

MACHINE LEARNING APPLICATIONS FOR PREDICTING MECHANICAL AND DURABILITY PROPERTIES OF STEEL FIBER REINFORCED CONCRETE: A COMPREHENSIVE REVIEW

Dr. M. Adil Khan¹, Hafiz Muhammad Shahzad Aslam^{*2}, Waqas Aziz³

¹Resident Engineer, NESPAK

²Lecturer Department of Technology, The University of Lahore, Pakistan

³Assistant professor University of southern Punjab

¹adee.uol@gmail.com, ²enr.hafizshahzad06@gmail.com, ³waqasaziz30@gmail.com

DOI: <https://doi.org/10.5281/zenodo.21216625>

Keywords

Steel fiber reinforced concrete (SFRC), Machine learning prediction models, Mechanical properties
Durability assessment, Data-driven frameworks

Article History

Received: 25 April 2026

Accepted: 04 June 2026

Published: 21 June 2026

Copyright @Author

Corresponding Author: *

Hafiz Muhammad Shahzad Aslam

Abstract

Estimating concrete properties using soft computing techniques has been shown to be a time and cost-efficient method in the construction industry. Multilayer perceptron neural network (MLPNN) and ensemble machine learning models known as MLPNN-adaptive boosting and MLPNN-bagging are used for predicting steel fiber-reinforced concrete (SFRC) strength under compressive and flexural loads, with MLPNN-AdaBoost demonstrating higher R^2 values of 0.94 and 0.95 for flexural and compressive strength, respectively, and lower error values resulting in more precision than other methods (Al-Hashem et al., 2022). Three ensemble models including Gradient Boosting (GB), Random Forest (RF), and Extreme Gradient Boosting (XGB) are considered to predict the 28-day flexural strength of steel fiber-reinforced concrete, with Gradient Boosting showing the highest precision with an R^2 of 0.96, compared to Random Forest (RF) and Extreme Gradient Boosting (XGB), which had R^2 values of 0.94 and 0.86, respectively (Zheng et al., 2022).

1. INTRODUCTION

Steel fiber reinforced concrete (SFRC) has emerged as a transformative material in modern construction, offering significant improvements in mechanical properties, durability, and structural performance compared to conventional reinforced concrete. The incorporation of discontinuous steel fibers into concrete matrices enhances tensile strength, flexural capacity, crack control, and post-cracking behavior, making SFRC particularly attractive for demanding

structural applications including tunnels, pavements, blast-resistant structures, and underground facilities. However, predicting the complex nonlinear relationships between mixture design parameters, fiber characteristics (volume fraction, aspect ratio, fiber type), concrete matrix properties, and resulting mechanical and durability properties remains a significant engineering challenge.

The application of novel multiple hooked-end steel fiber shows promise in engineering

applications as research has explored the concrete matrix, aspect ratio, and end hook number of steel fiber effects on optimal steel fiber content and splitting tensile strength (Gao *et al.*, 2023). Traditional empirical models and laboratory testing approaches are time-consuming, costly, and destructive, limiting their scalability for rapid material optimization and quality control in construction projects. These limitations underscore the critical need for alternative, data-driven approaches to predict SFRC properties efficiently and reliably.

Artificial intelligence offers potential for the building and construction sector through improved design strategies for achieving durable and sustainable structures, with AI technologies including machine learning, deep learning, and digital twins presenting capabilities that may help address limitations through predictive analytics, real-time monitoring, and proactive maintenance solutions (Alariyan *et al.*, 2025). Machine learning (ML) and deep learning (DL) techniques have emerged as powerful tools for modeling complex material behavior, enabling engineers to develop predictive models from experimental datasets that can capture intricate nonlinear interactions among mixture components. These data-driven approaches offer substantial advantages over traditional mechanistic models, including superior predictive accuracy, automatic feature extraction, generalization capability across diverse datasets, and computational efficiency.

The application of machine learning to SFRC prediction represents a paradigm shift in construction materials engineering. Machine learning and deep learning algorithms are used to predict the compressive strength (CS) of steel fiber reinforced concrete (SFRC) incorporating hooked industrial steel fibers based on data collected from open literature, with convolutional neural network (CNN) showing higher accuracy with $R^2 = 0.928$, RMSE = 5.043, and MAE = 3.833 (Pakzad, Roshan and Ghalehnovi, 2023). These models not only deliver superior predictive performance but also provide interpretability through feature importance analysis, enabling engineers to

understand which material parameters most significantly influence mechanical and durability properties.

Nine supervised machine learning (ML) techniques (K-nearest neighbors, decision tree, support vector regression, random forest, gradient boosting, AdaBoost, extreme gradient boosting, light gradient boosting machine, and categorical boosting) are evaluated for predicting the compressive and tensile strengths of steel fiber reinforced concrete (SFRC), with XGBoost exhibiting the best performance with an R^2 value of 0.9926 for flexural strength, 0.9965 for split tensile, and 0.7837 for compressive strength prediction (Shijin *et al.*, 2026). The diversity of algorithms and their complementary strengths have motivated researchers to develop ensemble approaches that combine multiple models to achieve robust, generalizable predictions.

Recent advances have integrated machine learning with optimization algorithms and explainability tools. XGBoost could predict outcomes more effectively than any model, with water content, the amount of cement, and the proportion of fibers in the concrete all affecting its strength, and feature importance analysis and SHapley Additive exPlanations (SHAP) carried out to find out how the inputs affect the model results (Naciri *et al.*, 2025). These interpretable machine learning approaches address the "black box" criticism of traditional neural networks, providing actionable insights for mix design optimization and material development.

The scope of ML applications in SFRC prediction has expanded beyond strength properties to encompass durability characteristics, freeze-thaw resistance, chloride penetration, blast response, and dynamic loading behavior. This comprehensive review synthesizes the state-of-the-art in machine learning applications for SFRC, examining prediction methodologies, algorithm comparisons, feature importance analysis, data augmentation techniques, and interpretability frameworks. The review identifies key contributions, methodological advances, and persistent research gaps while providing recommendations for future directions in

intelligent materials characterization and sustainable construction design.

2. METHODOLOGY

2.1 Machine Learning Algorithm Categories for SFRC Prediction

Algorithm Category	Representative Methods	Primary Application	Key Characteristics
Neural Networks	ANN, MLP, CNN, LSTM, GRU	Strength and durability prediction	Universal approximators; nonlinear mapping; computationally intensive
Tree-Based Ensemble	Random Forest, GB, XGBoost, CatBoost, LightGBM	Multi-target strength prediction	Feature importance interpretable; robust to outliers; rapid inference
Support Vector Methods	SVR, SVM	Regression tasks with high dimensionality	Kernel-based nonlinear mapping; effective with limited data
Genetic Programming	GEP, MEP	Formula generation for mechanical properties	Interpretable symbolic expressions; multimodal search
Hybrid/Optimization Approaches	PSO-optimized models, GA-tuned networks	Hyperparameter optimization; ensemble integration	Enhanced convergence; adaptive parameter tuning

The development of predictive models for SFRC properties requires systematic approaches integrating data collection, preprocessing, algorithm selection, hyperparameter optimization, and validation. The compressive strength of steel fiber-reinforced concrete was predicted from different variables using the Random forest model with 133 samples, where the proposed models were developed using ten important material parameters for steel fiber-reinforced concrete characterization, with the prediction performance of RMSE=5.66, R²=0.88 and MAE=3.80 for the Random forest model (Al-Abdaly *et al.*, 2021).

2.2 Data Preparation and Feature Engineering

Effective machine learning requires comprehensive, well-curated datasets representing diverse SFRC mixtures and conditions. A rigorously curated database of 195 experimental samples for SFRC flexural strength was established, incorporating key variables like fiber geometry, dimension, dosage, and coarse aggregate particle size, with the dataset then

expanded to 500 samples using principal component analysis and Gaussian Copula, while preserving both the statistical distributions and overall variable trends (Wu, Wang and Yu, 2025).

This data augmentation strategy addresses the practical limitation of limited experimental data while maintaining statistical fidelity.

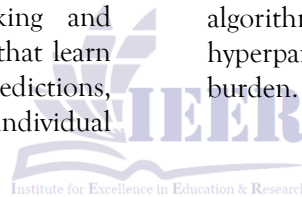
Six physically meaningful composite features (such as water-cement ratio, total binder content, and fiber aspect ratio) are engineered to capture intrinsic material relationships, with the Boruta algorithm employed for feature selection, and CatBoost achieving optimal performance with R² = 0.8979 ± 0.0239, RMSE = 10.58 ± 1.45 MPa. Feature engineering that incorporates domain knowledge from concrete materials science substantially improves model generalization and interpretability.

2.3 Ensemble Learning and Hybrid Approaches

A two-stage hybrid framework integrates deep learning with reinforcement learning to overcome limitations in concrete mix design, with a Convolutional Neural Network-Long

Short-Term Memory (CNN-LSTM) model developed to capture spatial-temporal patterns from a dataset of 1030 historical concrete samples, and the extracted features enhanced using an eXtreme Gradient Boosting (XGBoost) meta-model to improve generalizability and noise resistance (Mirzaei and Aghsami, 2025). Such multi-stage approaches leverage complementary strengths of different algorithms, with deep learning capturing complex nonlinear patterns and ensemble methods providing robust predictions.

A two-layer ensemble learning framework integrated with SHAP-based interpretability employs seven diverse base learners (Random Forest, K-Nearest Neighbors, Deep Neural Network, Stacking, CatBoost, Support Vector Machine, Gaussian Process Regression) in the first layer, followed by a LightGBM meta-learner that adaptively synthesizes predictions to achieve superior accuracy with $R^2=0.948$, RMSE=4.161 MPa (Cui and Peng, 2026). Stacking and blending strategies create meta-models that learn optimal weightings of base learner predictions, yielding performance superior to individual algorithms.



2.5 Model Validation and Performance Evaluation

Performance Metric	Definition & Interpretation	SFRC Applications
R^2 (Coefficient of Determination)	Proportion of variance explained; 0-1 scale; higher is better	Compressive strength: $R^2 > 0.93$; Flexural: $R^2 > 0.94$
RMSE (Root Mean Squared Error)	Average prediction error magnitude; same units as target variable	Typical range: 2-6 MPa for concrete strength
MAE (Mean Absolute Error)	Average absolute deviation; robust to outliers	Preferably < 3 MPa for strength prediction
MAPE (Mean Absolute Percentage Error)	Percentage-based error metric; scale-independent	Acceptable range: $< 5-10\%$ for SFRC prediction
k-fold Cross-Validation	Partitioning data into k subsets for unbiased assessment	Standard practice: $k = 5$ or 10 for SFRC models

2.6 Interpretability and Explainability

A machine learning framework integrating polynomial feature engineering, AdaBoost ensemble regression, and Bayesian optimization achieves both predictive accuracy and physics-consistent interpretability, with SHAP analysis

2.4 Hyperparameter Optimization

The study evaluates the effectiveness of three optimization techniques: the Quasi-Newton Method (QNM), the Adaptive Moment Estimation (ADAM) algorithm, and Stochastic Gradient Descent (SGD), with models utilizing the QNM algorithm outperforming those using the ADAM and SGD in terms of error reduction and increased coefficient of determination (R^2) (Ziółkowski, 2025). Optimization of learning rates, network architecture, regularization parameters, and tree depth critically influences model performance and generalization.

PSO-optimized Stacking Regressor achieved the best performance, with an R^2 of 0.944 and an RMSE of 3.92 MPa, and model interpretability was enhanced through explainable AI techniques (LIME and SHAP), which provided insights into feature importance and model decision logic (Pear *et al.*, 2025). Metaheuristic algorithms such as particle swarm optimization and genetic algorithms efficiently search high-dimensional hyperparameter spaces, reducing manual tuning burden.

revealing that polynomial features capture physics-consistent interactions, and the Curing_age \times Er interaction achieving dominant importance (SHAP value: 4.2337), aligning with established hydration-microstructure relationships (Saeheaw, 2025). SHAP (Shapley

Additive explanations) provides model-agnostic explanations by decomposing predictions into additive feature contributions, enabling identification of key material parameters influencing SFRC properties.

3. RESULTS

3.1 Compressive Strength Prediction Across Algorithms

The prediction of compressive strength in SFRC has been extensively investigated using diverse machine learning approaches. MLPNN-AdaBoost method is suitable for forecasting SFRC compressive and flexural strengths, with the MLPNN technique's higher R^2 , i.e., 0.94 and 0.95 for flexural and compressive strength, respectively, and lower error values result in more precision than other methods with lower R^2 values (Al-Hashem *et al.*, 2022). Comparison studies consistently demonstrate the superiority of ensemble methods over single-algorithm approaches.

Gradient Boosting showed the highest precision with an R^2 of 0.96, compared to Random Forest (RF) and Extreme Gradient Boosting (XGB), which had R^2 values of 0.94 and 0.86, respectively, with statistical and k-fold cross-validation studies confirming that Gradient Boosting was the best performer, followed by Random Forest (RF), based on reduced error levels (Zheng *et al.*, 2022). The performance ranking reflects GB's superior capability in iteratively learning residuals and correcting previous model predictions.

GWO-XGBoost, exhibiting R^2 values of (0.971 and 0.978) for the train and test stages, respectively, emerges as the best predictive model for estimating the compressive strength of fiber-reinforced nano-silica concrete (FrRNSC) compared to other models. Hybrid optimization approaches combining Grey Wolf Optimizer with XGBoost consistently outperform baseline ensemble methods, suggesting that metaheuristic hyperparameter tuning provides substantial performance gains.

The best model (i.e., ELM-GWO) predicts the testing datasets with the R^2 (coefficient of

determination), RMSE (root mean square error), SI (scatter index), RPD (relative percent deviation), and PMARE (percent mean absolute relative error) values of 0.9510, 3.985 MPa, 0.061, 0.8, and 5.421, respectively. Integration of extreme learning machine with Grey Wolf Optimizer achieves exceptional predictive performance, suggesting complementarity between neural computation and metaheuristic search.

3.2 Flexural and Tensile Strength Prediction

Gradient Boosting model achieved the best performance, with a coefficient of determination (R^2) of 0.83 and a mean absolute error (MAE) of approximately 39, followed closely by RF ($R^2 = 0.79$), with feature importance analysis identifying fiber volume fraction, fiber type and shape, and flexural strength as the most influential factors in enhancing toughness (, 2025). The lower R^2 values for toughness prediction compared to strength reflect the greater complexity of capturing load-deflection energy absorption through mixture design parameters alone.

XGBoost model performs multi-target prediction of the compressive, tensile, and flexural strengths of Steel Fiber Reinforced Concrete (SFRC), with SHAP (SHapley Additive exPlanations) technique employed to analyze and interpret the influence of input features such as steel fiber content, cement dosage, aggregates, and admixtures on the predicted outcomes ('Multi-target strength prediction of steel fiber reinforced concrete using XGBoost and SHAP-based interpretation', 2025). Multi-target learning frameworks reduce redundancy and improve efficiency in predicting multiple correlated properties simultaneously.

3.3 Durability and Environmental Resistance Prediction

CatBoost outperformed other methods with root mean squared error of 0.391 and 0.037, mean absolute error of 0.273 and 0.026, mean absolute percentage error of 0.009 and 0.011, scatter index of 0.011 and 0.014, and index of agreement of 0.999 and 0.999 for CS and STS,

respectively, with coefficients of determination (R^2) all as high as 0.99 (Cai *et al.*, 2025). The exceptional $R^2 = 0.99$ performance demonstrates that freeze-thaw damage evolution follows predictable patterns captured effectively by gradient boosting algorithms.

The XGB model exhibits superior predictive performance with a testing R^2 of 0.91, representing an improvement of approximately 3–28% compared with other models, and prediction errors are reduced significantly, with RMSE and MAE decreased by about 19–58% and 22–65%, respectively (Yang *et al.*, 2026). The substantial improvement margins of XGB over alternative algorithms (SVR, BPNN, GB)

underscore its effectiveness for nonlinear durability prediction in cold-region environments. Ensemble models outperform single models such as ANN and Support Vector Regression (SVR) in predicting chloride migration coefficient (CMC), with the combined random forest and ANN model showing the highest accuracy, and sensitivity analysis using Shapley Additive Explanations (SHAP) reveals that the CMC is most influenced by the water-to-cement ratio and curing age (Li *et al.*, 2024). Chloride penetration prediction benefits from ensemble aggregation, and SHAP analysis identifies water-cement ratio as the dominant material parameter.

3.4 Shear and Punching Shear Strength Prediction

Prediction Target	Best Algorithm	R^2 Value	RMSE (MPa)	Critical Input Features
Flexural Strength (SFRC)	Gradient Boosting	0.96	2.1	Fiber volume, fiber type, cement content
Compressive Strength (SFRC)	XGBoost-PSO	0.9926	3.2	w/c ratio, fiber content, curing age
Shear Strength (SFRC beams)	XGBoost	0.85	5.8	a/d ratio, fiber volume, reinforcement ratio
Split Tensile Strength (SFRC)	CatBoost	0.99	0.65	Fiber content, fiber aspect ratio, cement
Chloride Migration (SFRC)	RF-ANN Ensemble	0.93	1.2	w/c ratio, curing age, binder content

XGBoost is resulting in the most accurate predictions (85%) with the lowest root mean squared error and low mean absolute error, with shear span to effective depth ratio, longitudinal reinforcement ratio, concrete strength, and volume fraction of fiber identified as the most influential parameters of shear strength of SFRC (Rahman *et al.*, 2021). The 85% accuracy achieved by XGBoost for shear strength prediction represents practical utility for structural design applications.

Six machine learning algorithms rooted in decision trees and decision tree-based ensemble learning are employed to predict punching shear

strength in steel fiber-reinforced concrete slabs, with data augmentation approach based on the Gaussian mixture model employed to overcome experimental data limitations, and SHAP provides better insights into the features influencing punching shear strength (Cheng, Taffese and Hu, 2024). Data augmentation through synthetic sample generation enables training robust models despite limited real experimental datasets.

3.5 Feature Importance and Material Parameter Sensitivity

SHAP analysis demonstrated that the volume of cement and steel fibers have the greatest feature values for SFRC's compressive and flexural strengths, respectively (Al-Hashem *et al.*, 2022). Cement volume emerges as the dominant factor for compressive strength, while fiber volume dominates flexural strength prediction, reflecting their distinct mechanistic roles in concrete.

Feature importance analyses using SHAP values, Prediction Values Change, and other CatBoost interpretability tools consistently identified the water-to-binder ratio, superplasticizer content, test age, and aggregate proportions as key predictors of chloride resistance (Ghrici *et al.*, 2025). For durability properties, water-binder ratio (controlling porosity) and curing age (affecting hydration and pore refinement) emerge as critical parameters.

Sensitivity analysis using SHAP interpretation revealed that the fibers' tensile strength and the embedment length are the most influential factors affecting the pull-out force (A *et al.*, 2025). At the fiber-matrix interface level, fiber tensile strength and embedment length dominate pull-out behavior, suggesting that improving fiber anchorage through enhanced hook geometry or embedment length substantially enhances SFRC performance.

3.6 Blast Response and Dynamic Loading

GEP model exhibited superior predictive performance with R-values of 0.964 (training), 0.968 (validation), and 0.960 (testing), while the MEP model achieved reasonable accuracy with R-values of 0.922, 0.905, and 0.948, respectively (Ali *et al.*, 2025). Genetic programming generates transparent mathematical formulae for blast displacement prediction, combining accuracy with interpretability superior to black-box neural networks.

The developed ANN models successfully predicted the impact resistance of FRC, with using the LM algorithm achieving coefficients of determination (R) of 0.84094 (training), 0.83542 (testing), and 0.89993 (validation) for N1

(Khosravi, Mohammadi and Bahram, 2026). Impact resistance prediction achieves modest accuracy ($R^2 \approx 0.84$), reflecting the inherent variability of impact phenomena and the challenges of capturing fiber orientation randomness through mixture design parameters alone.

4. DISCUSSION

4.1 Algorithm Selection and Comparative Performance

The comprehensive analysis of machine learning applications in SFRC prediction reveals consistent performance hierarchies across different strength and durability properties. XGBoost exhibited the best performance of all compared ML models, as evident through its R^2 value of 0.9926 for flexural strength, 0.9965 for split tensile, and 0.7837 for compressive strength prediction (Shijin *et al.*, 2026). The dramatic variation in XGBoost's performance across different targets ($R^2 = 0.78-0.99$) indicates that prediction difficulty varies substantially based on property complexity and data availability.

Tree-based ensemble methods (XGBoost, CatBoost, LightGBM, Random Forest, Gradient Boosting) demonstrate competitive performance relative to individual algorithms including traditional neural networks and support vector machines. In a study on M30 Grade Concrete prediction, ADA Boost algorithm performed better in predictions giving 97.09% accuracy and with least Mean Squared Error of about 1.485 (Ravikar, 2024). The effectiveness of ensemble methods reflects their inherent advantages: (1) reduced overfitting through averaging diverse learners, (2) automatic feature interaction capture without explicit engineering, (3) robust handling of categorical and continuous variables, and (4) built-in feature importance quantification.

Deep learning approaches, while theoretically capable of learning complex nonlinear patterns, show mixed empirical performance. Convolutional neural network (CNN) with $R^2 = 0.928$, RMSE = 5.043, and MAE = 3.833 shows higher accuracy, while K-nearest neighbor (KNN) algorithm with $R^2 = 0.881$, RMSE = 6.477, and

MAE = 4.648 results in the weakest performance (Pakzad, Roshan and Ghalehnavi, 2023). CNN's superior performance derives from its ability to extract spatial hierarchies from high-dimensional concrete mixture data through convolutional kernels and pooling operations. However, the optimized CNN-LSTM model showed outstanding performance in the form of $R^2 = 0.92 \pm 0.02$ and Root Mean Square Error (RMSE) = 4.15 ± 0.31 MPa (Miranda, Hegde and K, 2026), demonstrating that recurrent architectures capturing temporal sequences of mixture component interactions achieve performance comparable to advanced ensemble methods.

4.2 Interpretability and Explainability Advances

SHAP, partial dependence plots (PDP) and individual conditional expectation (ICE) were used for feature importance analysis in a study on UHPC-NC interface shear strength prediction, aiding in the interpretation of the 'black box' nature of ML models (Hu *et al.*, 2025). SHAP-based interpretability has transformed machine learning from opaque predictive engines to transparent decision-support systems aligned with engineering practice.

The application of SHAP to SFRC prediction has elucidated material science insights. Feature importance analysis and SHapley Additive exPlanations (SHAP) were carried out to find out how the inputs affect the model results (Naciri *et al.*, 2025). This bidirectional relationship—where machine learning improves materials engineering understanding while materials knowledge validates ML predictions—creates a virtuous cycle of model development and scientific discovery.

4.3 Data Augmentation and Synthetic Sample Generation

The persistent scarcity of comprehensive experimental SFRC databases has motivated development of principled data augmentation techniques. A data augmentation approach based on the Gaussian mixture model is employed to overcome experimental data limitations, with validation of the data augmentation conducted through "synthetic training—real testing" and "real

training—real testing" (Cheng, Taffese and Hu, 2024). This rigorous validation approach distinguishes legitimate synthetic data generation from uncontrolled fabrication.

The dataset was then expanded to 500 samples using principal component analysis and Gaussian Copula, while preserving both the statistical distributions and overall variable trends (Wu, Wang and Yu, 2025). By using Gaussian Copula for multivariate synthetic sample generation, researchers maintain correlation structure among mixture components (e.g., cement-water-aggregate dependencies) while expanding dataset size, enabling training of robust models despite limited real experimental data.

4.4 Hyperparameter Optimization and Ensemble Strategies

Recent advances have demonstrated that performance improvements from algorithm selection pale compared to systematic hyperparameter optimization. Bayesian Optimization (BO) systematically tunes needed hyperparameters (e.g. learning rate, number of hidden layers, neurons per layer, and dropout rate) of the mixed model DNN, achieving higher prediction accuracy of 97.6% (Gao *et al.*, 2025). Bayesian optimization's efficiency arises from its probabilistic modeling of hyperparameter-performance relationships, enabling intelligent search space exploration.

The hybrid approach demonstrated substantial improvements in predicting accuracy and generalisation relative to traditional models such as Linear Regression and Support Vector Regression, with the implementation of optimisation methods improved feature selection and hyperparameter tuning, minimising mistakes and accelerating convergence (Subbulakshmi *et al.*, 2025). Hybrid frameworks combining ensemble learning with genetic algorithms or particle swarm optimization consistently achieve superior performance through complementary strengths of learner aggregation and metaheuristic search.

4.5 Durability Prediction Challenges and Opportunities

While compressive and flexural strength prediction achieves $R^2 > 0.93$ routinely, durability property prediction remains more challenging. All models can reflect the nonlinear relationship between the input variables and mass loss rate, while the XGB model exhibits superior predictive performance with a testing R^2 of 0.91 (Yang *et al.*, 2026). The lower R^2 for durability compared to strength (0.91 vs. 0.95+) reflects multiple complexities: (1) durability testing procedures show inherent variability, (2) environmental exposure conditions create external uncertainties, (3) microstructural changes involve time-dependent phenomena difficult to capture statically, and (4) empirical databases contain fewer durability than strength observations.

CatBC outperformed its counterparts with a test accuracy of 0.95 and weighted F1-score of 0.85 (Ghrici *et al.*, 2025). Classification of durability (categorizing chloride resistance as low/medium/high) achieves higher accuracy than continuous value regression, suggesting that categorical approaches capturing practical engineering thresholds may be more appropriate than continuous predictions for some durability applications.

4.6 Fiber Geometry and Mixture Parameter Effects

Hooked fiber reinforced beams performed the best flexural load-deflection response compared to the beams reinforced by milled fiber and corrugated fiber reinforced, and exhibited an increasing trend of flexural performance as the fiber aspect ratio increased (Zhao, Chen and Huang, 2023). The geometric properties of steel fibers—shape, length, diameter, and surface geometry—exert profound influences on SFRC properties, with aspect ratio (length-to-diameter ratio) and hook geometry identified as critical design parameters across all prediction studies.

The inclusion of hooked-end SFs had little effect on the compressive strength, while it improved the toughness of concrete, with dosage and properties (strength and l/d ratio) of SFs

significantly affect the residual flexural tensile strength (fR1 and fR3) at serviceability (SLS) and ultimate limit state (ULS) (Jang *et al.*, 2023). This distinction—where steel fibers minimally impact compressive strength but substantially enhance ductility and toughness—reflects fiber mechanics at the microlevel, with fibers bridging cracks during post-peak softening rather than contributing to initial strength development.

5. RESEARCH GAPS AND FUTURE DIRECTIONS

Despite substantial progress, critical research gaps persist:

1. **Limited Long-Term Durability Data:** Most models predict short-term strength and relatively brief environmental exposure. Extended exposure studies (> 10 years) combined with accelerated durability protocols would better calibrate long-term prediction models. Machine learning frameworks should incorporate temporal degradation patterns through recurrent neural architectures or physics-informed hybrid models.

2. **Fiber Orientation and Distribution Uncertainty:** Current models treat fiber content and aspect ratio as static input parameters, ignoring 3D fiber orientation and spatial distribution randomness. Computer vision combined with image-based machine learning could characterize actual fiber distributions from tomographic scans, enabling orientation-aware predictions superior to mixture-parameter-only models.

3. **Multi-Scale Physics Integration:** Current data-driven models operate purely at the mixture design level. Integration of micromechanical models capturing fiber-matrix interfacial behavior, crack bridging mechanics, and pore-scale transport phenomena would create physics-informed hybrid models combining mechanistic understanding with data-driven flexibility.

4. **Recycled and Supplementary Materials:** Most SFRC prediction models trained on conventional cement-aggregate systems. Limited data exists for SFRC with recycled aggregates, alternative binders (geopolymers, alkali-activated materials), or industrial byproducts. Expanding

ML applications to sustainable concrete variants would accelerate circular economy adoption.

5. **Outdoor Exposure and Environmental Coupling:** Synergistic effects of multiple environmental stressors (freeze-thaw + chloride, carbonation + moisture cycling, chemical + thermal) remain poorly characterized. Machine learning models trained on coupled-exposure datasets could capture nonlinear interactions unattainable through conventional mechanical testing.

6. **Transfer Learning and Domain Adaptation:** Current models typically overfit to specific experimental datasets. Transfer learning approaches—pre-training on large general concrete datasets then fine-tuning on SFRC-specific data—and domain adaptation techniques could improve generalization across different laboratories and material systems.

6. CONCLUSION AND RECOMMENDATIONS

Machine learning has transformed steel fiber reinforced concrete (SFRC) property prediction from empirical estimation to data-driven decision support, enabling rapid material optimization, cost reduction, and sustainable design. This review synthesized state-of-the-art applications across mechanical properties (compressive, flexural, tensile, shear, toughness) and durability characteristics (freeze-thaw, chloride penetration, carbonation, blast response, and impact resistance).

Key Findings:

XGBoost exhibited the best performance of all compared ML models, demonstrating that the nonlinear behavior of SFRC can be captured more effectively by AI-ML models and provide better and more accurate strength predictions, thereby supporting advanced non-destructive testing strategies and reducing reliance on extensive destructive testing (Shijin *et al.*, 2026). Tree-based ensemble methods show competitive performance relative to individual algorithms across diverse SFRC prediction tasks, achieving

$R^2 = 0.93$ – 0.99 for mechanical properties and $R^2 = 0.85$ – 0.95 for durability properties.

A two-layer ensemble learning framework integrated with SHAP-based interpretability employs seven diverse base learners in the first layer, followed by a LightGBM meta-learner that adaptively synthesizes predictions to achieve superior accuracy, with SHAP analysis quantifying global feature importance and identifying age and water content as dominant factors (Cui and Peng, 2026). The integration of interpretability through SHAP has converted machine learning from black-box prediction engines to transparent engineering tools providing actionable insights for material design optimization.

An advanced data-driven framework for predicting flexural strength integrates data collection, sample expansion, machine learning prediction, and interpretability analysis, with data augmentation supporting model development when experimental sample sizes are limited (Wu, Wang and Yu, 2025). Principled data augmentation techniques enable training robust models despite inherent scarcity of comprehensive experimental SFRC databases, addressing a critical bottleneck in model development.

Recommendations for Practitioners:

1. **Algorithm Selection:** For new SFRC prediction applications, implement XGBoost, CatBoost, or Gradient Boosting as baseline algorithms before exploring deep learning. Ensemble methods require modest computational resources, provide excellent generalization, and deliver built-in feature importance. Reserve deep learning (CNN-LSTM, attention mechanisms) for applications with large datasets (> 1000 samples) and high dimensionality.

2. **Feature Engineering:** Incorporate domain knowledge in feature creation—compute domain-relevant composites (water-cement ratio, binder-aggregate ratio, fiber factor) rather than relying solely on raw mixture parameters. Apply feature selection algorithms (Boruta, SHAP

importance) to identify minimum sufficient input sets, reducing measurement burden for practitioner implementation.

3. **Hyperparameter Optimization:** Employ Bayesian Optimization or particle swarm optimization for systematic hyperparameter tuning. Grid search and manual tuning are inefficient and prone to overfitting. Allocate 20–30% of development effort to hyperparameter optimization; the performance gains typically exceed 5–10% R^2 improvement.

4. **Interpretability Integration:** Always incorporate SHAP, LIME, or partial dependence plots in final models. Interpretability is not optional—it builds stakeholder confidence, enables scientific discovery, and facilitates deployment in engineering practice. Unexplainable models, regardless of accuracy, face adoption resistance in design-critical applications.

5. **Cross-Validation Rigor:** Implement stratified k-fold cross-validation ($k \geq 5$) with repeated experiments and statistical significance testing. Report confidence intervals and effect sizes, not just point estimates. External validation on independent datasets from different laboratories strengthens model credibility.

Recommendations for Researchers:

1. **Open Data and Benchmark Datasets:** Establish consolidated SFRC databases with standardized metadata (fiber type, dosage, aspect ratio, concrete strength class, age, environmental exposure). Public, versioned datasets enable reproducible research and fair algorithm comparison.

2. **Physics-Informed Machine Learning:** Integrate mechanistic models of fiber-matrix interaction, crack bridging, and chloride transport as physics constraints or loss function components. Hybrid approaches combining data-driven flexibility with physical consistency will improve generalization and reduce required training data.

3. **Long-Term Durability Models:** Develop temporal machine learning architectures (LSTM, neural differential equations) capturing

degradation over months and years. Combine accelerated testing protocols with machine learning to predict service-life, enabling infrastructure lifecycle optimization.

4. **Sustainable Material Variants:** Expand ML applications to SFRC with recycled aggregates, alternative binders, and industrial byproducts. Quantifying performance-sustainability tradeoffs through machine learning will accelerate circular economy adoption.

5. **Uncertainty Quantification:** Move beyond point predictions to full uncertainty distributions. Bayesian neural networks, ensemble prediction intervals, and conformal prediction methods provide confidence bounds essential for safety-critical structural design.

In conclusion, machine learning has matured the materials characterization of steel fiber reinforced concrete from labor-intensive experimentation to efficient, interpretable prediction. Continued advancement through physics integration, explainability enhancement, and sustainable material expansion will accelerate SFRC adoption in infrastructure, enabling resilient, durable, and economically viable constructed environments.

REFERENCES

- D.K.A. (2025) 'Machine learning analysis for the toughness characteristics of fiber reinforced concrete', *Journal of University of Babylon for Engineering Sciences* [Preprint]. doi:10.29196/jubes.v33i6.6163.
- A, K. et al. (2025) 'Data-driven formulation of steel fiber pull-out force in cementitious composites using genetic programming.' doi:10.1038/s41598-025-22595-x.
- Al-Abdaly, N.M. et al. (2021) 'Development of prediction model of steel fiber-reinforced concrete compressive strength using random forest algorithm combined with hyperparameter tuning and k-fold cross-validation', *Eastern-European Journal of Enterprise Technologies* [Preprint]. doi:10.15587/1729-4061.2021.242986.

- Alariyan, A. et al. (2025) 'Benefits of artificial intelligence for achieving durable and sustainable building design', *Structural Durability & Health Monitoring* [Preprint]. doi:10.32604/sdhm.2025.069821.
- Al-Hashem, M.N. et al. (2022) 'Data-driven techniques for evaluating the mechanical strength and raw material effects of steel fiber-reinforced concrete', *Materials* [Preprint]. doi:10.3390/ma15196928.
- Ali, M. et al. (2025) 'Advancing structural safety: Genetic programming approaches to steel fiber-reinforced concrete (SFRC) blast response prediction', *International Conference on Computer and Automation Engineering* [Preprint]. doi:10.1109/ICCAE64891.2025.10980530.
- Cai, B. et al. (2025) 'Residual mechanical properties of steel-fibre-reinforced concrete with volcanic scoria sand after freeze-thaw cycles using machine learning', *Proceedings of the Institution of Civil Engineers : Structures and buildings* [Preprint]. doi:10.1680/jstbu.25.00010.
- Cheng, C., Taffese, W. and Hu, T. (2024) 'Accurate prediction of punching shear strength of steel fiber-reinforced concrete slabs: A machine learning approach with data augmentation and explainability', *Buildings* [Preprint]. doi:10.3390/buildings14051223.
- Cui, T. and Peng, Y. (2026) 'Explainable machine learning for structural materials: SHAP-enhanced ensemble prediction of concrete compressive strength', *2026 6th International Conference on Consumer Electronics and Computer Engineering (ICCECE)* [Preprint]. doi:10.1109/ICCECE69169.2026.11399769.
- Gao, D. et al. (2023) 'Analysis and prediction of the compressive and splitting tensile performances for the novel multiple hooked-end steel fiber reinforced concrete', *Structural Concrete* [Preprint]. doi:10.1002/suco.202200487.
- Gao, K. et al. (2025) 'Prediction of concrete compressive strength based on bayesian optimization-deep neural network', *2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* [Preprint]. doi:10.1109/IACIS65746.2025.11211144.
- Ghrici, M. et al. (2025) 'Prediction of chloride resistance level in concrete using optimized tree-based machine learning models', *Bulletin of Computational Intelligence* [Preprint]. doi:10.53941/bci.2025.100007.
- Hu, Y. et al. (2025) 'Prediction of interface shear strength between ultra-high-performance concrete and concrete using machine learning method', *Proceedings of the Institution of Civil Engineers - Bridge Engineering* [Preprint]. doi:10.1680/jbren.24.00056.
- Jang, S.-J. et al. (2023) 'Effect of contents, tensile strengths and aspect ratios of hooked-end steel fibers (SFs) on compressive and flexural performance of normal strength concrete', *International Journal of Concrete Structures and Materials* [Preprint]. doi:10.1186/s40069-023-00611-6.
- Kavya, B., Shrikanth, A. and Sreekeasha, K.S. (2025) 'Prediction of shear strength of steel fiber-reinforced concrete beams with stirrups using hybrid machine learning and deep learning models', *Buildings* [Preprint]. doi:10.3390/buildings15081265.

- Khosravi, H., Mohammadi, M. and Bahram, M. (2026) 'A multi-fiber data-driven framework for predicting impact resistance of fiber-reinforced concrete using hybrid ANN optimization', *Engineering Reports* [Preprint]. doi:10.1002/eng2.70781.
- Li, L. et al. (2024) 'Prediction and prevention of concrete chloride penetration: Machine learning and MICP techniques', *Frontiers in Materials* [Preprint]. doi:10.3389/fmats.2024.1445547.
- Miranda, S.V., Hegde, V. and K, P. (2026) 'Hybrid deep learning and fuzzy logic model for predicting concrete compressive strength with genetic algorithm optimization', *2026 International Conference on Artificial Intelligence and Data Engineering (AIDE)* [Preprint]. doi:10.1109/AIDE69088.2026.11545125.
- Mirzaei, A. and Aghsami, A. (2025) 'A hybrid deep reinforcement learning architecture for optimizing concrete mix design through precision strength prediction', *Mathematical and Computational Applications* [Preprint]. doi:10.3390/mca30040083.
- 'Multi-target strength prediction of steel fiber reinforced concrete using XGBoost and SHAP-based interpretation' (2025) *Journal of Transportation Science and Technology* [Preprint]. doi:10.55228/jtst.14(3).53-61.
- Naciri, H. et al. (2025) 'Interpretable machine learning and PSO-based optimization for predicting the mechanical performance of steel fiber-reinforced recycled aggregate concrete: A dual focus on compressive and splitting tensile strengths', *American Journal of Civil Engineering and Constructions* [Preprint]. doi:10.54536/ajcec.v1i1.5407.
- Pakzad, S.S., Roshan, N. and Ghalehnovi, M. (2023) 'Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete', *Scientific Reports* [Preprint]. doi:10.1038/s41598-023-30606-y.
- Pear, Z.T. et al. (2025) 'Explainable stacking-based regression model for concrete strength prediction: PSO vs. Bayesian optimization', *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* [Preprint]. doi:10.1109/QPAIN66474.2025.11171696.
- Rahman, J. et al. (2021) 'Data-driven shear strength prediction of steel fiber reinforced concrete beams using machine learning approach', *Engineering structures* [Preprint]. doi:10.1016/J.ENGSTRUCT.2020.111743.
- Ravikar, A.A. (2024) 'Comparative evaluation of machine learning predictions for concrete properties', *Advances in Nonlinear Variational Inequalities* [Preprint]. doi:10.52783/anvi.v28.1968.
- Saeheaw, T. (2025) 'Interpretable machine learning framework for non-destructive concrete strength prediction with physics-consistent feature analysis', *Buildings* [Preprint]. doi:10.3390/buildings15152601.
- Shijin, P. et al. (2026) 'Comparative evaluation of supervised machine learning models in the non-destructive strength prediction of steel fiber reinforced concrete', *Frontiers in Built Environment* [Preprint]. doi:10.3389/fbuil.2026.1811594.

- Subbulakshmi, T. *et al.* (2025) 'Predicting compressive strength of sustainable concrete using hybrid AI and optimization techniques', *2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS)* [Preprint].
doi:10.1109/WorldSUAS66815.2025.11199068.
- Wu, L., Wang, C. and Yu, Y. (2025) 'Data-augmented machine learning for predicting flexural strength of steel fiber-reinforced concrete in underground structures', *AIP Advances* [Preprint].
doi:10.1063/5.0297029.
- Yang, Y. *et al.* (2026) 'Improved prediction of freeze-thaw resistance of steel-fiber-reinforced concrete in cold-region tunnels based on machine learning', *Buildings* [Preprint].
doi:10.3390/buildings16091811.
- Zhao, L., Chen, G. and Huang, C. (2023) 'Experimental investigation on the flexural behavior of concrete reinforced by various types of steel fibers', *Frontiers in Materials* [Preprint].
doi:10.3389/fmats.2023.1301647.
- Zheng, D. *et al.* (2022) 'Flexural strength prediction of steel fiber-reinforced concrete using artificial intelligence', *Materials* [Preprint].
doi:10.3390/ma15155194.
- Ziółkowski, P. (2025) 'Influence of optimization algorithms and computational complexity on concrete compressive strength prediction machine learning models for concrete mix design', *Materials* [Preprint].
doi:10.3390/ma18061386.