Engineering and Technology Journal for Research and Innovation (ETJRI)
ISSN 2581-8678, Vol. VII, Issue II, July 2025

Application of Machine Learning for the Prediction of Concrete Compressive Strength: A Comprehensive Review

Sachin Tiwari[1], Saurabh Dubey[2*], and Imran Ahmad Khan[3]
[1]*Department of Civil Engineering, Amity University, Gwalior, stiwari@gwa.amity.edu*
[2*]*Department of Civil Engineering, Amity University, Gwalior, sdubey3@gwa.amity.edu*
[3]*Department of Civil Engineering, Amity University, Gwalior, iakhan@gwa.amity.edu*

***Abstract*— The prediction of concrete compressive strength is a fundamental task in civil engineering, traditionally relying on empirical relationships and extensive laboratory experimentation. In recent years, Machine Learning (ML) has emerged as a powerful data-driven approach for accurately predicting concrete properties by capturing complex nonlinear interactions among mix design parameters. This review presents a comprehensive overview of recent advancements in ML-based prediction of concrete compressive strength. The study covers conventional concrete mixes as well as sustainable alternatives, including Recycled Aggregate Concrete (RAC), Geopolymer Concrete (GPC), and High-Performance Concrete (HPC). Various ML algorithms reported in the literature are critically analyzed, including standalone models such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Random Forest (RF), ensemble learning techniques such as Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM), and deep learning approaches including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. In addition, optimization techniques such as the Grey Wolf Optimizer (GWO), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) are discussed for improving model performance. The review highlights that ensemble and hybrid models generally provide superior predictive accuracy, often achieving coefficient of determination (R^2) values greater than 0.95. Finally, key challenges such as data heterogeneity, limited datasets, and model interpretability are identified, and future research directions are proposed to enhance the reliability and practical applicability of ML-driven concrete mix design.**

***Keywords*— Concrete Compressive Strength, Deep Learning, Ensemble Learning, Geopolymer Concrete, Machine Learning, Metaheuristic Optimization, Recycled Aggregate Concrete, Sustainable Construction.**

I. INTRODUCTION

Compressive strength is the most critical parameter in the structural design and quality control of concrete. Traditionally, engineers have relied on empirical equations derived from regression analysis or conducted physical tests, which are often labor-intensive, costly, and time-consuming [1,8]. The complexity of concrete, characterized by the non-linear interactions between its constituents—cement, water, aggregates, and admixtures—poses a significant challenge for traditional modeling [5,10].

The advent of artificial intelligence and machine learning (ML) has provided a more robust alternative for predicting these mechanical properties. ML models can process large, heterogeneous datasets to identify patterns that are not easily captured by conventional formulas [2-3]. These techniques have been applied across various domains of concrete technology, from optimizing standard mixes to predicting the strength of sustainable alternatives like recycled aggregate concrete (RAC) and geopolymer composites [6,11,21].

This review synthesizes findings from recent scholarly works to provide a detailed overview of the ML landscape in concrete strength prediction. We explore the transition from simple linear models to complex ensemble and hybrid frameworks, evaluating their performance across different concrete types and identifying the most influential factors driving strength development [7,13,24].

II. OBJECTIVE OF THE REVIEW

The objective of this review paper is to critically analyze and summarize the recent developments in machine learning techniques used for predicting the compressive strength of concrete. The study aims to review various machine learning and deep learning models applied in the literature, including traditional algorithms such as Artificial Neural Networks, Support Vector Machines, Random

---

* Corresponding Author: - sdubey3@gwa.amity.edu

Forest, and advanced ensemble learning approaches. In addition, this review highlights the application of machine learning in sustainable concrete materials, particularly recycled aggregate concrete and geopolymer concrete. The paper also evaluates commonly used performance metrics, optimization techniques, and current research challenges, and identifies potential future research directions for improving prediction accuracy and promoting sustainable construction practices.

III. METHODOLOGY OF LITERATURE REVIEW

A systematic literature review was conducted to identify relevant studies on the application of machine learning techniques in predicting concrete compressive strength. Scientific databases such as Scopus, Web of Science, ScienceDirect, and Google Scholar were used to collect research articles.

Keywords used during the search included “concrete compressive strength prediction,” “machine learning in concrete,” “deep learning for concrete strength,” “recycled aggregate concrete prediction,” and “geopolymer concrete machine learning.”

Only peer-reviewed journal articles and conference papers published between 2015 and 2025 were considered. Studies focusing on traditional experimental methods without ML models were excluded. The selected literature was analyzed based on the type of ML model, dataset size, input parameters, and performance evaluation metrics.

IV. MACHINE LEARNING ALGORITHMS AND OPTIMIZATION TECHNIQUES

Standalone Machine Learning Models: Standalone algorithms serve as the foundational building blocks for predictive modeling in concrete strength. These include:

- *Artificial Neural Networks (ANN)*: Inspired by biological neural systems, ANNs are widely used for their ability to model complex non-linear relationships [1, 8]. Studies have shown that ANNs, such as Back Propagation Neural Networks (BPNN), can achieve high accuracy ($R^2 > 0.95$) when properly tuned [4,14].
- *Support Vector Machines (SVM/SVR)*: SVMs are effective in handling high-dimensional data and non-linearities through kernel functions [5-6]. They are often competitive with neural networks in smaller datasets [10].
- *Random Forest (RF)*: A tree-based ensemble method that uses bagging to improve stability and accuracy. RF is frequently used as a benchmark due to its robustness against overfitting [10,15].
- *Decision Trees (DT) and K-Nearest Neighbors (KNN)*: These models are simpler and more interpretable but often yield lower accuracy compared to complex ensembles [1,11].
- *Gene Expression Programming (GEP)*: An evolutionary algorithm that creates mathematical expressions to model strength, providing better interpretability than "black-box" models like ANNs [20,27].

Ensemble and Boosting Methods: Ensemble methods combine multiple base learners to produce a more accurate and stable model. These have consistently been reported as top performers:

- *XGBoost (eXtreme Gradient Boosting):* Known for its efficiency and high accuracy, XGBoost frequently outperforms other models in strength prediction tasks, achieving R^2 values as high as 0.96 [1,12].
- *LightGBM and CatBoost:* Gradient boosting frameworks that focus on speed and handling categorical features, respectively. They have shown superior performance in large-scale dataset analyses [12,22].
- *AdaBoost:* An adaptive boosting technique that focuses on training subsequent models on the errors of previous ones. It has been highlighted for its suitability in predicting strength with lower error rates [2,5].
- *Stacking and Bagging:* Stacking involves using a meta-learner (e.g., RF or XGBoost) to aggregate predictions from diverse base models, often resulting in significant improvements in generalization [3,13,15,24].

Deep Learning Architectures: Deep learning (DL) techniques, particularly those involving multiple hidden layers, are increasingly applied to civil engineering problems:

- *Convolutional Neural Networks (CNN):* While typically used for image processing, CNNs have been adapted to predict concrete strength from tabular data, sometimes optimized by genetic algorithms to determine the best architecture [19].
- *Recurrent Neural Networks (RNN) and LSTM:* These models are effective for sequential data and have been used to model the time-dependent development of concrete strength [18].
- *Keras Neural Networks:* Advanced neural network frameworks are used to build custom architectures that can achieve near-perfect accuracy on well-curated experimental datasets [18].

Optimization and Metaheuristic Techniques: The performance of ML models is highly dependent on their hyperparameters. Recent research emphasizes the use of metaheuristic algorithms for tuning:

- *Grey Wolf Optimizer (GWO):* Used to optimize hyperparameters for RF and XGBoost, leading to improved predictive fidelity [24,26].
- *Genetic Algorithms (GA):* Frequently applied to optimize CNN architectures and ANFIS (Adaptive Neuro-Fuzzy Inference System) models [19, 23].
- *Sine Cosine Algorithm (SCA) and Harris Hawks Optimization (HHO):* These nature-inspired optimizers have been successfully paired with RF models to predict geopolymer strength with high reliability [21].
- *Imperialist Competitive Algorithm (ICA) and Firefly Algorithm (FFA):* These techniques have been used to optimize LightGBM and ANFIS models, significantly outperforming standalone versions [22-23].
- *Bayesian Optimization:* Techniques such as the Tree-structured Parzen Estimator (TPE) and Gaussian Process Regression are used for efficient hyperparameter search, reducing training time while maximizing accuracy [11,18].

V. APPROACHES IN THE SOURCES: SPECIALIZED CONCRETE TYPES

Recycled Aggregate Concrete (RAC): Predicting the strength of RAC is more challenging than standard concrete due to the variability of recycled coarse aggregates (RCA) and the presence of old mortar.

- *Models:* XGBoost, LightGBM, and stacked ensembles are the primary choices for RAC strength prediction [11-13].
- *Performance:* Models consistently achieve high R^2 values (0.93–0.97). For instance, a DE-XGBoost model (Differential Evolution-XGBoost) improved prediction accuracy by 1.12% and reduced RMSE by 16.09% compared to standalone XGBoost [12].
- *Influential Factors:* Curing age, cement content, and the replacement ratio of RCA are critical. Studies report an inverse relationship between RCA content and compressive strength [18].

Geopolymer and Eco-Friendly Concrete: Geopolymer concrete, an eco-friendly alternative to Portland cement, involves chemical reactions between aluminosilicate precursors and alkaline activators.

- *Models:* Hybrid models like ICA-LightGBM, SCA-RF, and FFA-ANFIS are preferred for capturing the complex chemical interactions [21-23].
- *Performance:* Hybrid models often achieve R^2 values exceeding 0.98. For example, the FFA-ANFIS model achieved an RMSE of 1.0322 and R^2 of 0.994 [23].
- *Influential Factors:* NaOH molarity, precursor proportions (Fly Ash/GGBS), and curing temperature are the most significant predictors [21,25].

High-Performance and Confined Concrete

- *HPC:* High-performance concrete requires precise modeling of multiple additives. GA-optimized CNNs have been shown to outperform traditional methods on large HPC datasets (1030 records) [19].
- *Confined Concrete:* For concrete filled in steel tubes or confined with FRP, ML models like ANN and GEP provide accurate structural strength predictions ($R^2 > 0.99$) [7,20].

VI. FINDINGS AND COMPARATIVE ANALYSIS

Model Performance Metrics: Researchers typically use a combination of statistical metrics to evaluate models:

- *Coefficient of Determination (R^2):* The most common metric, with top models ranging from 0.85 to 0.99 [1,4,6,21].
- *Root Mean Square Error (RMSE) and Mean Absolute Error (MAE):* Used to quantify the magnitude of error. In geopolymer studies, RMSE values as low as 1.03 MPa have been reported [23], while standard concrete models typically see RMSE in the range of 3.36–3.54 MPa [1,10].
- *Other Metrics:* Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Variance Accounted For (VAF) are also frequently cited to ensure comprehensive evaluation [12,22].

Feature Importance and Sensitivity Analysis: Feature importance techniques, such as SHAP (SHapley Additive exPlanations) and sensitivity analysis, are vital for model transparency:

- *Primary Factors:* Water-cement ratio and curing age are universally recognized as the most critical features across all concrete types [3,12,17].
- *Mix Constituents:* Cement content and superplasticizer dosage are strongly positively correlated with strength [1].
- Sustainable Materials: In RAC, the RCA replacement ratio is a key negative influencer [18]. In Geopolymer concrete, NaOH molarity is often the most important parameter [25].

Comparative Performance of Hybrid vs. Standalone Models: A consistent finding across the literature is the superiority of hybrid and optimized models:

- *Optimized Ensembles:* Hybridizing metaheuristics with ensemble models (e.g., GWO-RF or DE-XGBoost) consistently yields lower error rates and better generalization than default settings [12,24,26].
- *Ensemble vs. Individual:* XGBoost and Stacking models generally outperform standalone ANN or SVM models by integrating the strengths of multiple learners [1,12,15].
- *Neural Network Performance:* While ANNs are powerful, they are sensitive to architecture choices; thus, GA-CNN or optimized BPNNs are preferred for high-fidelity tasks [14,19].

VII. LIMITATIONS OF THE EVIDENCE

Despite the success of ML in this field, several limitations persist:

- *Data Heterogeneity and Covariate Shift:* Models trained on one dataset often perform poorly on another due to differences in material properties and testing conditions [10].
- *Small Dataset Sizes:* Many specialized studies rely on small experimental datasets (e.g., <200 samples), which increases the risk of overfitting [6,9,14].
- *Reporting Inconsistencies:* There is a lack of standardization in the reporting of input features and evaluation metrics, making cross-study comparisons difficult [4,6].
- *Synthetic Data Utility:* While synthetic data can augment training, its performance may degrade if the data generation process does not perfectly capture real-world physics [9].

VIII. CHALLENGES AND RESEARCH GAPS

Despite the progress made in machine learning applications for concrete strength prediction, several challenges remain:

- Limited availability of large and high-quality datasets
- Lack of standardized datasets across studies Overfitting issues in complex ML models
- Limited application of explainable AI techniques in concrete engineering

IX. FUTURE DIRECTIONS AND RESEARCH GAPS

- *Explainable AI (XAI):* There is a growing need to move beyond "black-box" models. Using SHAP and partial dependence plots can provide physical insights into how ML models make predictions [3,12].
- *Large-Scale Multi-National Datasets:* Combining datasets from different regions and laboratories can improve model generalization and robustness [10].

- *Uncertainty Quantification:* Implementing prediction intervals and outlier detection (e.g., Mahalanobis distance) is essential for engineering applications to assess confidence in a prediction [10].
- *Standardized Feature Engineering:* Developing a standardized list of input features, including chemical properties of materials, would facilitate more reliable model development [3].
- *Optimization for Multi-Objective Mix Design:* Future work should focus not only on strength but also on optimizing for cost, CO2 emissions, and workability simultaneously [15].

X. DISCUSSION

The reviewed literature clearly demonstrates that machine learning has significantly enhanced the ability to predict concrete compressive strength by capturing the complex nonlinear relationships between mix constituents, curing conditions, and mechanical properties. Traditional empirical models and regression-based approaches often struggle to represent these nonlinear interactions, whereas ML models can effectively learn such patterns from experimental datasets.

Among the various algorithms explored in the literature, ensemble learning methods such as XGBoost, Random Forest, LightGBM, and CatBoost consistently outperform standalone machine learning models. These algorithms benefit from combining multiple decision trees and reducing overfitting through boosting or bagging strategies, resulting in higher prediction accuracy and stability. Several studies report coefficient of determination (R^2) values exceeding 0.95, indicating strong predictive capability for both conventional and sustainable concrete mixtures.

Deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks have also demonstrated promising performance, particularly in modeling complex relationships and time-dependent strength development. However, these approaches generally require larger datasets and greater computational resources, which can limit their practical adoption in many engineering applications.

Another key observation is the increasing use of hybrid models combining machine learning with metaheuristic optimization techniques, such as Genetic Algorithms (GA), Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), and Differential Evolution (DE). These optimization strategies help identify optimal hyperparameters for ML models, thereby improving prediction accuracy and reducing training errors. Hybrid approaches such as DE-XGBoost and GWO-RF have been shown to significantly outperform their standalone counterparts.

Despite these advancements, several challenges remain. One major issue is data heterogeneity and limited dataset availability. Many studies rely on small experimental datasets or publicly available datasets such as the UCI concrete dataset, which may not fully represent real-world variability in materials and environmental conditions. This limitation may lead to reduced generalization when models are applied to field conditions.

Another challenge is the lack of standardized validation protocols. Different studies employ varying evaluation metrics, training/testing splits, and feature sets, making it difficult to directly compare model performance across studies. Moreover, many high-performing ML models operate as black-box systems, which limits their interpretability and acceptance among civil engineers.

Recent research has started addressing this issue through Explainable Artificial Intelligence (XAI) techniques such as SHAP and sensitivity analysis. These methods help identify influential parameters such as the water–cement ratio, curing age, cement content, and supplementary cementitious materials, providing valuable insights for mix design optimization.

Overall, while machine learning models demonstrate strong predictive capabilities, their widespread adoption in the construction industry will depend on improvements in dataset quality, model interpretability, and standardized evaluation practices.

XI. CONCLUSION

This review provides a comprehensive overview of the application of machine learning techniques for predicting concrete compressive strength. The analysis of recent literature indicates that machine learning offers a reliable and efficient alternative to traditional empirical and experimental approaches by enabling accurate strength prediction based on mix composition and curing conditions.

Among the various algorithms reviewed, ensemble learning methods such as XGBoost, Random Forest, and LightGBM consistently achieve the highest predictive accuracy, particularly when combined with metaheuristic optimization techniques. Deep learning architectures, including CNN and LSTM, also show significant potential for modeling complex relationships in advanced concrete materials such as recycled aggregate concrete and geopolymer concrete.

The findings also highlight the importance of feature importance analysis, which consistently identifies parameters such as the water–cement ratio, curing age, cement content, and supplementary materials as the most influential factors affecting compressive strength. These insights can support engineers in optimizing concrete mix design and improving sustainability in construction materials.

However, several limitations still hinder the practical implementation of ML-based prediction models. These include limited and heterogeneous datasets, inconsistent validation methods, and the lack of interpretability in complex models. Addressing these issues will require collaborative efforts to develop large-scale standardized datasets, robust validation frameworks, and explainable AI approaches that enhance trust and transparency in predictive models.

Future research should focus on integrating multi-objective optimization, where machine learning models simultaneously optimize compressive strength, cost efficiency, durability, and environmental impact. In addition, the incorporation of real-time field data and digital construction technologies could further enhance the applicability of ML models in practical engineering environments.

In conclusion, machine learning has the potential to become an essential tool in modern concrete technology by enabling accurate prediction, efficient mix design, and sustainable construction practices, provided that current challenges related to data quality, model interpretability, and standardization are effectively addressed.

REFERENCES

[1] Thapa, "Concrete Compressive Strength Prediction by Ensemble Machine Learning Approach," *Journal of Engineering and Sciences*, vol. 3, no. 1, 2024. DOI: 10.3126/jes2.v3i1.66249

[2] Khandagale et al., "Optimizing Concrete Compressive Strength Prediction: A Comparative Study of Machine Learning Algorithms," in *Proc. IEEE International Conference on Electrical, Electronics and Computer Technology*, 2024. DOI: 10.1109/iceect61758.2024.10738970

[3] Tak et al., "Advanced Machine Learning Techniques for Predicting Concrete Compressive Strength," *Infrastructures*, vol. 10, no. 2, 2025. DOI: 10.3390/infrastructures10020026

[4] Jiang et al., "A Study on the Application of Machine Learning for Predicting Concrete Compressive Strength," *Science and Engineering*, 2023. DOI: 10.57237/j.se.2023.05.003

[5] Logeswaran et al., "Machine Learning Model for Concrete Strength Prediction," in *Proc. IEEE International Conference on Computing, Communication and Networking Technologies*, 2024. DOI: 10.1109/icccnt61001.2024.10724333

[6] Poudel et al., "Machine Learning-Based Prediction of Compressive Strength of Sustainable Concrete Incorporating Waste Glass as a Partial Replacement for Cement," *Preprints*, 2025. DOI: 10.20944/preprints202504.0644.v1

[7] Papakonstantinou et al., "The application of data science and machine learning techniques in predicting the compressive strength of confined concrete," *Technical Annals*, 2024. DOI: 10.12681/ta.37190

[8] Sah et al., "Performance Comparison of Machine Learning Models for Concrete Compressive Strength Prediction," *Materials*, vol. 17, no. 9, 2024. DOI: 10.3390/ma17092075

[9] Alghamdi, "Prediction of Concrete's Compressive Strength via Artificial Neural Network Trained on Synthetic Data," *Engineering, Technology & Applied Science Research*, 2023. DOI: 10.48084/etasr.6560

[10] Mandal, "Predicting Compressive Strength of Concrete Using Advanced Machine Learning Techniques: A Combined Dataset Approach," *Research Square*, 2024. DOI: 10.21203/rs.3.rs-5401974/v1

[11] Zhang et al., "Prediction of compressive strength of recycled aggregate concrete using machine learning and Bayesian optimization methods," *Frontiers in Earth Science*, 2023. DOI: 10.3389/feart.2023.1112105

[12] Guo et al., "Compressive strength prediction of recycled aggregate concrete based on DE-XGBoost algorithm model and explainable machine learning," *Research Square*, 2023. DOI: 10.21203/rs.3.rs-3277626/v1

[13] Qi et al., "A Two-Level Machine Learning Prediction Approach for RAC Compressive Strength," *Buildings*, vol. 14, no. 9, 2024. DOI: 10.3390/buildings14092885

[14] Bu et al., "Prediction of the Compressive Strength of Recycled Aggregate Concrete Based on Artificial Neural Network," *Materials*, vol. 14, no. 14, 2021. DOI: 10.3390/MA14143921

[15] Golafshani et al., "Sustainable mix design of recycled aggregate concrete using artificial intelligence," *Journal of Cleaner Production*, 2024. DOI: 10.1016/j.jclepro.2024.140994

[16] Yang et al., "Prediction on compressive strength of recycled aggregate self-compacting concrete by machine learning method," *Journal of Building Engineering*, 2024. DOI: 10.1016/j.jobe.2024.109055

[17] Tipu et al., "Enhancing Concrete Properties Through the Integration of Recycled Coarse Aggregate: A Machine Learning Approach for Sustainable Construction," in *Proc. IEEE International Conference on Innovative Projects in Technology and Management*, 2024. DOI: 10.1109/iciptm59628.2024.10563490

[18] Jaber et al., "Harnessing AI For Sustainable Construction: Predicting and Optimizing Recycled Aggregate Concrete with Krn-PSO," *International Journal of Environmental Sciences*, 2025. DOI: 10.64252/vt6cy239

[19] Ranjbar et al., "A combination of deep learning and genetic algorithm for predicting the compressive strength of high-performance concrete," *Structural Concrete*, 2022. DOI: 10.1002/suco.202100199

[20] Chen et al., "Axial compressive strength predictive models for recycled aggregate concrete filled circular steel tube columns using ANN, GEP, and MLR," *Journal of Building Engineering*, 2023. DOI: 10.1016/j.jobe.2023.107439

[21] Feng et al., "Proposing Optimized Random Forest Models for Predicting Compressive Strength of Geopolymer Composites," *Infrastructures*, vol. 9, no. 10, 2024. DOI: 10.3390/infrastructures9100181

[22] Wang et al., "ICA-LightGBM Algorithm for Predicting Compressive Strength of Geo-Polymer Concrete," *Buildings*, vol. 13, no. 9, 2023. DOI: 10.3390/buildings13092278

[23] Chang et al., "Soft Computing Techniques to Model the Compressive Strength in Geo-Polymer Concrete: Approaches Based on an Adaptive Neuro-Fuzzy Inference System," *Buildings*, vol. 14, no. 11, 2024. DOI: 10.3390/buildings14113505

[24] Patro, "Prediction of compressive strength of geopolymer concrete using a hybrid ensemble of grey wolf optimised machine learning estimators," *Journal of Building Engineering*, 2023. DOI: 10.1016/j.jobe.2023.106521

[25] Chen et al., "A Novel MBAS-RF Approach to Predict Mechanical Properties of Geopolymer-Based Compositions," *Gels*, vol. 9, no. 6, 2023. DOI: 10.3390/gels9060434

[26] Zhang et al., "Prediction of Compressive Strength of Geopolymer Concrete Landscape Design: Application of the Novel Hybrid RF–GWO–XGBoost Algorithm," *Buildings*, vol. 14, no. 3, 2024. DOI: 10.3390/buildings14030591

[27] Khatri et al., "Prediction Of Strength Properties Of Geopolymer Concrete By Using Artificial Intelligence And Machine Learning Techniques," *KUEY*, vol. 30, no. 5, 2024. DOI: 10.53555/kuey.v30i5.5124