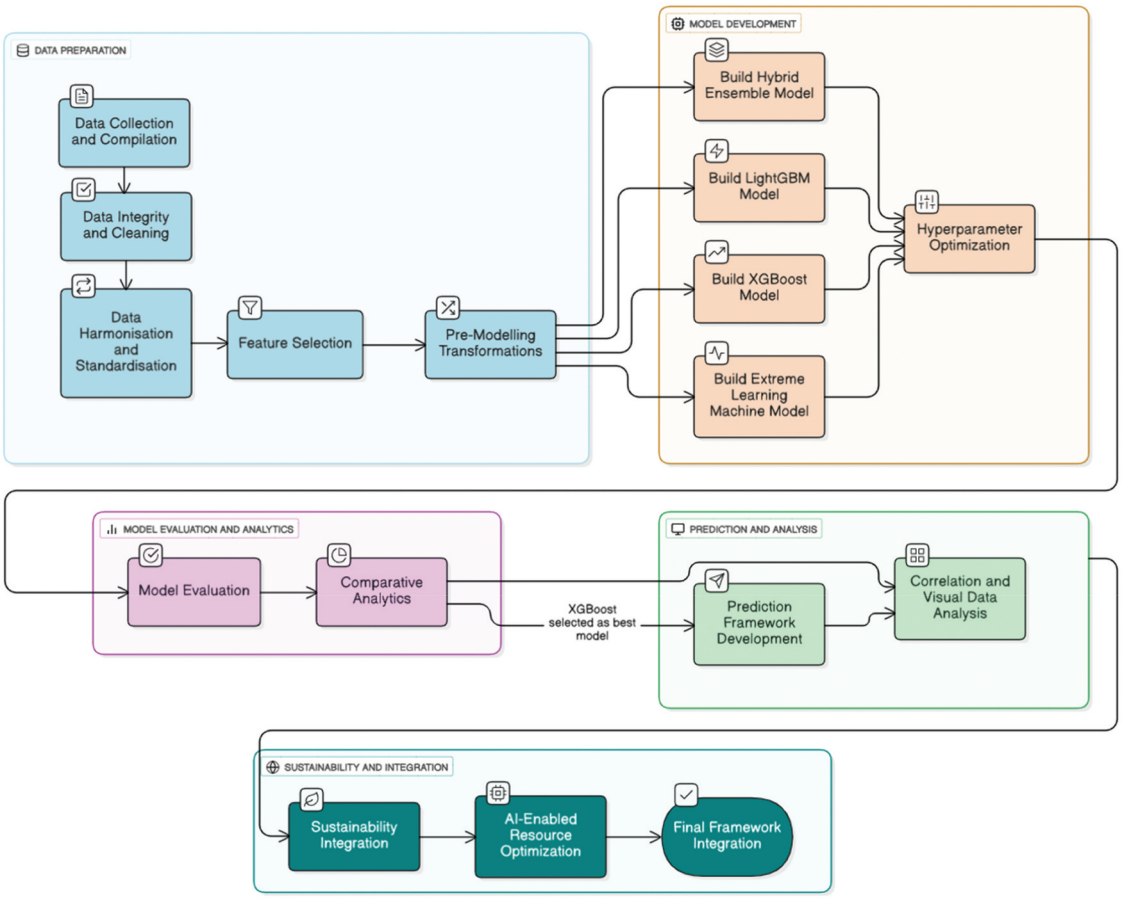


Figure 1. Methodology of the study.

and data normalization. Feature selection and Pre-Modeling Transformation are completed so that high-quality and reliable data are available to input into the ML model. Four ML models (Hybrid Ensemble Model, Light GBM, XGBoost, and ELM) were created, and each model was systematically hyperparameter-tuned. Performance evaluation was conducted through the application of various statistical metrics; Comparative Analytics were performed to identify the models that generated the highest accuracy, and it was identified that XG Boost generated the highest accuracy of the four models tested. The XGBoost model was developed to generate a predictive framework that combined correlation data analysis and visual data analysis. The final step is to complete an integration of sustainability assessment and AI-enabled optimization of resources to develop the most energy-efficient mix designs for CFRC, as well as to provide decision-makers with the data they need to support their sustainability-based decisions.

Materials and dataset development

The study utilizes a unified literature-derived data set of 475 various CFRC mix designs collected from peer-reviewed journals published between 2002 and 2025. Each entry in the data set represents a specific mix design with the experimentally determined compressive strength and the related mix design parameters. Only those studies that utilized standard methods for determining the compressive strength of the mixes were included in this database so that the results could be compared across different studies.

Variables identified as key factors in the governing equations of the database are the amount of cement, the quantity of fine aggregates, the quantity of coarse aggregates, the amount of cactus fiber, the w/b ratio, and the curing age (Murillo et al. 2024). These six input variables were also used as the input variables to the ML models developed in this research project. Compressive strength was established as the only output variable of interest for the purposes of this research project. Other variables (fiber length, slump,

compaction factor, water absorption, split tensile strength, and flexural strength) that were available in some studies and that may relate to the other variables were assessed using an exploratory data analysis (EDA). No additional variables beyond the six input variables were used as input to the ML models to avoid "information leakage" and to provide an accurate measure of the predictive capabilities of the models.

During the process of cleaning the data set, duplicates of the same entry, entries that lacked sufficient detail regarding the mix design being tested, entries that lacked one or more of the primary input variables, or entries that contained results from non-standardized testing procedures were deleted from the data set. While there were limited amounts of durability or mechanical properties represented by secondary parameters in the studies utilized to create this data set, all of the entries in the data set that had valid values for the six governing input variables and the compressive strength were retained for use in developing the ML models. The dataset isolates fiber dosage and curing-age effects under consistent base mix conditions, enabling robust modeling of fiber-dependent nonlinear behavior. The classification of input and output variables used in this study is presented in Table 1.

The descriptive statistics used to evaluate the compiled CFRC data set are shown in Table 2. This data set consists of 475 different mix design combinations that have a fiber addition percentage that ranges from 0% to 4% by weight of binder, and a fiber content that ranges from 0 kg/m^3 to 16.8 kg/m^3. The compressive strengths range from 34.78 MPa to 49.24 MPa and average to 42.76 MPa; thus, there is sufficient variability for use in regression models. The water-to-binder ratio was held constant throughout the data set at 0.42. Fine aggregate content varied slightly within the data set. This variation is attributed to minor adjustments in fine aggregate content to accomodate fibre addition, while maintaining overall mix consisitency. The consistency of split tensile strength and flexural strength distributions supports the reliability of the data set. As a result of the variations in fiber addition percentage and mechanical properties, it can be concluded that nonlinear relationships exist among variables, thus supporting the selection of advanced ML methods for developing predictive models for this type of research.

Data integrity and quality assurance

It was necessary to assess the reliability of the data before modeling, so the data could be trusted. The retrieved values for all of the quantities were verified to ensure they were correct. Data points where the format used to report values was ambiguous, testing conditions were unclear, or values were missing a unit of measure were removed to minimize the potential for the model to learn biased information. To verify that the compressive strength data collected was consistent with that collected using standard testing procedures under similar curing conditions, several consistency tests were performed (Lamichhane, Lamichhane, and Gyawali 2024; X. Wang et al. 2024). The duplicate screening method ensured that multiple entries from the same authors were not included in the analysis. To determine if there were trends between the various data points, which would reflect the known behavior of materials, the data points were compared. These processes ensured the data set was "clean" and reliable and reflected real material behavior as opposed to error.

Data harmonization and feature standardization

The variations of the data were reported across different research studies, and a standard translation process was developed, which allowed each combination of materials to be measured in terms of the same unit (cubic meters of concrete) and represented uniformly. Statistical analysis was also conducted to identify potential anomalies from study results; these possible outliers were reviewed at source (the original articles) to confirm whether they should be removed or left in place (Liu et al. 2023). All variables were normalized

Table 1. Dataset variable classification.

Category	Variables
Primary ML Inputs	Cement, Fine Aggregate, Coarse Aggregate, Fibre Dosage, Water–Binder Ratio, Curing Age
ML Output	Compressive Strength
Secondary (EDA Only)	Slump, Compaction Factor, Water Absorption, Split Tensile Strength, Flexural Strength

Table 2. Descriptive statistics of the compiled CFRC dataset.

Statistic	Fiber_pct_ofbinder	Cement (kg/m^3)	Water (kg/m^3)	Fine Aggregates (kg/m^3)	Coarse Aggregates (kg/m^3)	Fiber (kg/m^3)	WB ratio	Compressive (MPa)	Split Tensile (MPa)	Flexural (MPa)	Water Absorption_pct	Slump (mm)	Compaction Factor
Count	475	475	475	475	475	475	475	475	475	475	475	475	475
Mean	1.500	420.0	176.0	673.70	1200.0	6.300	0.420	42.755	3.622	5.290	3.775	76.261	0.867
Std	1.416	0.0	0.0	5.946	0.0	5.946	~0.000	3.787	0.309	0.414	0.337	13.018	0.046
Min	0.000	420.0	176.0	663.20	1200.0	0.000	0.420	34.780	2.988	4.472	3.277	48.600	0.778
25%	0.500	420.0	176.0	671.60	1200.0	2.100	0.420	40.525	3.452	5.061	3.483	69.750	0.843
50% (Median)	1.000	420.0	176.0	675.80	1200.0	4.200	0.420	42.810	3.613	5.300	3.715	81.200	0.881
75%	2.000	420.0	176.0	677.90	1200.0	8.400	0.420	46.015	3.905	5.664	3.956	86.450	0.902
Max	4.000	420.0	176.0	680.00	1200.0	16.800	0.420	49.240	4.190	5.961	4.429	93.800	0.932

(to remove size) where appropriate to ensure that there would be no scale differences between the variables of cement, aggregate, and fiber content. The above process ensures that the patterns generated by the model will be consistent with the actual physical behavior of materials and not just due to varying scales in the numerical values used (R. Kumar et al. 2025; Subramanya et al. 2025).

Variable encoding and pre-modelling transformations

Standardized data transformations (standardization) were used for all input variables before using ML to improve both model consistency and reliability. The mixtures of materials that had been represented in variable units and reference systems were standardized into a single unitary representation (per cubic meter) for each study (Calò et al. 2025). Standardization of input variables was performed using Equation (1):

$$X_{std} = \frac{X - \mu}{\sigma} \tag{1}$$

where:

- X = original feature value
- μ = mean of the feature
- σ = standard deviation of the feature

A significant scale variation among the numerical features (such as cement content, aggregate fractions, and fiber dosage) was minimized through normalization so that the learning algorithms would not be biased by their magnitudes. All categorical characteristics (fiber type or treatment, etc.) were modeled with accuracy where possible; however, the majority of the information contained within the data set was numeric. Instead of removing outlier patterns, outliers were identified and corrected or removed from the original literature if they were considered to be erroneous. Through these initial data processing steps, each programming model could effectively utilize the data without introducing distortion due to different scales or data structures.

Identification of governing parameters

The input features were selected for their mechanical importance in CS (cement) concrete. The governing factors are cement proportion, proportions of fine and coarse aggregates, CS proportion, water binder ratio, and curing time. All of these factors can influence the hydrating process: fiber matrix interactions and pore structures that affect the compressive strength (R. Kumar et al. 2025). By defining only one output variable (compressive strength), the modeling method is simplified. Rationalization of features will reduce redundancy in the model and make it easier to interpret.

Construction of multi-model learning architecture

The study uses a multi-model learning architecture to use four state-of-the-art ml techniques due to their unique abilities.

Hybrid ensemble: SVR – MLP

Two-stage model combining an error-margin based regressor with a non-linear neural network (Nafees et al. 2022). The SVR can identify the structured relationships found in the data, while the MLP identifies the complex non-linear relationships associated with the CSs embedded within the cementitious matrix. This hybrid approach balances both accuracy and generalization.

The SVR regression function is defined as shown in Equation (2).

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{2}$$

The neuron output of the MLP component is computed using Equation (3).

$$y = \phi\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{3}$$

where:

- w_i = synaptic weights
- x_i = input features
- b = bias
- $\phi(\cdot)$ = activation function (ReLU/tanh)

The hybrid SVR – MLP architecture balances margin-based regression with nonlinear neural learning, enabling robust generalization across heterogeneous CFRC datasets. Margin-based regression (via the SVR) provides the ability to constrain the solution space by minimizing the structural risk of the model to prevent overfitting of the model to the training data. Structural risk is an especially large concern for the type of data collected from heterogeneous studies, such as those related to CFRC. This allows the model to produce a consistent set of results even if there is variability within the data due to factors such as fiber geometry, curing conditions, and mixing inconsistencies. The MLP component is able to complement the function of the SVR component by capturing the nonlinear relationships between the input variables, such as the relationship between fiber dosage and workability or the effect of the water-binder ratio on the hydration rate of cement.

LightGBM gradient booster

LightGBM employs a leaf-wise boosting strategy that effectively captures complex interactions between fiber properties and mixture proportions. The LightGBM algorithm has the advantages of faster training speed and higher accuracy on heterogeneous data, so it is especially suitable for natural fibers reinforced concrete. The LightGBM prediction function is expressed in Equation (4).

$$y^{i} = \sum_{m=1}^{M} f_m(x_i) \tag{4}$$

where:

- f_m = individual decision trees
- M = number of boosting iterations

The leaf-wise approach of LightGBM has allowed it to be more efficient at learning complex feature interaction than traditional level-wise boosting methods. For CFRC, many different factors interact with each other (fiber content, proportion of aggregates, curing age) and affect how strong the compressive strength of the material is. LightGBM identifies the splits of the most informative leaves first, so it does not spend too much time finding the most important variable affecting strength. In addition, LightGBM uses histograms to decide which side of a split is best and also uses a method called gradient-based one-sided sampling to greatly speed up the training process. The advantages of these features make LightGBM an excellent choice for training models on medium-sized data sets that have non-linear relationships and need a good trade-off between model accuracy and the number of computational resources needed.

XGBoost regularized booster

A control boosting model with built-in regularization to avoid overfitting. The XGBoost algorithm can handle non-linear relationships in the mixture design and enhance the reliability of prediction through combining multiple weak learners. The XGBoost objective function is defined in Equation (5).

$$\mathcal{L} = \sum_{i=1}^{N} l(y_i, y_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

The regularization component of XGBoost is presented in Equation (6):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{6}$$

where:

- $l(\cdot)$ = loss function
- T = number of leaves
- w_j = leaf weights
- γ, λ = regularization parameters

The addition of specific regularization terms in XGBoost's loss function enhances the ability of XGBoost to learn a generalizable model. Generalizability is achieved through the control of the complexity of the model through regularization, which is particularly important for modeling CFC due to the data being noisy with fiber dispersion variability and inconsistencies in testing environments that exist between the various studies reported in the literature. In particular, XGBoost helps prevent the model from overfitting and fitting spurious patterns of data (due to excessive tree depth or large leaf weights) while at the same time helping capture the meaningful nonlinear relationships existing within the data. Further, the sequential nature of XGBoost's iterative learning process allows XGBoost to effectively model the non-monotonic relationship between fiber content and compressive strength. The robustness of XGBoost to multicollinearity allows XGBoost to effectively learn from data that includes interdependent variables (such as fine aggregate and fiber content).

Extreme learning machine

A rapid single-layer neural model with randomly generated hidden nodes. ELM provides an efficient baseline for comparison and aids in determining the influence of rapid learning approaches on compressive strength prediction.

The hidden layer output matrix of ELM is calculated using Equation (7).

$$H\beta = T \tag{7}$$

The output weights are determined analytically as shown in Equation (8).

$$\beta = H^{\dagger} T \tag{8}$$

where:

- H = hidden layer output matrix
- β = output weights
- T = target output
- $H^{\dagger}$ = Moore – Penrose pseudoinverse

The ELM has an advantage due to its fast learning, which occurs via an analytical computation of the output layer based on randomly assigned input layer weight values. This may limit the model's ability to capture highly complex nonlinear relationships within the dataset. Therefore, the random initialization of parameters within the hidden layers can produce unreliable predictions, especially when the data set exhibits high variability. This results in the ELM exhibiting high levels of variance and being highly sensitive to small changes in the input data, as shown in the residual analysis and sensitivity analysis.

Hyperparameter optimization strategy

The model performance of each ML algorithm's hyperparameters has to be carefully tuned separately. To determine the most suitable model configuration, an optimization strategy was systematically implemented to evaluate model performance (accuracy) vs. model complexity. The Hybrid SVR-MLP model, as well as the kernel type, regularization factor, and the number of layers in the hidden layer, were iteratively modified to find the optimal balance between these competing factors. The LightGBM and XGBoost models were also

optimized through the selection of different parameters, including the learning rate, maximum tree depth, number of estimators, and sub-sampling ratios. A goal of this approach was to optimize for non-linear interaction effects within the data while avoiding over-fitting. The parameters used in the ELM model were optimized using a systematic hyperparameter search strategy until consistent convergence was achieved through calibration of the number of hidden units as well as the activation functions. Hyperparameter optimization of these three models was performed using validation metrics from the training data set. A systematic optimization strategy was used to improve the reliability of predictions generated by each model and to ensure that each model was working at its optimal level.

Multi-metric performance assessment

A set of statistical measures was utilized to assess the model performance for clear evaluation. The metrics that were used included R^2, RMSE, MAE, WMAPE, NSE, VAF, RSR, and LMI. These various metrics are used to measure different aspects of the model's behavior (such as accuracy, dispersion, error scaling, etc.) and ability to make predictions. The multi-metric technique enables detailed evaluation of all four models while avoiding reliance on a single performance indication. To evaluate model accuracy and robustness, statistical performance metrics were calculated as defined in Equations (9)–(16).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum (y_i - \hat{y}_i)^2} \tag{10}$$

$$MAE = \frac{1}{n}\sum |y_i - \hat{y}_i| \tag{11}$$

$$WMAPE = \frac{\sum |y_i - \hat{y}_i|}{\sum y_i} \times 100 \tag{12}$$

$$NSE = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{13}$$

$$VAF = \left(1 - \frac{\mathrm{var}(y - \hat{y})}{\mathrm{var}(y)}\right) \times 100 \tag{14}$$

$$RSR = \frac{RMSE}{SD_{\text{observed}}} \tag{15}$$

$$LMI = 1 - \frac{\sum |y_i - \hat{y}_i|}{\sum y_i} \tag{16}$$

where:

y_i = Literature value,
$\hat{y}_i$ = predicted value,
$\bar{y}$ = mean literature value.

Cross-model comparative analytics

Each model results were compared to establish a performance hierarchy. The performance hierarchy was used as a basis for comparing each method in terms of how well they handled the non-linear and fiber dependent behavior of cactus fiber concrete. The comparison of single value measures that produced the

same trend for multiple measures highlighted the advantage of using boosting methods, the stability of the hybrid ensemble method, and the limitation of rapid learning methods when dealing with highly non-linear data sets.

Development of the prediction framework

An application-based predictive tool was developed utilizing the most suitable ML model that was identified from comparative analysis. The objective of this tool is to develop an accurate predictor tool for engineers who are able to utilize readily accessible mix design information to accurately predict the compressive strength of CFRC. The predictive tool was developed using the previously trained model by reexamining the trained model over the entire dataset to establish consistency and to minimize the difference between predicted values from training and test datasets. The predictive model was run with the optimal hyperparameters determined as providing the greatest generalization performance to ensure that the predictive model will perform consistently in a field environment.

The predictive tool was also developed to accept variable inputs that represent the characteristics of the mixture design (cement content, proportion of aggregates, water-to-cement ratio, CS dosage, and curing time). By identifying and analyzing trends in residuals and distributions of errors, confidence in the predictions made by the model has been improved so that the completed model can be utilized as a multi-step methodology to potentially use in determining the mix design, selecting materials, or in pre-feasibility studies of CFRC. The simplicity of this format will enable researchers and engineers to use their own calculations to estimate the compressive strength of CFRC without undertaking an extensive amount of laboratory testing. Therefore, it promotes the utilization of CFRC for sustainable building applications.

Correlation heatmap

The heatmap in Figure 2 illustrates the relative interdependencies (both positive and negative) among the most important data fields of the CFRC dataset. Positive associations are represented by dark red and negative associations by dark blue. Compressive strength shows a strong positive relationship with split tensile and flexural strength, suggesting that mixes with higher compressive strength also tend to perform better in other mechanical tests. Water absorption is strongly and negatively correlated with all strength parameters, indicating that mixes absorbing more water tend to have lower strength due to weaker fiber – matrix bonding and higher porosity. Fiber percentage shows a strong negative relation with slump and compaction factor, meaning increasing fiber content reduces workability. Fine aggregate and fiber content appear nearly perfectly negatively correlated because they are inversely adjusted in the mix design. The heatmap highlights meaningful trends that align with expected material behavior, confirming the internal consistency of the data.

Visual exploratory data analysis

A combined correlation and scatter-matrix visualization is depicted in Figure 3 for all variables of the CFRC dataset. In the color-coded correlation matrix in the upper triangular region, red represents positive correlations of high strength and blue represents negative correlations of high strength. Therefore, it is possible to see easily how individual parameters of the composite material (such as compression, splitting, and flexural strength), and which mix parameters influence the properties of the composite material. Patterns can be clearly identified from the structures in the heatmap.

The scatter-plots in the lower triangular region represent the distribution and the association of each pair of variables. These charts show nonlinear trends, clustering, and the distribution of literature values. Fiber percentage, slump, and compaction factor all exhibit different patterns, with increasing fiber dose reducing workability and strength qualities clustering within the expected range. The diagonal elements contain histograms that depict the distribution of each variable. The scatter plots, heatmap, and histograms offer a detailed visual exploration of data structure, correlations, and variable behavior, confirming both dataset consistency and expected material trends.

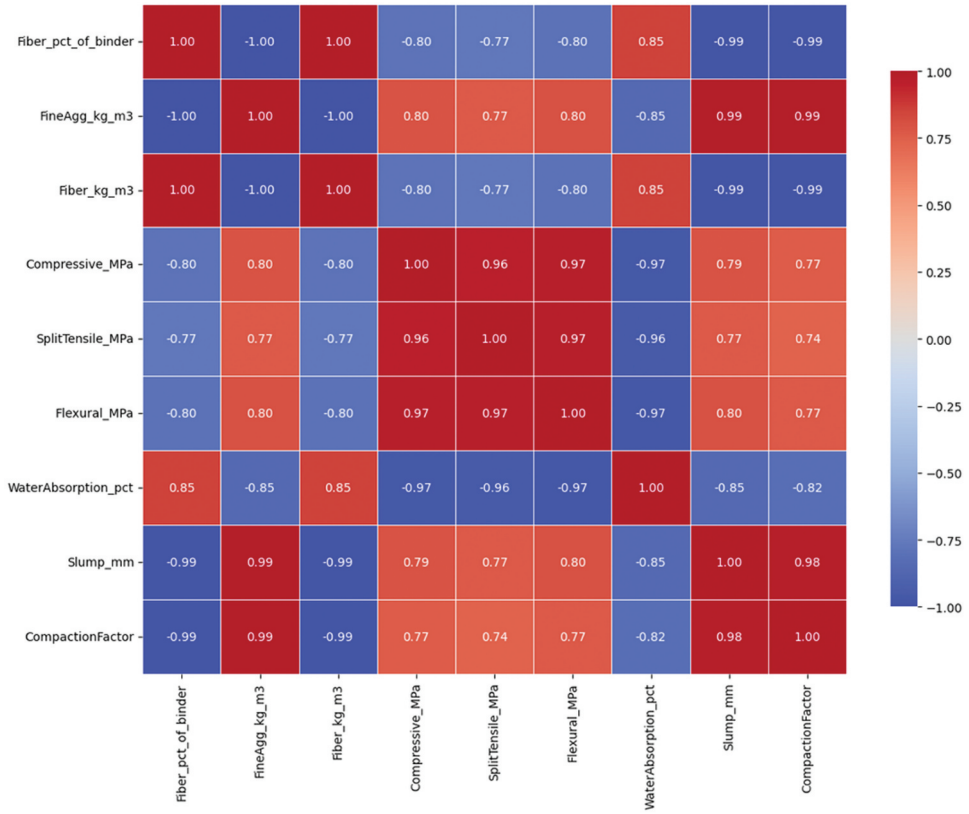


Figure 2. Correlation heatmap showing relationships among mix parameters and mechanical properties of CFRC.

Results

Model performance comparison

Model performance metrics

Table 3 demonstrates that all four models are capable of making good predictions during their training phase, with the Gradient Boosting Models performing the best. Among these GBMs, XGBoost has achieved the greatest accuracy with a high R^2 value of 0.9748; LightGBM is second with an R^2 value of 0.9713, while the latter also has the lowest values for RMSE and MAE among all models. Although the ensemble approach (SVR and MLP) displays sufficient learning behavior, it is associated with a much greater degree of error than all boosting methods. While ELM has done fairly well, it is still significantly less accurate and efficient at reducing errors than the two best-performing GBMs, i.e., XGBoost and LightGBM. The validation of the superior learning performance of GBMs is provided through NSE, VAF, and RSR.

The testing results further highlight the generalization strength of the models. XGBoost and LightGBM retain high predictive accuracy on unseen data, each achieving an R^2 of approximately 0.97 with relatively low RMSE, MAE, and WMAPE values, indicating robust and stable performance. The ensemble model exhibits reliable generalization with slightly reduced accuracy compared to the boosting methods, while ELM shows a modest decline in predictive quality, reflected by higher error indices and lower robustness metrics. The comparison demonstrates that gradient boosting algorithms provide the most accurate and generalizable predictions for compressive strength, followed by the hybrid ensemble model, whereas ELM is comparatively less effective for this dataset. All performance metrics reported in this study are based on a consistent train – test split, and evaluation values correspond exclusively to the hold-out testing dataset.

Comparative performance evaluation of machine learning models

Table 4 compares the performance of the four ML models used to predict the compressive strength of CFRC based on their strengths and limitations. XGBoost is identified as the best-performing model, achieving the highest prediction accuracy and the lowest error values, although it shows slight overfitting due to very high

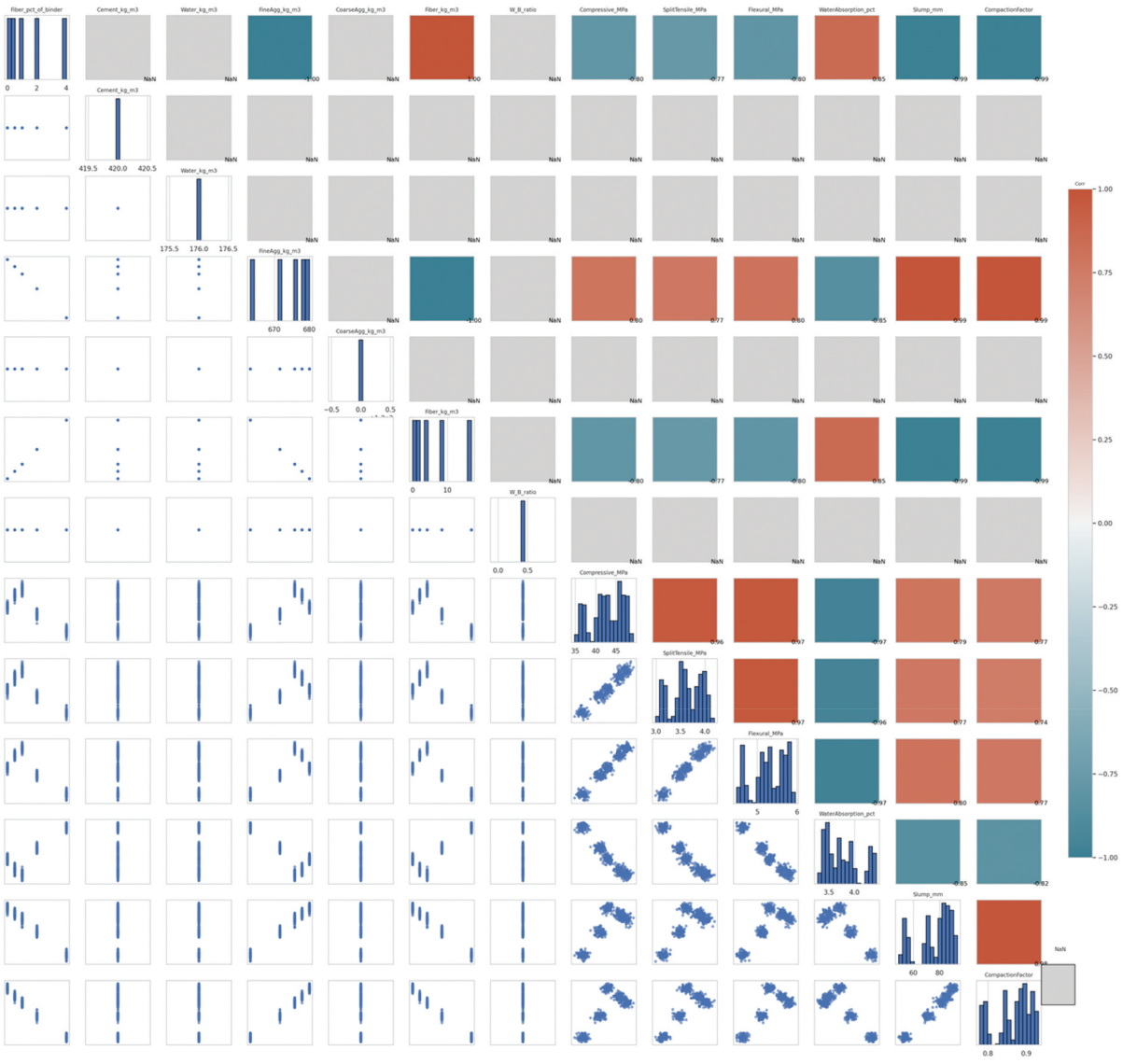


Figure 3. Pairwise scatter matrix for CFRC.

Table 3. Performance metrics of ML models for compressive strength prediction.

Metric	Ensemble (SVR+MLP)	LightGBM	XGBoost	Extreme Learning Machine
R^2 (Training)	0.9143	0.9713	0.9748	0.9542
R^2 (Testing)	0.9284	0.9720	0.9720	0.9669
RMSE (Training)	1.0843	0.6273	0.5884	0.7929
RMSE (Testing)	1.0487	0.6564	0.6561	0.7133
MAE (Training)	0.8569	0.4971	0.4661	0.6254
MAE (Testing)	0.8543	0.5253	0.5102	0.5584
WMAPE (%) (Training)	1.9922	1.1556	1.0838	1.4541
WMAPE (%) (Testing)	2.0472	1.2587	1.2225	1.3381
NSE (Training)	0.9143	0.9713	0.9748	0.9542
NSE (Testing)	0.9284	0.9720	0.9720	0.9669
VAF (%) (Training)	91.4449	97.1318	97.4759	95.4254
VAF (%) (Testing)	92.8713	97.3266	97.2749	96.7296
RSR (Training)	0.2928	0.1694	0.1589	0.2141
RSR (Testing)	0.2676	0.1675	0.1674	0.1820
LMI (Training)	0.0099	0.0058	0.0055	0.0073
LMI (Testing)	0.0104	0.0064	0.0062	0.0067

Table 4. Strengths and limitations of ML models.

Model	Strengths	Weaknesses
XGBoost	Best overall: Highest R^2 (0.9720), lowest RMSE/MAE/RSR	Slight overfitting (train $R^2 = 0.9720$)
LightGBM	Very close to XGBoost, faster training	Slightly higher error than XGBoost
Ensemble (SVR+MLP)	Robust, good generalization	Underfits slightly vs tree models
ELM	Fast training, simple	Poorest performance, high variance

training accuracy. LightGBM performs nearly as well as XGBoost with faster training speed but with slightly higher error. The Ensemble (SVR+MLP) model produces stable and consistent predictions with strong generalization; however, it performs somewhat worse than tree-based models. On the other hand, although it is computationally fast, the ELM model has the poorest predictive capability because it can have high variability and does not generalize very well.

Model comparison Radar chart

The Radar chart of Figure 4 illustrates how each of the four ML models: ELM, XGBoost, LightGBM, and the Ensemble Model (SVR + MLP) performed on performance metrics and achieved ±5 MPa accuracy. Each metric is represented on an individual axis, with the larger shaded area near the edge representing better-than-average prediction performance. XGBoost consistently ranked highest across most performance metrics; therefore, it had the best R^2 value and the lowest error. LightGBM was the second-best performer in all metrics; its performance was consistently strong and very close to that of XGBoost on both accuracy and error. Ensemble generated a smaller polygon than LightGBM or XGBoost, but maintained similar accuracy and good performance in many other metrics. ELM generated a much smaller and irregular polygon than the other three methods, suggesting large errors, low accuracy, and generally poor predictive ability. The clear visual separation of the polygons also illustrates the relative ranking of the models' performance; XGBoost as the top performing model, followed by LightGBM, and then the Ensemble model, and finally ELM, as the worst performing.

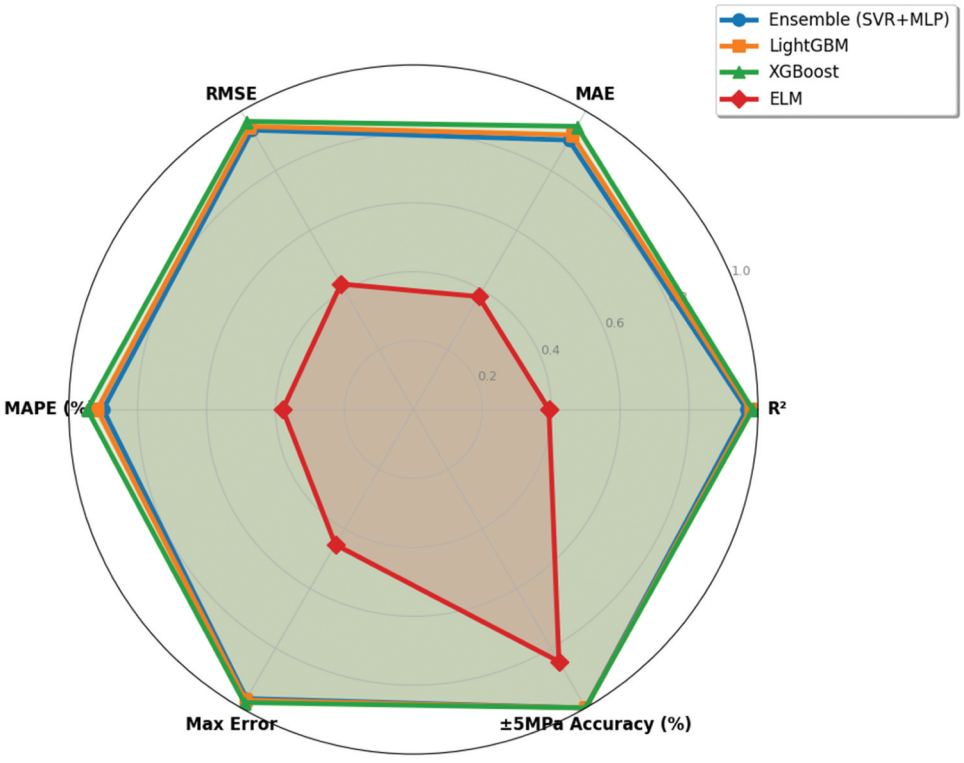


Figure 4. Radar chart comparing predictive performance metrics.

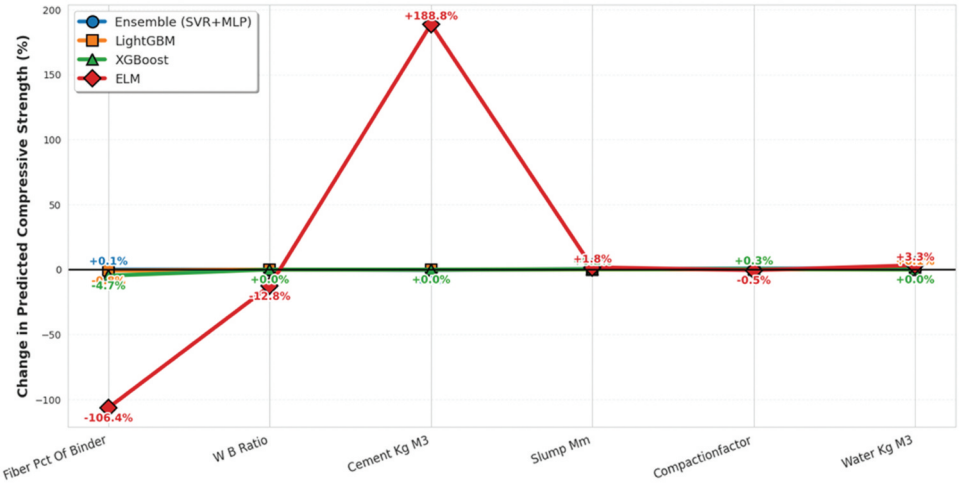


Figure 5. Hoffman and Gardner sensitivity analysis showing feature influence on predicted compressive strength across ML models.

Hoffman and Gardner sensitivity analysis

The Hoffman-Gardner sensitivity analysis of Figure 5, which compares the predicted compressive strengths of the Ensemble (SVR + MLP), LightGBM, XGBoost, and ELM models, demonstrates how the models respond to changes in their input parameters by showing the variability of the predictions when one input parameter at a time is systematically varied, thus enabling comparisons of the relative impacts of different input parameters among the models. Positive percentage values represent an increase in the expected compressive strength with an increase in the input parameter, while negative percentages represent an increase in the expected compressive strength with a decrease in the input parameter. Comparing the responses to input parameter changes of the Ensemble, LightGBM, and XGBoost models, they show similar levels of sensitivity to changes in their input parameters and demonstrate stable behavior with minimal changes in their predicted values resulting from small changes in their input parameters. The response curves for these three models are generally flat and near zero for all input parameters, indicating a high level of robustness and a lack of dependence on any individual parameter. The sensitivity index is calculated according to Equation (17).

$$S_i = \frac{\Delta \grave{y}}{\Delta X_i} \times 100 \tag{17}$$

where:

- S_i = sensitivity of input variable X_i

The ELM model, however, shows very large sensitivity values to most of its input parameters. For example, the ELM model shows dramatic increases and decreases in predicted compressive strength with small changes in fiber percent, cement content, and slump. These large swings highlight the instability of the ELM model and indicate that the ELM model does not generalize reliable input-output relationships. The comparison between the smooth, nearly flat response lines of the top-performing models and the steep deviation of the ELM response lines illustrates the significant differences in model stability and predictive reliability.

Effect of cactus fiber content on mechanical strength

The mechanical performance of concrete is heavily affected by the concentration of CSs (CS content), and strength is typically enhanced at a low to moderate dosage of CSs as they create additional benefits of crack bridging and stress redistribution and increase toughness (Refer to Figure 6). The CSs bond together within the cement matrix to delay crack initiation and crack extension, and will increase compressive, tensile, and flexural strength, most notably in bending and tensile conditions. Beyond the optimal CS dose, the

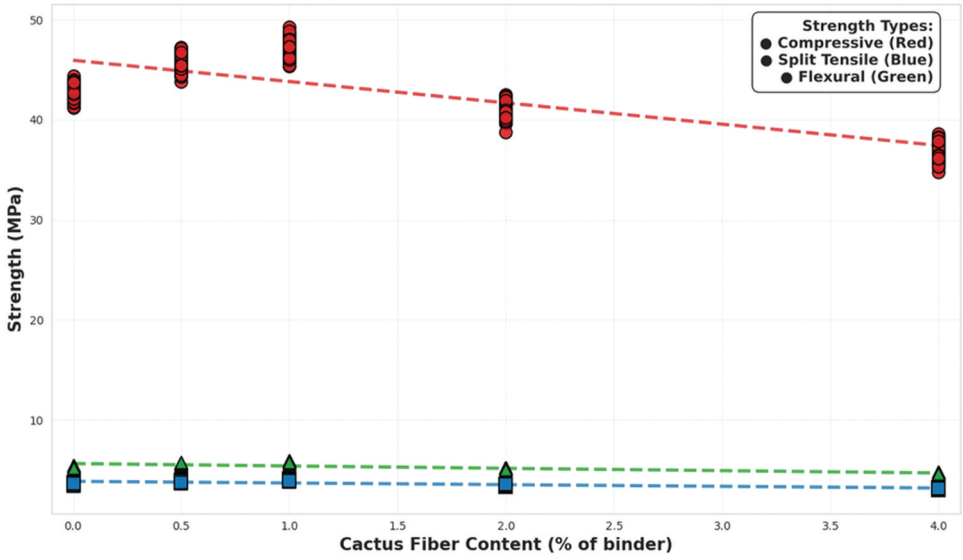


Figure 6. Effect of cactus Fibre content on mechanical strengths.

mechanical properties of the concrete decrease due to the tendency for CSs to cluster, decreased workability, and increased porosity, which creates weak links between the cactus fibers and the cement matrix. In addition to creating poor bonds between the CS and the cement matrix, excess CSs impede proper compaction, which results in voids being created and discontinuities in the matrix. The results indicate that CSs enhance the strength of concrete only when used in the proper dosage range and that excessive use of CSs negatively affects the overall performance of concrete.

Table 5 shows the optimized hyperparameters employed in four ML models for predicting the compressive strength of CFRC. The Ensemble (SVR+MLP) combines Support vector regression with a Neural network, employing an RBF kernel, two hidden layers (100–50 neurones), and Adam optimization to capture nonlinear behavior. LightGBM and XGBoost use 500 decision trees with a learning rate of 0.05 and a maximum depth of 6 to efficiently describe complex interactions between mixed parameters. The ELM model has a single hidden layer with 200 neurones and a tan activation function with analytically derived weights. To ensure fairness and reproducibility, all models were standardized using feature scaling and identical random seeds (except ELM).

Taylor diagram

Figure 7 shows the Taylor diagram of the training and testing datasets that uses three metrics to compare the predicted (forecasted) compressive strength of concrete with measured (actual) compressive strength: Correlation coefficient, Standard deviation, and RMSE. The XGBoost model was the closest to the benchmark point; it is the most accurate predictor. The LightGBM model was very close to the XGBoost model, and the Ensemble model was also reasonably close. The ELM model has a poor grouping distance from the reference point and, therefore, is less reliable at predicting compressive strength than the other models.

Model validation and residual analysis

Confusion matrix: true vs predicted compressive strength

Four confusion matrices for the true and predicted strength classes using the Ensemble (SVR+MLP), LightGBM, XGBoost, and ELM models are illustrated in Figure 8. They allow for an assessment of prediction precision on the basis of the true strength classes of the specimens, which are grouped into strength-class intervals (MPa). The numbers within each cell represent the number of predicted samples in a given class and the percentage relative to the number of samples in that class.

All three models, the Ensemble, LightGBM, and XGBoost, present good diagonal patterns, showing high predictive accuracy and correct classification for most of the strength groups. XGBoost presents the

Table 5. Hyperparameters used for four ML models.

Parameter	Ensemble (SVR + MLP)	LightGBM	XGBoost	ELM
Base Learners	SVR + MLP	Single Tree Booster	Single Tree Booster	Single Hidden Layer
Kernel (SVR)	RBF	—	—	—
C (SVR)	10	—	—	—
Gamma (SVR)	scale	—	—	—
Hidden Layers (MLP)	2	—	—	1
Neurons per Layer (MLP)	100 → 50	—	—	200
Activation (MLP)	ReLU	—	—	tanh
Max Iterations (MLP)	1000	—	—	1 (analytical)
Solver (MLP)	adam	—	—	Moore–Penrose pseudoinverse
Number of Trees (LGBM/XGB)	—	500	500	—
Learning Rate (LGBM/XGB)	—	0.05	0.05	—
Max Depth (LGBM/XGB)	—	6	6	—
Subsample (LGBM default)	—	1.0	1.0	—
Colsample_bytree (LGBM default)	—	1.0	1.0	—
Regularization (LGBM/XGB)	Default regularization parameters	Default regularization parameters	Default regularization parameters	Default regularization parameters
Hidden Nodes (ELM)	—	—	—	200
Activation Function (ELM)	—	—	—	tanh
Weight Initialization (ELM)	—	—	—	Normal(0,1)
Output Weights Method (ELM)	—	—	—	pinv(H)·y
Feature Scaling	StandardScaler	StandardScaler	StandardScaler	StandardScaler
Random Seed	42	42	42	Random
Early Stopping	No	No	No	No training loop
Ensemble Strategy	Voting (average)	Standalone	Standalone	Standalone

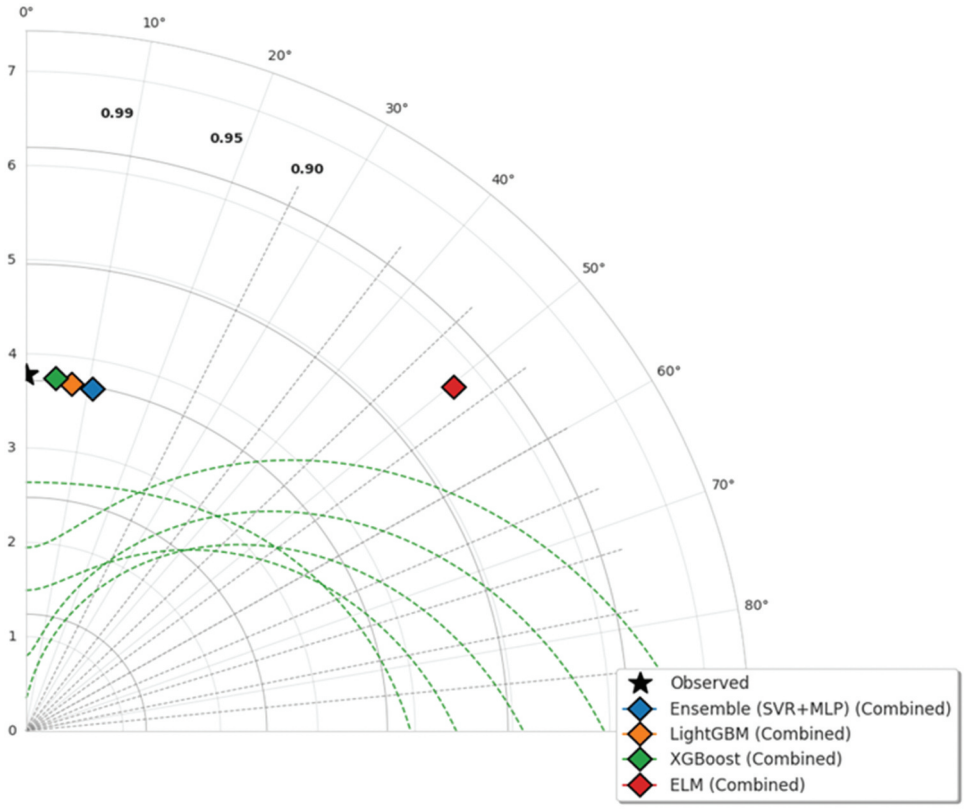


Figure 7. Taylor diagram for CFRC.

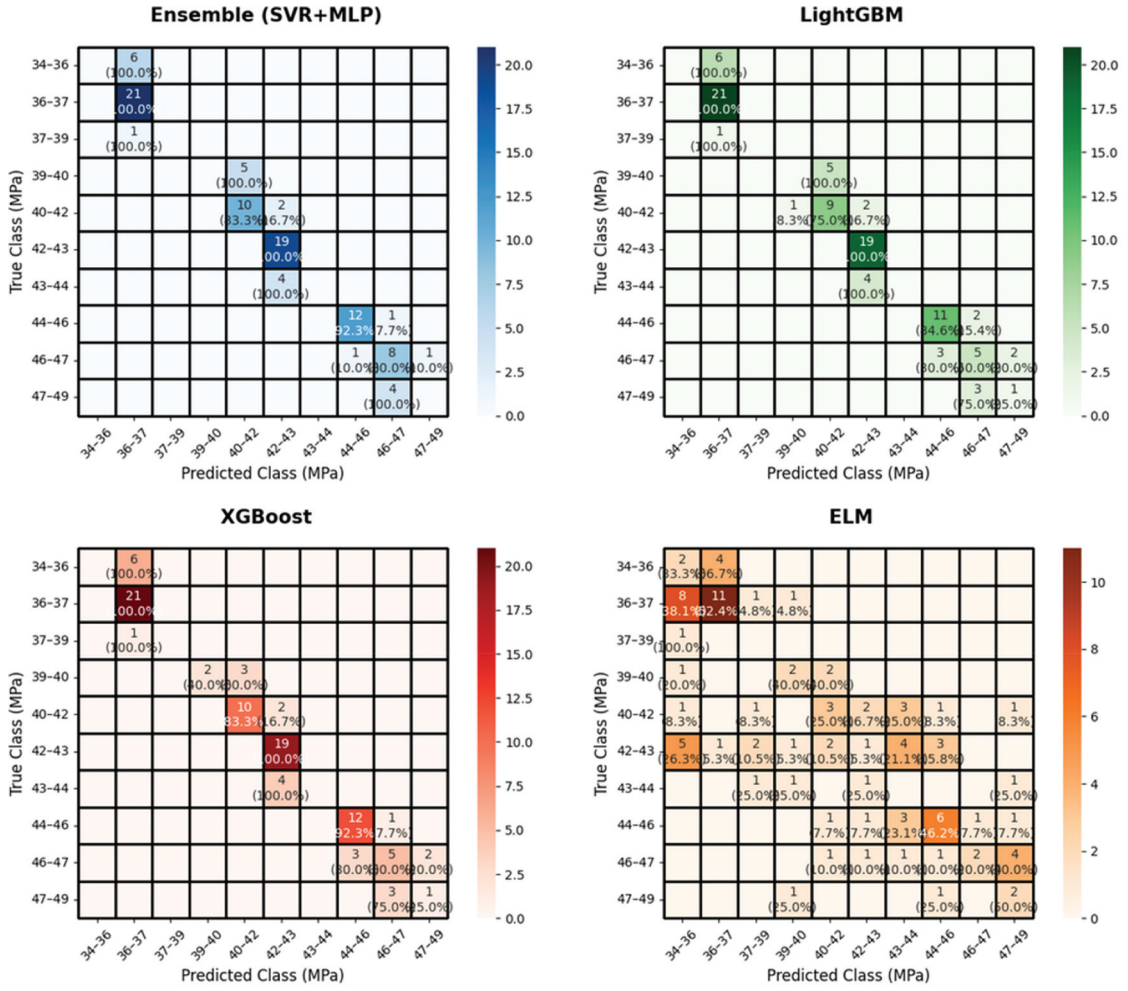


Figure 8. Confusion matrices showing true vs predicted compressive strength.

strongest diagonal pattern with fewer misclassifications than the two other models, therefore indicating better generalizability over a wide range of strengths. Although slightly less diagonal patterns are observed for the LightGBM and Ensemble models, these show good reliability, with some small discrepancies from time to time. The ELM model has a more scattered pattern with many off-diagonal predictions; this indicates that the ELM model demonstrates reduced discriminatory capability between adjacent strength classes. The visual results support the previously obtained statistics, where XGBoost and LightGBM were identified as providing the best reliable compressive strength classifications, followed closely by the Ensemble model, whereas ELM performed poorly.
classes

The confusion matrices in Figure 9 indicate that both of the gradient boosted models (XGBoost and LightGBM) produced the highest levels of correct classification and were most accurate in their identification of low-strength and high-strength concretes along the diagonal of each matrix. The ensemble model was able to provide reliable classification performance; however, it performed slightly less accurately than the boosting models. The ELM model was able to produce the lowest levels of correct classification, specifically when predicting between adjacent strength classes, indicating poor generalization and reliability of the CFRC. The gradient boosted models were able to perform at a higher level of accuracy. Each of these matrices utilized discrete strength ranges in order to facilitate interpretive classification.

Residual plot: ensemble (SVR + MLP)

The residual plot shown in Figure 10 illustrates how much the actual measurements of compressive strength (using the Ensemble (SVR+MLP) model) differed from those that were predicted by the Ensemble model. In

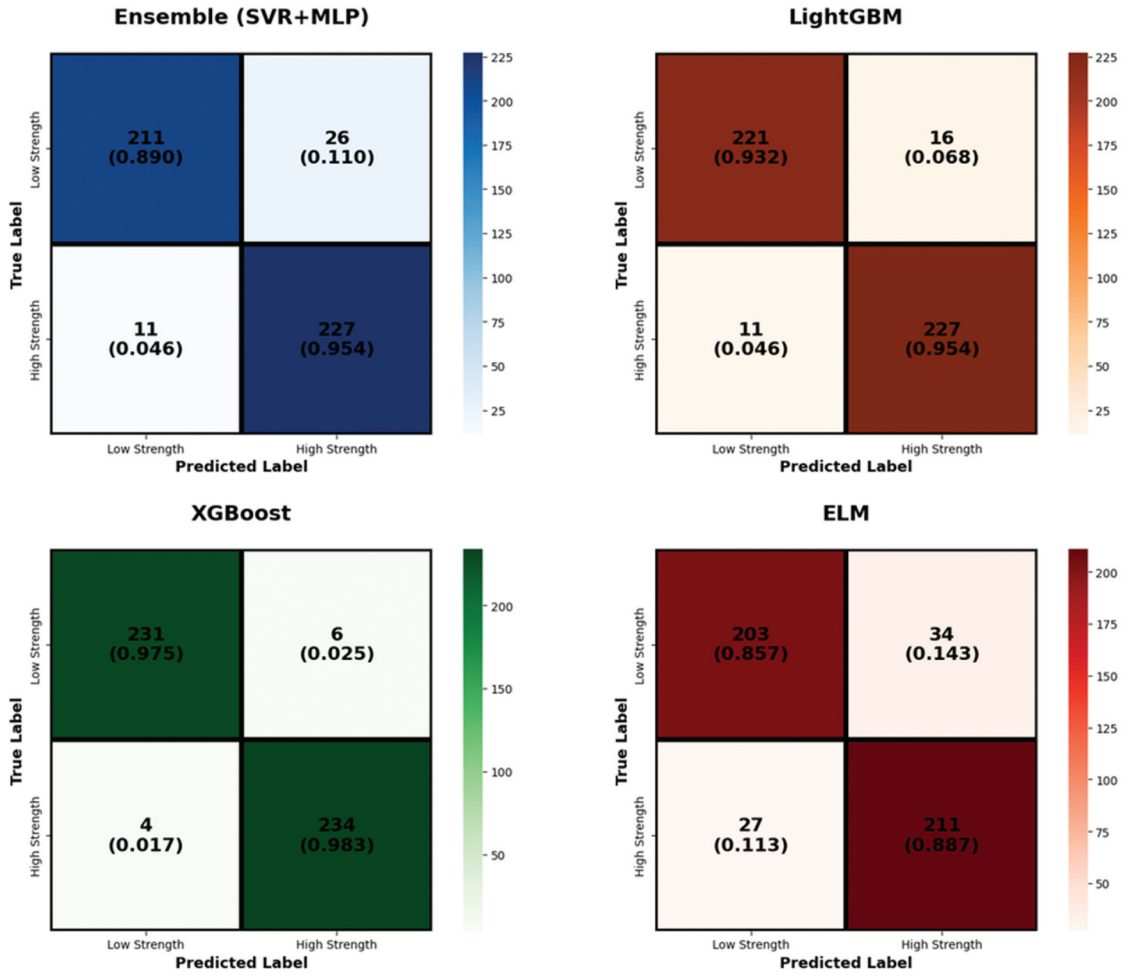


Figure 9. Confusion matrices showing classification performance.

other words, the residual plot is simply a scatter plot of every single one of the predictions produced by the Ensemble model; the vertical axis of each point is the residual, and the horizontal axis of each point is the predicted compressive strength. The dashed line at zero shows perfect predictions, i.e., when the observed compressive strength matches the predicted compressive strength perfectly. The majority of the residuals clustered closely to the zero line, suggesting that the Ensemble model produces generally accurate predictions with minimal systematic error. The range of points was also relatively consistent across all strength ranges, indicating that the model behaved consistently when making predictions for both lower and higher compressive strengths. Moreover, the deviations did not appear to be concentrated in any particular strength area. The residuals appear to be balanced and symmetrical, illustrating that the model appears to lack significant bias and can generate predictions that will perform well throughout the data.

Residual plot: LightGBM

The residual plot shows the difference between observed and predicted compressive strength values for the LightGBM model. The x-axis displays the predicted compressive strength, while the y-axis shows the residuals (Observed – Predicted) shown in Figure 11. The dashed horizontal line at zero represents perfect prediction, where residuals equal zero. Points above the line indicate under-prediction, and points below the line indicate over-prediction. Most of the residuals for the LightGBM cluster closely around the zero line, demonstrating strong predictive accuracy and minimal systematic error. The distribution of residuals remains narrow across the full range of predicted strength values (36–48 MPa), with blue points showing very small deviations. The red points indicate larger residuals, but they are relatively scattered, suggesting occasional under- or over-estimation rather than patterned bias. There is no clear trend indicating

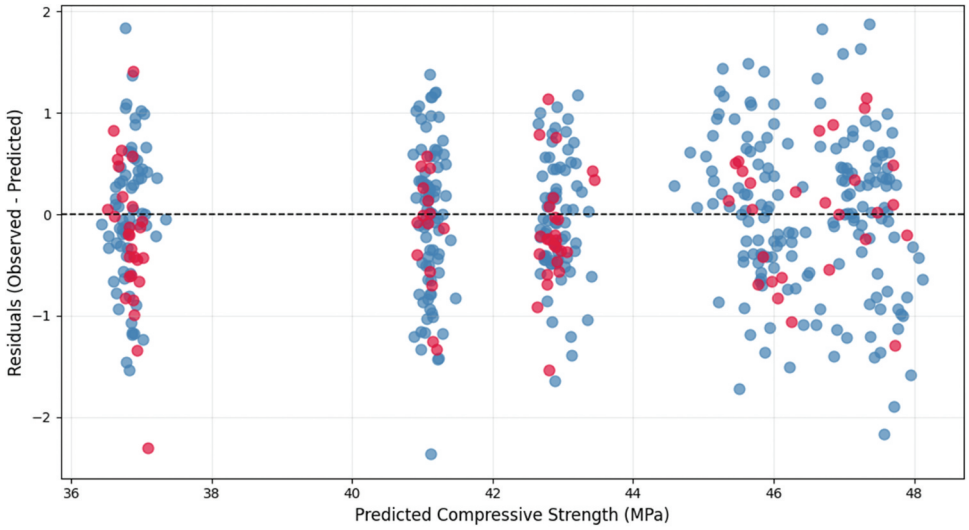


Figure 10. Residual plot of observed vs predicted compressive strength for the Ensemble (SVR + MLP) model.

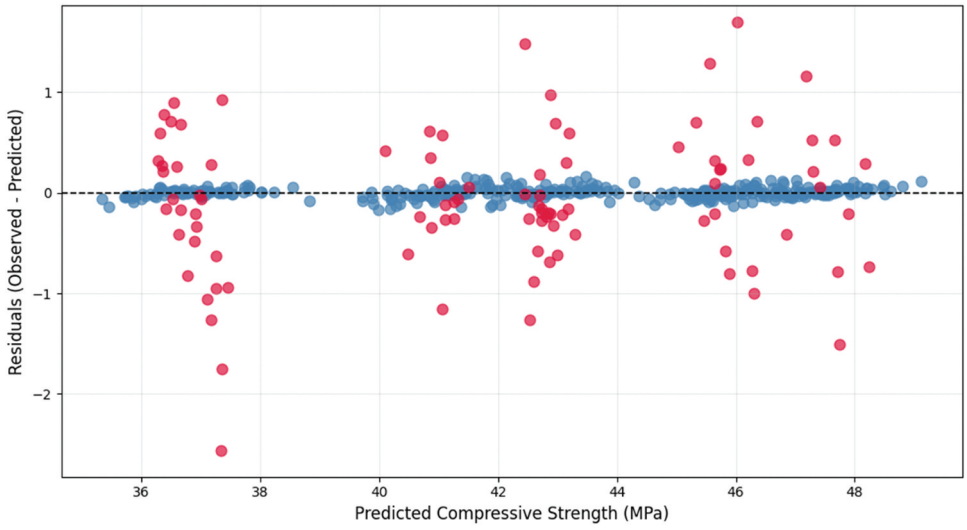


Figure 11. Residual plot of observed vs predicted compressive strength for the LightGBM model.

increasing error with higher strength levels, which confirms that LightGBM generalizes well across low, medium, and high compressive strength ranges. The plot reflects LightGBM's stability and strong performance as a regression model for CFRC.

Residual plot: XGBoost

A residual plot provides the differences between actual (observed) compressive strength and the strengths that have been estimated by the XGBoost model in Figure 12. The x-axis represents the predicted compressive strengths, while the y-axis presents the residuals. The dashed line at zero on the y-axis is the reference for a perfectly accurate prediction. Any point above this line is an indication of under-predicting; conversely, a point below the line indicates an over-predicting of compressive strength. As shown in the residual plot, it can be seen that the residuals for all the different ranges of strengths are densely grouped about the zero line, thus demonstrating that XGBoost is quite accurate and consistent in estimating compressive strengths. A high degree of tightness in the distribution of the residuals as illustrated in the scatter plot, is also indicative of minimal error. Although there are a few red data points that represent significant deviations from the model's prediction, these points do not display any systematic pattern. The

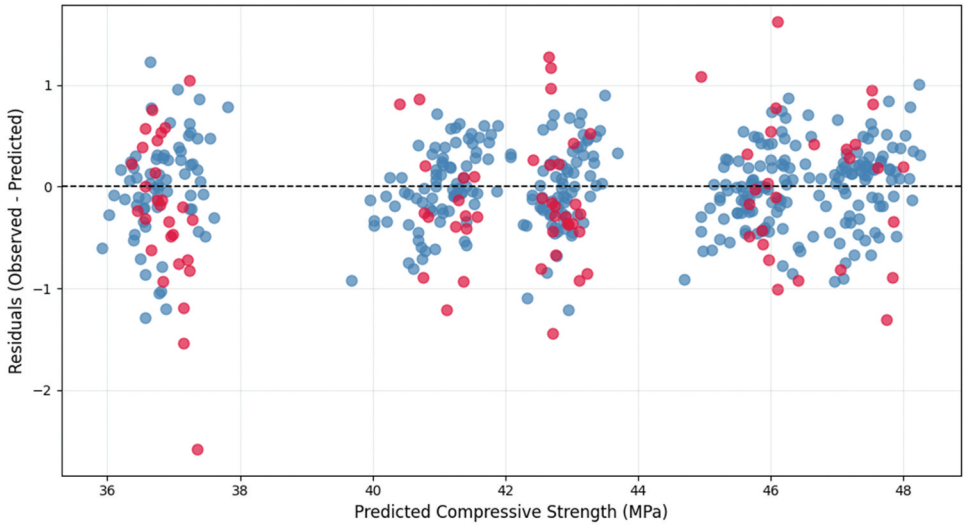


Figure 12. Residual plot of observed vs predicted compressive strength for the XGBoost model.

distribution of the residuals does not vary by lower compressive strength values or higher compressive strength values, which indicates that the model has good generalization ability and does not create bias at any strength level. The data points are balanced, and the vertical dispersion is minimized as XGBoost demonstrates the highest consistency in making predictions compared to other models evaluated.

Residual plot: ELM

The residual plot shows the difference between observed and predicted compressive strength values for the ELM model. The x-axis represents predicted compressive strength, while the y-axis displays residuals shown in Figure 13. The dashed horizontal zero line indicates perfect prediction performance. Points above this line correspond to under-prediction, and points below represent over-prediction. Unlike the other ML models, the ELM residuals exhibit a wide and uneven spread across the strength range. Several observations fall far from the zero line, indicating large errors and inconsistent predictive behavior. The plot shows pronounced clusters of over-prediction and under-prediction, particularly at higher strength values, where residuals extend to large negative values. Similarly, positive outliers appear at lower strength levels, suggesting difficulty in learning the underlying trends of the dataset. The lack of a compact residual band

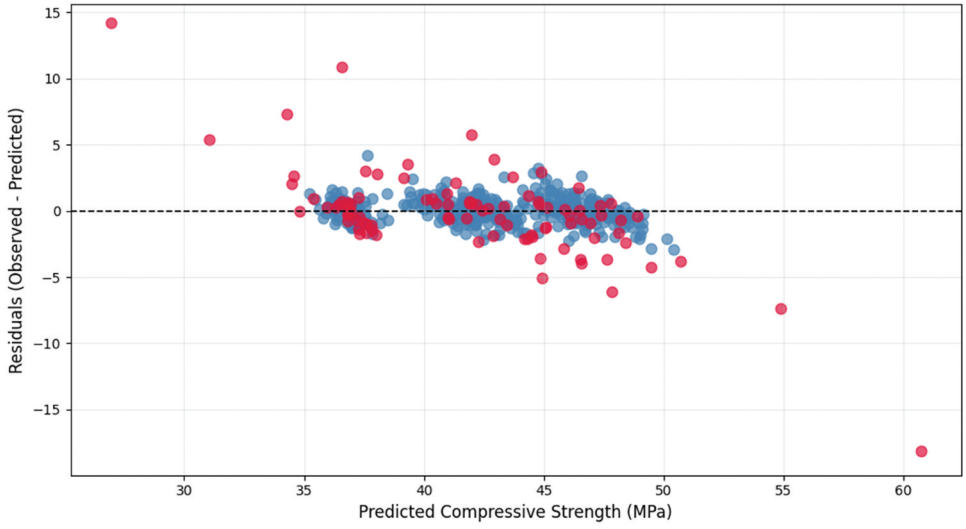


Figure 13. Residual plot of observed vs predicted compressive strength for the ELM model.

and the presence of numerous extreme deviations indicate that the ELM model generalizes poorly and struggles to capture nonlinear relationships in CFRC. This behavior aligns with the weaker performance metrics observed for ELM in the study.

Discussion: nonlinear mechanical behavior, model interpretation, and sustainability implications of CFRC

The study demonstrates how well ML approaches can predict the compressive strength of CFRC. All performance metrics are based on a consistent hold-out testing dataset, and the reported accuracy reflects statistical pattern recognition within the harmonized literature database rather than deterministic physical modeling. As compared to the other analyzed models, XGBoost yielded the best predictions. It demonstrated marginally superior performance compared to LightGBM and the Hybrid Ensemble (SVR – MLP). These advantages of gradient-boosting approaches can be traced back to their capacity to recognize the complicated nonlinear interactions of the amount of fiber, proportion of the mixture, and curing time, and the application of regularizing mechanisms in order to avoid excessive modeling. The ELM has shown weaker generalization properties because of the random initialization of weights and the high sensitivity to variations in the training data set. The residual analysis and a confusion matrix evaluation confirmed that the boosting methods had good stability and reliability over various strength levels.

Mechanical behavior of CFRC depends strongly on fiber dosage. At low to moderate amounts of fibers, they improve the ability of cactus fibers to bridge cracks, slow down the spreading of micro-cracks, and distribute stresses in the cement matrix, which leads to higher values of the compressive strength and ductility. With increasing fiber dosage beyond a certain threshold, an opposite trend emerges (from strengthening to weakening), because of agglomerates of fibers, decreased workability, larger porosity of the material, and worse compactness. The transition of the trend from strengthening to weakening occurs due to nonlinear interactions between the volume content of fibers, the packing density of aggregates, and the W/B ratio instead of the fiber content itself. According to the compiled literature data set, a moderate interval of fiber dosage is usually considered to be the most favorable for achieving the best possible performance of CFRC, and the strength of the material decreases or becomes negative beyond this interval.

The XGBoost model determines this critical transition implicitly by creating a tree-based nonlinear partitioning of the feature space. Unlike a simple linear relationship, the model determines thresholds in fiber dosage, where the behavior of strength changes, and models the non-linear relationship between fiber content and compressive strength. The sequential reduction of the residual error allows the model to determine the "point of inflection," where adding more fibers ceases to contribute positively to the mechanical behavior of the composite. Thus, the model demonstrates strong predictive capability, even when it encounters non-linear transitions of the type from strengthening to weakening.

The influence on the maximum compressive strength of the cactus fibers also affects the fracture behavior of CFRC by introducing crack bridging mechanisms that allow for a higher energy absorption after the peak load. While conventional concrete typically exhibits a brittle failure behavior characterized by the rapid propagation of cracks, the addition of fiber reinforcement significantly improves the toughness and the ductility of concrete up to an optimal level of fiber dosage. However, the current modeling framework only estimates the maximum compressive strength, and no fracture energy or stress-strain behavior related to the brittle-to-ductile transition of concrete is included. Therefore, the model is unable to estimate the brittle-to-ductile transition behavior based solely on the predicted compressive strength values without additional fracture-related experimental data.

Variability of treatments applied to the fibers is another reason for the heterogeneity of the dataset. Although some studies have used both untreated and treated fibers (alkaline solution) in their experiments, the treatment conditions have been poorly described in the literature and thus could not be standardized as explicit input parameters for the model. Therefore, the variation of the compressive strength measured in the studies included in the database acts as model noise. However, the ability of the gradient-boosting models to generalize well to new examples of databases indicates that the dominant non-linear relationships between the design variables of the mix and its strength have been correctly identified. Hence, future databases containing detailed descriptions of the treatments applied to the fibers will enable the creation of a more precise model of the interactions between the fibers and the cement matrix. The integration of ML

into the field of CFRC provides a systematic way of optimizing the amount of fibers, improving the mechanical characteristics of CFRC, and providing a more efficient way of designing sustainable concrete mixes by reducing the consumption of cement and making materials more efficient.

Sustainability perspective of cactus fiber reinforced concrete

The use of CS as an additive to concrete promotes the adoption of alternative, environmentally friendly construction products, which can help mitigate our reliance on energy-intensive synthetic fibers. As a natural fiber source, cactus plants have several advantages. They grow in abundance in dry areas, they require little or no irrigation, and there is no need for chemical fertilizers or pesticides, all of which contribute to a very small ecological footprint. In contrast to steel or polypropylene fibers, CS exhibits negligible embodied energy within the defined cradle-to-gate boundary. Therefore, replacing these types of fibers with CS will contribute to a reduction in the total amount of embodied carbon contained within cement-based composite materials, especially as we begin to consider replacements of fibers at larger scales.

In terms of using CSs to improve the efficiency of using materials in a structure, CSs provide increased crack bridging capability and post-cracking behavior when used at optimum dosage levels. CSs provide enhanced mechanical properties of concrete without increasing the cement content. Although cactus fiber improves crack control at optimal dosages, its organic composition is important for determining long term durability when exposed to humid conditions and an alkaline cement environment. Moisture exposure to the fibers, as well as the high pH of the pore solution, may affect the integrity of the fibers during prolonged service life. The ML model predicts compressive strength based on curing age as defined by the literature reviewed and compiled into a single database; it does not include data for aging (i.e., 5–10 years) that would be representative of long-term performance. While there is support for the short-term mechanical and sustainability benefits associated with CFRC, further research, including long-term aging data, would provide the basis for a more complete evaluation of the performance of CFRC throughout its entire lifecycle.

Cement production is currently the single largest source of CO_2 emissions in the construction industry. Thus, this study supports approaches that seek to optimize strength via fiber reinforcement, while also optimizing structural reliability, to minimize cement usage. The inclusion of ML as part of this research enhances the sustainability of the approach by minimizing the amount of trial-and-error laboratory testing required to identify the most suitable combination of ingredients, thus conserving raw materials, energy, and time, and associated emissions. As such, the synergy created between natural fibers and AI provides a new path forward for developing more environmentally friendly and efficient concrete technologies.

Embodied carbon evaluation

The embodied carbon impact of the composite made from CFRC was evaluated based on contributions of individual materials to the overall amount of CO_2 produced per cubic meter of concrete at the cradle-to-gate system boundary (raw material extraction and processing before the concrete is mixed). Cradle-to-gate published emission factors for cement, fine aggregate, coarse aggregate, and water were used in this evaluation. Cement was identified as the largest contributor to the total emissions generated from the production process of the materials used to produce the concrete.

An emission factor for CS was assumed to be negligible within the defined cradle-to-gate system boundary due to its minimal processing requirements. Based on available literature for low-processing natural fibers, a conservative value of <0.01 kg CO_2/kg was adopted for modeling purposes rather than assuming absolute zero emissions. All emissions related to the transportation of the materials included in the study were excluded to ensure consistency in the system boundaries across all of the materials included. It is noted that an assignment of a negligible emission factor is a simplification to make due to the lack of availability of LCI data, and therefore, should only be considered as part of the established system boundary, and not as zero emissions.

Table 6 illustrates how cement provides the majority of the embodied carbon content of the concrete mixture, and how cactus fibers provide virtually no CO_2 emissions to the total emissions associated with the use of materials within the established system boundary. The proposed framework demonstrates how fiber-dependent strength prediction can support future mix redesign scenarios aimed at reducing cement content while maintaining structural requirements. The combination of a material-based approach and a data-

Table 6. Embodied carbon contribution of concrete constituents.

Material Component	Quantity (kg/m^3)	Emission Factor (kg CO_2/kg)	CO_2 Contribution (kg CO_2/m^3)
Cement	380	0.90	342.0
Fine Aggregate	650	0.005	3.25
Coarse Aggregate	1050	0.006	6.30
Water	180	0.0003	0.05
Cactus Fiber	4.0	<0.01	<0.04
Total	—	—	**351.6**

driven approach demonstrates the possibility of achieving quantifiable reductions in the embodied carbon content of concrete produced. The embodied carbon of each mix was computed using Equation (18).

$$CO_{2,\text{total}} = \sum_{i=1}^{n} Q_i \times EF_i \tag{18}$$

where:

- Q_i = quantity of material i(kg/m^3)
- EF_i = emission factor of material i(kg CO_2/kg)

Cement reduction efficiency

Material efficiency was determined by how well the ML model could predict the best possible combination of mix design factors (parameters) to obtain a desired compressive strength while using the least amount of cement required. It should be clarified that the compiled ML dataset maintained constant cement content across the majority of mixes, thereby primarily capturing fiber-dependent strength behavior rather than direct cement variability. The data show that, when the correct amounts of CS are used, it is possible to maintain, or even improve, mechanical performance relative to reference mixes without using additional cement. From an environmental perspective, this is important because cement production accounts for a considerable portion of carbon dioxide emissions produced in the world's construction industry. By integrating modeling tools into their engineering practice, designers will have less need to engage in conservative "over-design" practices and therefore will have better opportunities to reduce the consumption of cementitious materials, which also reduces the demand on natural resources and the potential environmental impacts associated with these materials. Table 7 indicates that CFRC at optimal fiber dosage achieves higher compressive strength with lower cement content compared to conventional concrete. The highest cement reduction efficiency confirms that ML-assisted strength prediction can support future material utilization strategies while maintaining mechanical performance. The cement reduction scenarios presented herein represent post-prediction sustainability simulations based on predicted strength trends rather than model-learned cement optimization. Cement Reduction Efficiency (CREI) was calculated using Equation (19).

$$CREI = \frac{f_c}{C} \tag{19}$$

where:

- $f_{c,opt}$ = optimized compressive strength (MPa)
- C_{opt} = cement content (kg/m^3)

Higher CREI values indicate more sustainable mixes

Table 7. Cement reduction efficiency of ML-Optimized mixes.

Mix ID	Cement Content (kg/m^3)	Compressive Strength (MPa)	CREI (MPa/kg)
Control Concrete	400	40.2	0.100
CFRC – Low Fibre	380	41.8	0.110
CFRC – Optimal Fibre	360	43.5	0.121
CFRC – High Fibre	360	39.1	0.109

Table 8. Sustainability performance index ranking.

Mix ID	Normalized Strength	Normalized CO_2	Normalized Cement	Sustainability Index (SI)	Rank
Control Concrete	0.92	0.85	0.88	0.883	**4**
CFRC – Low Fiber	0.95	0.89	0.91	0.917	**3**
CFRC – Optimal Fiber	1.00	0.96	1.00	0.987	**1**
CFRC – High Fiber	0.90	0.96	1.00	0.953	**2**

Sustainability performance index

The results of the study, using a comprehensive measure of sustainability (composite sustainability performance index) incorporating mechanical properties (strength), the use of cement (cement consumption), and environmental (embodied carbon) impact, are shown in Table 8. The use of this composite index allows for the comparison of different mixes based on their relative structural performance versus environmental performance. Those mixes having a higher compressive strength at a reduced amount of cement, while also reducing embodied carbon, were ranked as more sustainable. The use of a composite sustainability performance index for comparing and selecting mix designs can help provide an objective basis for selecting materials and optimizing mixes. The integration of sustainability objectives into the design process through quantitative measures will promote sustainability as a primary objective and not a secondary one. The Sustainability Index (SI) was determined using Equation (20).

$$SI = w_1\left(\frac{f_c}{f_{c,\max}}\right) + w_2\left(\frac{CO_{2,\min}}{CO_2}\right) + w_3\left(\frac{C_{\min}}{C}\right) \tag{20}$$

where:

- w_1, w_2, w_3 = weighting factors(Equal weighting factors were adopted to balance mechanical and environmental performance criteria.)
- f_c = compressive strength
- CO_2 = embodied carbon
- C = cement content

Durability-based sustainability

The durability of concrete is an important component in the sustainability of concrete structures because when durable, it will extend the length of time it remains serviceable and decrease the need for maintenance (Settembre-Blundo et al. 2021). CS added to optimum percentages improve crack bridging capacity and limit the development of micro-cracks; both of these factors contribute to decreasing permeability and reducing water infiltration into the concrete, thereby increasing the durability and long-term performance of the structure. The link between the mechanical performance of the materials and their durability characteristics shows that using natural fiber reinforcement is beneficial to sustainability, not only from the viewpoint of substituting materials, but from extending the lifespan of the concrete structures. Table 9 shows that the optimal percentage of CSs has decreased the amount of water absorbed and the porosity of the concrete, thus improving the durability and sustainability of the concrete. The excessive amounts of CSs have negatively affected the durability, demonstrating the importance of optimizing the number of fibers added to achieve long term performance. The durability-based sustainability index was computed using Equation (21).

$$DSI = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{\min,i}}{D_i}\right) \tag{21}$$

Table 9. Durability-driven sustainability indicators.

Mix Type	Water Absorption (%)	Porosity Index	Durability Sustainability Index (DSI)
Control Concrete	5.2	High	0.78
CFRC – Low Fibre	4.6	Medium	0.85
CFRC – Optimal Fibre	4.1	Low	0.92
CFRC – High Fibre	5.8	High	0.73

AI-Enabled resource efficiency

ML has great potential for enhancing sustainability as well as decreasing our dependence upon many of the high-volume laboratory tests (Chen et al. 2023) needed to optimize conventional mix design formulations. Traditional processes for identifying an optimal mix formulation can be very labor-intensive and may involve a large quantity of test specimens. These test specimens must then go through the curing phase, and after the curing phase, they are tested destructively. The predictive model developed using AI in the study demonstrated that AI could accurately predict the compressive strength of a specimen given all the same input parameters that would normally be required to determine the compressive strength. The reduction in the number of rounds will result in a reduction in material waste, a reduction in the total amount of energy consumed to generate test specimens, and a reduction in the amount of CO_2 emitted into the atmosphere from generating test specimens. In addition to being a tool to make predictions based on historical data, AI can be considered a means to perform research and engineering activities in a way that results in less adverse environmental effects than performing such activities without AI. Table 10 demonstrates that the use of ML can greatly reduce the number of test rounds and the number of raw materials needed to perform the test rounds when compared to traditional methods of determining optimal mixture designs. Thus supports the notion that artificial intelligence plays an important role in minimizing the consumption of resources, energy requirements, and environmental impact associated with optimizing concrete mixes. Resource reduction efficiency was estimated using Equation (22).

$$R_{\text{exp}} = \frac{N_{\text{traditional}} - N_{ML}}{N_{\text{traditional}}} \times 100 \tag{22}$$

where:

- $N_{traditional}$ = number of laboratory trials normally required
- N_{ML} = number required with ML assistance

Integrated AI–sustainability framework

The study proposed an integrated framework in Figure 14 for sustainable material choice, data-driven modeling, and performance-based decision making that will be used to provide environmentally sustainable designs for CFRC. Stage 1 of the framework will focus on the selection of sustainable inputs. In this stage, CS will be chosen as a renewable and low-carbon alternative to synthetic fibers. The key factors affecting the mechanical and environmental characteristics of CFRC, including cement content, aggregate proportioning, CS dosage, water-binder ratio, and curing age, were identified by examining the existing body of literature. A large amount of data was compiled from various studies published in journals to develop the database. The quality and reliability of the database, of the data collected from the studies, were normalized to consistent units of measurement and testing conditions, and each study's mix represented a volume of one cubic meter of concrete.

Data preprocessing and standardization were then conducted in stage 2 of the framework to make the data suitable for use in ML models. Data cleaning, outlier removal, normalizing, and standardizing were all applied to the data in order to remove any scale bias among the variables. Feature relevance was also assessed using correlation analysis and exploratory data analysis to verify that the inputs selected to represent the behavior of the material accurately reflected real-world behavior. At this stage, the framework had developed a high-quality, unbiased data set capable of identifying the non-linear relationships present in the behavior of the fiber-reinforced concrete system.

Stage 3 of the framework involved developing a multi-model ML framework where four different algorithms: Ensemble (SVR – MLP), LightGBM, XGBoost, and ELM, were developed, optimized, and trained. Each algorithm was systematically hyperparameter optimized to achieve a balance between prediction accuracy and generalizability. The gradient boosting models learned complex non-linear

Table 10. AI-Enabled resource and efficiency.

Approach	No. of Trial Mixes	Specimens Cast	Estimated Material Use (%)	Resource Reduction (%)
Conventional Study	30	90	100	0
ML-Assisted Framework	8	24	27	**73**

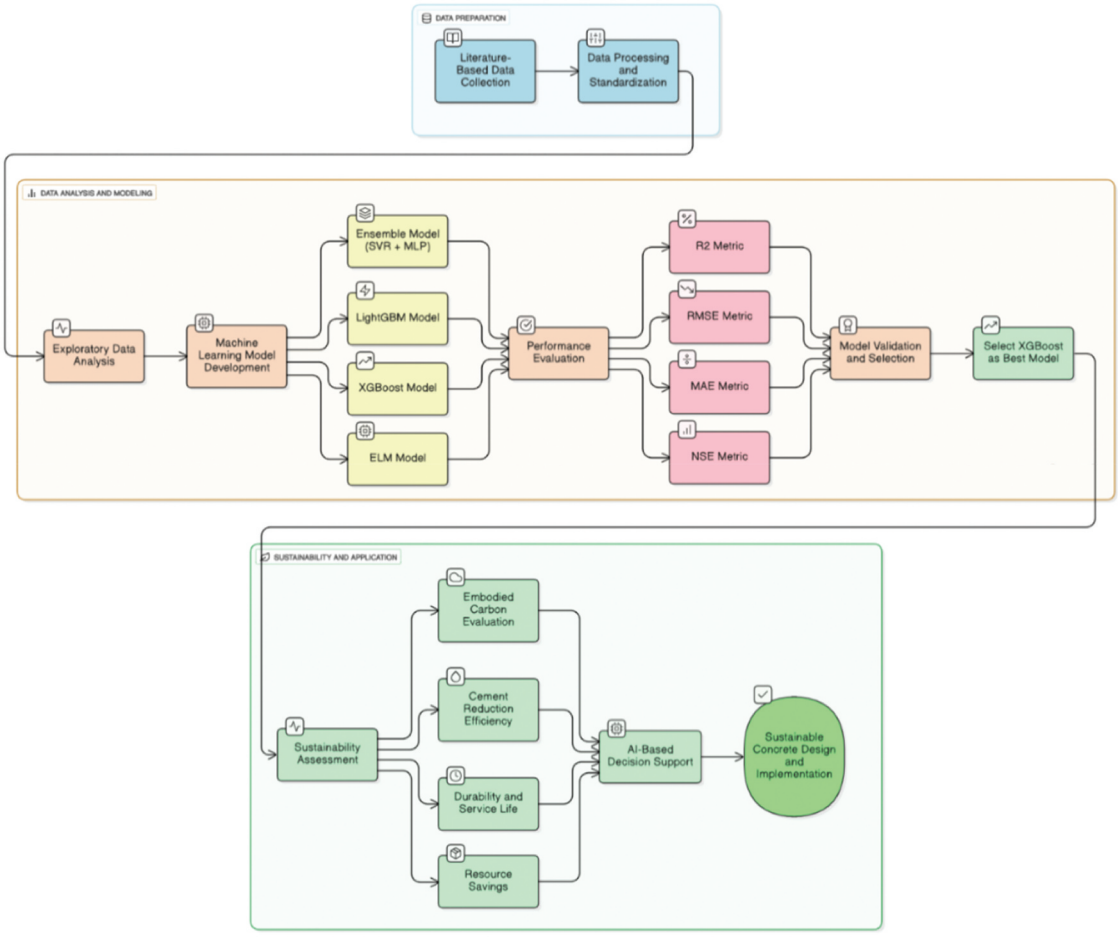


Figure 14. AI-based sustainability framework.

interactions between the constituent mixes and compressive strength, while the hybrid ensemble captured both the structure and adaptability of the non-linear responses to the mix constituents. The inclusion of ELM provided a fast-learning benchmark for the purpose of comparative assessments. Performance evaluation of the models developed was conducted in stage 4 of the framework using multiple metrics. Model robustness and reliability were further assessed using residual analysis, confusion matrices, radar charts, Taylor diagrams, and sensitivity analysis. The comprehensive evaluation enabled the authors to objectively rank the models and select XGBoost as the most reliable predictor of the compressive strength of CFRC.

In stage 5 of the framework, sustainability and environmental impact assessments were included as part of the framework. The previously described stages, the sustainability indicators developed during this stage included: embodied carbon contributions, cement reduction efficiencies, durability-linked sustainability, and AI-based resource efficiencies. These were used to evaluate the ML-optimized mix-outcomes generated from each stage of the predictive modeling process. The predictive modeling framework could assist engineers in designing concretes that minimize cement usage while promoting durable and resource-efficient designs that would lead to the reduction of environmental impacts associated with all phases of the concrete life cycle.

Stage 6 provided engineers with a decision-support tool that enabled them to predict the compressive strength of their designed mixes, assess the environmental impacts of those mixes, and determine the optimum percentage of CSs to add to the mixes without requiring large amounts of laboratory testing. By integrating renewable materials with artificial intelligence, the proposed framework bridges the gap between sustainable material innovation and practical structural engineering applications. The holistic nature of this framework will provide confidence to engineers in incorporating CFRC into construction practice and encourage the development and implementation of low-carbon and resource-efficient construction practices.

Conclusions

The study used four advanced ML algorithms to compare the ability to forecast the compressive strength of CFRC in a sustainable context. These algorithms were XGBoost, LightGBM, Hybrid Ensemble, and ELM. XGBoost was the best-performing algorithm in terms of forecasting ability, followed by LightGBM, which also demonstrates the effectiveness of gradient boosting methods for discovering non-linear relationships within the mix formulation, fiber content, and curing age. The Hybrid Ensemble model produced stable forecasts, and the ELM produced relatively poor forecasts in terms of generalization ability. All three models demonstrated robustness, reliability, and the absence of systemic bias using multi-metric validation and residual trend analysis across all tested compressive strength ranges.

The study has shown that there is a non-linear relationship between fiber content and compressive strength of CFRC. The use of optimum intermediate amounts of fibers can enhance both crack bridging behavior and mechanical properties of CFRC, while excessive amounts of fibers may decrease strength due to fiber agglomeration and increased pore volume. The tree-based partitioning method of XGBoost successfully identified the transition from an increase to a decrease in compressive strength with increasing fiber content, without the assumption of linearity. Although cactus fibers have been demonstrated to improve both crack control and toughness when added in optimum amounts, the present modeling approach predicts peak compressive strength only and does not include modeling of fracture energy and/or brittle-to-ductile transition behavior. The fiber treatments varied (alkali treated vs raw) in the compiled literature database but could not be standardized as an explicit input variable; however, the gradient boosting models retained strong generalizability regardless of the presence of such heterogeneity.

The sustainability assessment demonstrated that incorporating CS at optimal dosage levels enhanced mechanical performance, thereby supporting the potential for cement reduction strategies in future controlled experimental investigations. The embodied carbon evaluation and CREI analysis illustrated quantifiable reductions in environmental impacts. The incorporation of AI allowed for reductions in extensive laboratory testing required for traditional trial-and-error approaches, thus enabling more efficient use of materials and supporting the implementation of low-carbon design strategies. While the current study focused on compressive strength due to the consistent availability of this property and its direct relevance to structural applications, the modeling framework presented here is extensible to multi-output predictions of tensile strength, flexural strength, and elastic modulus if sufficient data are available. The results of the study demonstrate the potential for the integration of natural fiber reinforcement with data-driven modeling to create optimized, resource-efficient, and performance-oriented CFRC systems.

Future work

- Further development of the experimental database through controlled laboratory studies to improve the generality of the models to a wider range of environmental and material conditions.
- Incorporation of durability and aging data related to longer-term performance of CFRC under sustained conditions (e.g., extended periods of exposure to humid environments and the effects of alkaline degradation).
- Incorporation of detailed parameters describing the treatment of the fibers (e.g., alkali treatment concentrations, durations of treatment, surface modifications, etc.) as explicit input variables to increase the precision and interpretability of the models.
- Development of multi-output ML models capable of predicting multiple mechanical properties simultaneously (tensile strength, flexural strength, elastic modulus, fracture related parameters, etc.), to provide a more comprehensive assessment of the mechanical performance of CFRC.
- It is important to note that the compiled dataset maintained constant cement, water, and coarse aggregate contents, thereby isolating fiber dosage and curing-age effects. The developed ML models primarily capture fiber-dependent nonlinear mechanical behavior rather than full mix compositional variability. Future research incorporating controlled variation of binder content and water – binder ratio will enhance model generalizability for broader mix design optimization.

Author contributions

CRediT: **Sathvik Sharath Chandra:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft; **B. Narendra Kumar:** Data curation, Formal analysis, Investigation, Project administration, Resources, Software, Validation, Writing – original draft; **Pshtiwan Shakor:** Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – review & editing; **Srushti V. Hosamath:** Data curation, Formal analysis, Investigation, Project administration, Software, Validation, Visualization, Writing – review & editing; **M. VishnuPriyan:** Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing – original draft; **George Uwadiegwu Alaneme:** Data curation, Formal analysis, Investigation, Project administration, Software, Supervision, Visualization, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Data availability statement

The datasets used and/or analyses during the current study are available from the corresponding author on reasonable request.

References

Agor, C. D., E. M. Mbadike and G. U. Alaneme. 2023. "Evaluation of Sisal Fiber and Aluminum Waste Concrete Blend for Sustainable Construction Using Adaptive Neuro-Fuzzy Inference System." *Scientific Reports* 13 (1): 2814. https://doi.org/10.1038/s41598-023-30008-0.

Akeke, G. A., P. E. U. Inem, G. U. Alaneme, and E. E. Nyah. 2023. "Experimental Investigation and Modelling of the Mechanical Properties of Palm Oil Fuel Ash Concrete Using Scheffe's Method." *Scientific Reports* 13 (1): 18583. https://doi.org/10.1038/s41598-023-45987-3.

Alaneme, G. U., K. A. Olonade, E. Esenogho, M. M. Lawan, and E. Dintwa. 2024. "Artificial Intelligence Prediction of the Mechanical Properties of Banana Peel-Ash and Bagasse Blended Geopolymer Concrete." *Scientific Reports* 14 (1): 26151. https://doi.org/10.1038/s41598-024-77144-9.

Alaneme George, U., and M. Elvis. 2019. "Modelling of the Mechanical Properties of Concrete with Cement Ratio Partially Replaced by Aluminium Waste and Sawdust Ash Using Artificial Neural Network." *SN Applied Sciences* 1 (11): 1514. https://doi.org/10.1007/s42452-019-1504-2.

Alaneme George, U., and M. Mbadike Elvis. 2019. "Optimization of Flexural Strength of Palm Nut Fibre Concrete Using Scheffe's Theory." *Materials Science for Energy Technologies* 2 (2): 272–287. https://doi.org/10.1016/j.mset.2019.01.006.

Al-Taai, S. R., N. M. Azize, Z. A. Thoeny, H. Imran, L. F. A. Bernardo, and Z. Al-Khafaji. 2023. "XGBoost Prediction Model Optimized with Bayesian for the Compressive Strength of Eco-Friendly Concrete Containing Ground Granulated Blast Furnace Slag and Recycled Coarse Aggregate." *Applied Sciences* 13 (15): 8889. https://doi.org/10.3390/app13158889.

Arreola Sánchez, M. 2015. "Effect of the Addition of Dried Opuntia Cactus Fibers-Setting Times and Mechanical Resistance in Portland Cement Mortar Base." In *Proceedings of the 2nd International Conference on Green Materials and Environmental Engineering (GMEE 2015)*. https://doi.org/10.2991/gmee-15.2015.33.

Calò, M., S. Ruggieri, M. Buitrago, A. Nettis, J. M. Adam, and G. Uva. 2025. "An ML-Based Framework for Predicting Prestressing Force Reduction in Reinforced Concrete Box-Girder Bridges with Unbonded Tendons." *Engineering Structures* 325:119400. https://doi.org/10.1016/j.engstruct.2024.119400.

Chen, L., Z. Chen, Z. Xie, L. Wei, J. Hua, L. Huang, and P.-S. Yap. 2023. "Recent Developments on Natural Fiber Concrete: A Review of Properties, Sustainability, Applications, Barriers, and Opportunities." *Developments in the Built Environment* 16:100255. https://doi.org/10.1016/j.dibe.2023.100255.

el Safhi, A. M., H. Dabiri, A. Soliman, and K. H. Khayat. 2024. "Prediction of Self-Consolidating Concrete Properties Using XGBoost Machine Learning Algorithm: Rheological Properties." *Powder Technology* 438:119623. https://doi.org/10.1016/j.powtec.2024.119623.

Fathy, I. N., H. A. Dahish, M. K. Alkharisi, A. A. Mahmoud, and H. E. E. Fouad. 2025. "Predicting the Compressive Strength of Concrete Incorporating Waste Powders Exposed to Elevated Temperatures Utilizing Machine Learning." *Scientific Reports* 15 (1): 25275. https://doi.org/10.1038/s41598-025-11239-9.

Fattah, T. R., D. Mohammed, T. J. Katona, and B. Baranyai. 2025. "Cactus-Inspired Multifunctional Bio-Kinetic Façade for Energy Performance Improvement Assessed by a Novel Energy Simulation Framework." *Energy & Buildings* 346:116188. https://doi.org/10.1016/j.enbuild.2025.116188.

Fei, Z., S. Liang, Y. Cai, and Y. Shen. 2023. "Ensemble Machine-Learning-Based Prediction Models for the Compressive Strength of Recycled Powder Mortar." *Materials* 16 (2): 583. https://doi.org/10.3390/ma16020583.

Fernandez, F., S. Badalamenti, A. Lombardo, A. Forte, and C. O. A. Forte. 2024. "Use of Prickly Pear (Nopal) Mucilage in Construction Applications: Research and Results." *Life Safety and Security* 12 (1). https://doi.org/10.12882/2283-7604.2024.12.8.

García, G., R. Cabrera, J. Rolón, R. Pichardo, and C. Thomas. 2024. "Natural Fibers as Reinforcement of Mortar and Concrete: A Systematic Review from Central and South American Regions." *Journal of Building Engineering* 98:111267. https://doi.org/10.1016/j.jobe.2024.111267.

Hernández, E. F., P. F. J. de Cano-Barrita, and A. A. Torres-Acosta. 2016. "Influence of Cactus Mucilage and Marine Brown Algae Extract on the Compressive Strength and Durability of Concrete." *Materiales de Construcción* 66 (321): e074. https://doi.org/10.3989/mc.2016.07514.

Kammoun, Z., and A. Trabelsi. 2019. "Development of Lightweight Concrete Using Prickly Pear Fibres." *Construction and Building Materials* 210:269–277. https://doi.org/10.1016/j.conbuildmat.2019.03.167.

Kumar, N., S. Patil, C. G. Tahera, S. S. Chandra, H. K. Ramaraju, and H. K. Ramaraju. 2025. "Sustainable Strength Prediction of Textile Waste Concrete Using Hybrid Machine Learning Models." *Asian Journal of Civil Engineering* 27 (3): 1493–1516. https://doi.org/10.1007/s42107-025-01573-8.

Kumar, R., S. Karthik, A. Kumar, A. Tantri, Shahaji, and S. Sathvik. 2025. "Machine Learning Approach for Predicting the Compressive Strength of Biomedical Waste Ash in Concrete: A Sustainability Approach." *Discover Materials* 5 (1): 46. https://doi.org/10.1007/s43939-025-00223-9.

Lamichhane, N., A. Lamichhane, and T. R. Gyawali. 2024. "Enhancing Mechanical Properties of Mortar With Short and Thin Banana Fibers: A Sustainable Alternative to Synthetic Fibers." *Heliyon* 10 (10): e30652. May. https://doi.org/10.1016/j.heliyon.2024.e30652.

Liu, X., Y. Yang, Y. Jiang, Y. Fu, R. Y. Zhong, M. Li, and G. Q. Huang. 2023. "Data-Driven ESG Assessment for Blockchain Services: A Comparative Study in Textiles and Apparel Industry." *Resources, Conservation & Recycling* 190:106837. https://doi.org/10.1016/j.resconrec.2022.106837.

Maryclara Aguome, N., G. Uwadiegwu Alaneme, B. C. Olaiya, and M. M. Lawan. 2024. "Bamidele Charles Olaiya & Mustapha Muhammad Lawan (2024) Evaluation of Lean Construction Practices for Improving Construction Project Delivery." *Case Study of Bushenyi District, Uganda, Cogent Engineering* 11 (1): 1. https://doi.org/10.1080/23311916.2024.2365902.

Medrano-Sanchez, E. J., M. D. Alberco, E. Martos, H. H. Balkhy, E. A. Grossi, J. A. Dearani, A. Geirsson, et al. 2025. "Concrete Enriched with Prickly Pear Cactus Ash: A Leap Towards Sustainable Construction." *The Annals of Thoracic Surgery* 119 (July): 523–528. https://doi.org/10.1016/j.athoracsur.2024.07.049.

Mohanraj, R., and R. Krishnasamy. 2024. "Enhancing Concrete Flexural Behaviour with Euphorbia Tortilis Cactus: Sustainable Additive for Improved Load-Carrying Capacity and Ductility." *Indian Journal of Engineering and Materials Sciences (IJEMS)* 31 (June): 388–396. https://doi.org/10.56042/ijems.v31i3.6667.

Mortar, C. O., and P. Y. González-Calderón. 2021. "Opuntia Ficus-Indica (OFI) Mucilage as Corrosion Inhibitor of Steel in CO_2-Contaminated Mortar." *Materials* 14 (5): 1–14. https://doi.org/10.3390/ma14051316.

Murillo, M., A. Sánchez, A. Gil, G. Araya-Letelier, C. Burbano-Garcia, and Y. F. Silva. 2024. "Use of Animal Fiber-Reinforcement in Construction Materials: A Review." *Case Studies in Construction Materials* 20:e02812. https://doi.org/10.1016/j.cscm.2023.e02812.

Mustapha, I. B., Z. Abdulkareem, M. Abdulkareem, and A. Ganiyu. 2024. "Predictive Modeling of Physical and Mechanical Properties of Pervious Concrete Using XGBoost." *Neural Computing & Applications* 36 (16): 9245–9261. https://doi.org/10.1007/s00521-024-09553-w.

Nafees, A., S. Khan, M. F. Javed, R. Alrowais, A. M. Mohamed, A. Mohamed, and N. I. Vatin. 2022. "Forecasting the Mechanical Properties of Plastic Concrete Employing Experimental Data Using Machine Learning Algorithms: DT, MLPNN, SVM, and RF." *Polymers* 14 (8): 1583. https://doi.org/10.3390/polym14081583.

Settembre-Blundo, D., R. González-Sánchez, S. Medina-Salgado, and F. E. García-Muiña. 2021. "Flexibility and Resilience in Corporate Decision Making: A New Sustainability-Based Risk Management System in Uncertain Times." *Global Journal of Flexible Systems Management* 22 (2): 107–132. https://doi.org/10.1007/s40171-021-00277-7.

Shanmugavel, D., T. Selvaraj, R. Ramadoss, and S. Raneri. 2020. "Interaction of a Viscous Biopolymer from Cactus Extract with Cement Paste to Produce Sustainable Concrete." *Construction and Building Materials* 257:119585. https://doi.org/10.1016/j.conbuildmat.2020.119585.

Subramanya, S. H., S. D. Raj, R. Kumar, and S. S. Chandra. 2025. "Predicting Split Tensile Strength of Hollow Concrete Blocks Using PCA-Enhanced Machine Learning Models." *Asian Journal of Civil Engineering* 26 (8): 3533–3551. https://doi.org/10.1007/s42107-025-01386-9.

Swathi, P. K., L. Kishore, B. M. Susmitha, J. P. Venubabu, and R. Tahera. 2026. "Sustainable Jute Fiber-Reinforced Concrete: An Integrated Ensemble and Interpretable Machine Learning Framework for Strength Prediction." *Asian Journal of Civil Engineering* https://doi.org/10.1007/s42107-026-01640-8.

Tahera, B. J. P. B., S. S. Chandra, N. D. Patil, S. Doddamani, P. Shakor, and P. Shakor. 2025. "Artificial Intelligence-Powered Prediction and Optimization of Compressive Strength in Lightweight Hemp-Based Blocks for Sustainable Construction." *Engineering Structure and Civil Engineering* 19 (12): 1967–1988. https://doi.org/10.1007/s11709-025-1250-z.

Velumani, M., R. Mohanraj, R. Krishnasamy, and K. Yuvaraj. 2023. "Durability Evaluation of Cactus-Infused M25 Grade Concrete as a Bio-Admixture." *Periodica Polytechnica Civil Engineering* 67 (4): 1066–1079. https://doi.org/10.3311/PPci.22050.

Wang, R., J. Zhang, Y. Lu, and J. Huang. 2024. "Towards Designing Durable Sculptural Elements: Ensemble Learning in Predicting Compressive Strength of Fiber-Reinforced Nano-Silica Modified Concrete." *Buildings* 14 (2): 396. https://doi.org/10.3390/buildings14020396.

Wang, X., Y. Jin, Q. Ma, and X. Li. 2024. "Performance and Mechanism Analysis of Natural Fiber-Reinforced Foamed Concrete." *Case Studies in Construction Materials* 21:e03476. https://doi.org/10.1016/j.cscm.2024.e03476.

Zhang, J., R. Wang, Y. Lu, and J. Huang. 2024. "Prediction of Compressive Strength of Geopolymer Concrete Landscape Design: Application of the Novel Hybrid RF–GWO–XGBoost Algorithm." *Buildings* 14 (3): 591. https://doi.org/10.3390/buildings14030591.