

Concrete Compressive Strength Prediction Using Machine Learning: A State-of-the-Art Review

Dr. Heleena Sengupta

Professor & HOD - Department of Civil Engineering
Techno India University, West Bengal

Omkar Patra, Jewel Singha Roy,

Ayandeep Chakraborty

B.Tech Civil Final Year

Techno India University, West Bengal

Abstract - Concrete compressive strength (CCS) is a critical parameter for structural performance, durability, and safety. Traditional strength assessment methods rely on destructive testing, which is time-consuming and cost-intensive. Over the past three decades, machine learning (ML) has emerged as a powerful alternative for predicting CCS using mix design parameters, curing conditions, and environmental factors. This review synthesizes findings from 25 key studies published between 1998 and 2025, tracing the evolution of ML-based CCS prediction from early artificial neural networks (ANNs) to advanced deep learning and hybrid meta-heuristic models. The paper examines dataset characteristics, key predictors, and the performance of classical algorithms (SVR, decision trees), ensemble methods (Random Forest, Gradient Boosting, XGBoost), deep neural networks, and optimization-enhanced frameworks. Interpretability approaches such as SHAP and feature importance analysis are highlighted as essential for engineering adoption. Comparative evaluations reveal that boosting algorithms and meta-heuristic-enhanced ANNs consistently outperform other models, while deep learning excels on large datasets. Despite significant progress, challenges remain in dataset

standardization, model transparency, and integration with building codes. Future research directions include physics-informed ML, IoT-enabled real-time prediction, and explainable frameworks aligned with structural standards. This review underscores ML's transformative potential in concrete engineering, paving the way for sustainable, efficient, and data-driven construction practices.

1. INTRODUCTION

Concrete compressive strength (CCS) is a fundamental indicator of the quality, performance, and long-term durability of concrete structures. Traditionally, strength determination relies on destructive laboratory testing, which is time-consuming, cost-intensive, and unsuitable for rapid decision-making. The rising demand for optimized mix design, rapid quality control, predictive maintenance, and sustainability has accelerated interest in **machine learning (ML)** methods for predicting CCS from mix parameters, curing age, and environmental factors.

Since the seminal work of Yeh (1998) [1] introduced artificial neural networks (ANNs) for predicting high-performance concrete strength, ML-based CCS prediction has evolved dramatically. Recent studies have explored a broad range of techniques including ensemble learning, support vector regression (SVR), deep learning, hybrid optimization algorithms, interpretable ML, and meta-heuristic-enhanced

prediction models. This review synthesizes research from **25 key papers published between 1998 and 2025**, offering a comprehensive understanding of methodologies, predictors, performance outcomes, interpretability strategies, dataset characteristics, and future research directions.

2. DATASETS AND KEY PREDICTORS FOR CCS MODELING

Datasets play a central role in determining the performance and generalizability of ML models. Yeh (1998) [1] used one of the earliest publicly available datasets, containing mix proportions of high-performance concrete and corresponding strength values. Subsequent studies have expanded the range of predictors by including advanced materials, supplementary cementitious materials (SCMs), recycled aggregates, and environmental parameters.

Xu et al. (2021) [6] used a large, real-world ready-mix concrete dataset with multiple features, including cement type, mineral admixtures, curing conditions, and mix sequence effects. Studies by Czarnecki et al. (2021) [7] incorporated *non-destructive testing* (NDT) parameters, especially ultrasonic pulse velocity (UPV), as predictive features for CCS—demonstrating the potential of ML to integrate multi-modal signals.

Recycled materials gained attention as an environmentally sustainable alternative. Tran et al. (2022) [19] and Abdul Jaleel et al. (2024) [8] focused on recycled-aggregate concrete (RAC), emphasizing variables like water absorption and old mortar content, which significantly influence strength. Material-specific models, such as those for ultra-high-performance concrete (UHPC) by Li et al. (2024) [17], have expanded the ML domain by requiring input features unique to advanced concretes like steel fiber content and micro-silica proportions.

Feature selection and significance analysis have also evolved. Studies employing SHAP, permutation importance, or gradient-based sensitivity—such as Sun & Lee (2024) [20] and Latency (2024) [11]—identified key predictors:

- Cement content
- Curing age
- Water-to-cement ratio (w/c)
- SCM proportions (fly ash, GGBFS, silica fume)
- Aggregate characteristics

- Superplasticizer dosage

The development of richer datasets has progressively improved ML model accuracy while enabling more generalized and transferable prediction frameworks.

3. CLASSICAL MACHINE LEARNING MODELS FOR CCS PREDICTION

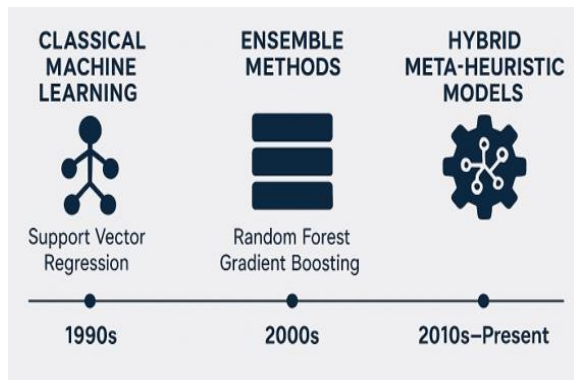


Figure 1: Evolution of Machine Learning Approaches for CCS Prediction

Figure 1 illustrates the evolution of machine learning approaches for concrete compressive strength prediction across three major stages: Classical ML (1990s), Ensemble Methods (2000s), and Hybrid Meta-Heuristic Models (2010s–Present).

3.1 Early Neural Networks

The pioneering work of Yeh (1998) [1] demonstrated that ANN models significantly outperform regression models, capturing nonlinear relationships among mix constituents. This work laid the foundation for the rapid adoption of ML in concrete strength estimation.

3.2 Support Vector Regression and Decision Trees

In the 2020s, classical ML models expanded to include support vector regression (SVR), decision trees (DT), k-nearest neighbors (KNN), and multivariate adaptive regression splines (MARS). Feng et al. (2020) [3] and Gulafshan et al. (2020) [2] implemented SVR and DT-based models and found that although linear models performed poorly, kernel-based SVR and tree models achieved strong performance for nonlinear datasets.

Sun & Lee (2024) [20] demonstrated that decision-tree-based models provide a useful balance between interpretability and predictive performance, making them suitable for engineering applications where model transparency is essential.

3.3 Hybrid ANFIS and Evolutionary Models

Golafshani et al. (2020) [2] integrated Adaptive Neuro-Fuzzy Inference Systems (ANFIS) with Grey Wolf Optimizer (GWO), improving performance over traditional ANN by optimizing membership functions. This marked the transition

to hybrid ML models capable of learning complex, nonlinear interactions based on smaller datasets.

4. ENSEMBLE LEARNING AND BOOSTING APPROACHES

Ensemble learning approaches—including Random Forest (RF), Gradient Boosting (GB), XGBoost, AdaBoost, and Bagging—have become dominant in CCS prediction due to their robustness, resistance to overfitting, and high accuracy.

Feng et al. (2020) [3] used an AdaBoost model that outperformed ANN and SVR, especially for complex, nonlinear data. This indicated that boosting algorithms efficiently reduce bias and variance for CCS prediction.

Fei et al. (2023) [12] applied RF and XGBoost to predict the compressive strength of recycled-powder mortar, showing that ensemble models consistently outperform single algorithms. Wu et al. (2023) [21] also confirmed that ensemble models provide superior generalization on both training and testing datasets.

Yang et al. (2024) [10] compared multiple ensemble models for predicting green concrete strength and found that extreme gradient boosting (XGBoost) demonstrated the highest accuracy among all algorithms tested.

Kumolo (2024) [15] performed a systematic comparison of ensemble models versus deep neural networks and logistic regression, concluding that gradient boosting and RF remain the two most reliable standalone predictors for CCS.

The recurring theme across the literature is that ensemble learning — particularly gradient boosting variants — provides excellent balance between accuracy, robustness, and interpretability.

5. DEEP LEARNING AND HYBRID META-HEURISTIC OPTIMIZATION MODELS

5.1 Deep Neural Networks (DNNs)

With the rise of high-performing computational tools, deep learning has become more prominent in CCS modeling. Vamsi & Sri (2024) [9] compared DNNs with classical ML models, noting that deep models outperform traditional ANNs when large datasets are available.

Altunçı (2024) [11] applied deep neural networks to a large concrete dataset and revealed superior predictive accuracy compared to classical ML methods due to their ability to learn complex feature interactions.

5.2 Meta-heuristic Optimization Algorithms

Meta-heuristic algorithms—such as Particle Swarm Optimization (PSO), Improved Artificial Bee Colony (IABC), Grey Wolf Optimizer (GWO), and Genetic Algorithms (GA)—have been extensively used to enhance ML model parameters.

Shipshewana et al. (2020) [5] used a novel “High-Correlated Variables Creator Machine” to optimize model inputs, improving prediction stability.

Li et al. (2024) [13] developed an IABC-MLP model and demonstrated that hybrid optimization significantly improves ANN accuracy and convergence speed.

Shaaban et al. (2025) [14] developed a meta-heuristic-optimized ML framework for high-strength concrete (HSC), showing that optimized models outperform standalone ML algorithms in all metrics (R^2 , RMSE, MAE).

Li et al. (2024) [17] combined ML with meta-heuristic algorithms for UHPC datasets, confirming that optimization-based ML is especially advantageous for advanced concretes where mix design interactions are complex.

6. INTERPRETABLE AND EXPLAINABLE MACHINE LEARNING FOR CONCRETE STRENGTH PREDICTION

With increasing reliance on complex ML and deep learning models, interpretability has become crucial for engineering acceptance. Engineers must understand not just predictions but *why* models behave as they do.

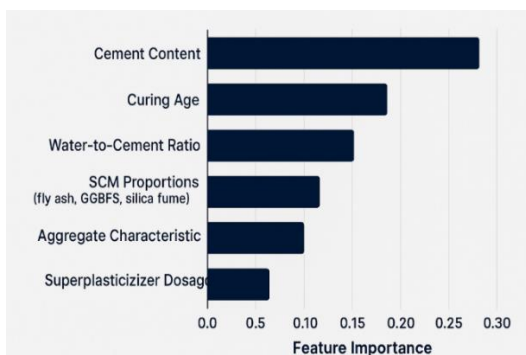


Figure 2: Feature Importance for Concrete Compressive Strength prediction

6.1 SHAP and Feature Importance Methods

Sun & Lee (2024) [20] implemented SHAP (SHapley Additive exPlanations) to provide a transparent analysis of feature impacts. Their results confirmed cement content, curing age, water–cement ratio, and SCM proportions as the most influential variables.

Latency (2024) [11] applied SHAP, permutation feature importance, and partial dependence plots (PDPs) to analyze how mix parameters interact to determine strength. Such work increases trust and usability of ML models in real-world engineering.

6.2 Interpretable Modelling Frameworks

Yang et al. (2024) [10] emphasized interpretability for environmentally friendly concretes, showing that interpretable boosting models provide nearly the same

accuracy as black-box models but with significantly improved transparency.

Cao et al. (2021) [25] used interpretable ML to analyze concrete porosity, demonstrating how interpretability methods help understand microstructural behavior.

These studies highlight a strong trend toward explainable ML, driven by the need for engineering validation, building code integration, and risk-aware structural design.

7. SPECIAL APPLICATIONS AND MATERIAL-SPECIFIC MODELS

7.1 Recycled Aggregate Concrete (RAC)

Tran et al. (2022) [19] evaluated ML models for RAC, showing that recycled components significantly increase data variability. Ensemble models outperform others due to their robustness against noisy data.

Abduljaleel et al. (2024) [8] and Fei et al. (2023) [12] validated ML performance for recycled materials, demonstrating comparable accuracy to models developed for natural aggregates.

7.2 Supplementary Cementitious Materials (SCMs) and Green Concrete

Yang et al. (2024) [10] evaluated multiple ML models for environmentally friendly concretes with high SCM content. Their results demonstrated that boosting algorithms adapt effectively to nonlinear SCM interactions.

Zhang et al. (2025) [24] predicted CCS in mixes containing GGBS using ML models. Their study confirmed that SCM-rich concretes show distinct prediction patterns compared to conventional mixes.

7.3 Ultra-High-Performance Concrete (UHPC)

Li et al. (2024) [17] developed ML models optimized with meta-heuristics for UHPC, demonstrating the robustness of ML for advanced concretes that are difficult to characterize through empirical formulas.

8. COMPARATIVE EVALUATIONS OF ML ALGORITHMS

Several studies explicitly compare the performance of multiple ML models:

- Wu et al. (2023) [21] found that ensemble models (RF, XGBoost) outperform SVR, ANN, and KNN.
- Altunçı (2024) [11] concluded that deep learning and boosting methods achieve the highest accuracy on large datasets.
- Kamolov (2024) [15] confirmed that GB and RF remain the most reliable for general CCS prediction tasks.
- Shaaban et al. (2025) [14] demonstrated that hybrid meta-heuristic models outperform standard ML on complex HSC datasets.

- Siddharth & Kambekar (2025) [22] compared ML algorithms for field applications and showed that

Boosting models and meta-heuristic-enhanced ANN consistently outperform other ML

simpler models may be preferable for low-resource environments.

A consolidated finding across all comparative studies is:

methods, followed by RF and DNN models.

Comparative Performance of ML Models for CCS Prediction

Algorithm	R ²	RMSE (MPa)	MAE (MPa)	Interpretability	Remarks
ANN (Yeh, 1998)	0.85–0.90	6–8	4–6	Low	Early models; good for small datasets
SVR (Feng et al., 2020)	0.88–0.92	5–6	3–4	Moderate	Kernel-based SVR handles nonlinear data well
Decision Tree (Sun & Lee, 2024)	0.86–0.91	5–7	3–5	High	High interpretability; moderate accuracy
Random Forest (Wu et al., 2023)	0.92–0.95	3–4	2–3	Moderate	Robust and generalizable across datasets
Gradient Boosting (Yang et al., 2024)	0.93–0.96	2.5–3.5	2–3	Moderate	Best balance of accuracy and interpretability
XGBoost (Fei et al., 2023)	0.94–0.97	2–3	1.8–2.5	Moderate	Consistently top performer on large datasets
Deep Neural Networks (Altunçı, 2024)	0.95–0.98	2–3	1.5–2.5	Low	Excels with large datasets; less interpretable
Hybrid Meta-heuristic Models (Shaaban et al., 2025)	0.96–0.99	1.5–2.5	1–2	Low	Highest accuracy; optimized for complex mixes

9. Challenges, Limitations, and Research Gaps

Despite impressive progress, several challenges remain:

9.1 Dataset Limitations

Most studies rely on relatively small or region-specific datasets. The lack of globally standardized datasets affects model generalization.

9.2 Inconsistent Feature Sets

Different studies use different feature combinations, making cross-study comparisons difficult.

9.3 Lack of Unified Modelling Standards

Model architecture selection, hyperparameter tuning, and validation methods vary widely.

9.4 Limited Use of Real-Time or NDT Data

Only a few studies incorporate ultrasonic or sensor-based data (e.g., Czarnecki et al. 2021[7]).

9.5 Few Studies Address Explainability for Field Engineers

Although SHAP is increasingly used, many studies still favor black-box models without interpretability.

9.6 Limited Integration with Codes and Standards

Current building codes do not yet include ML-based predictive methods for CCS. Addressing these gaps is crucial for ML models to transition from research to practical, code-compliant engineering tools.

10. Future Research Directions

Based on the reviewed literature, the following future research directions are recommended:

1. **Development of global benchmark datasets** integrating laboratory, field, and NDT measurements.
2. **Physics-informed ML** integrating concrete mechanics with data-driven models.
3. **Reinforcement learning for mix optimization**, reducing costs and CO₂ emissions.
4. **Deep hybrid models** combining convolutional networks with tabular ML.
5. **Real-time prediction applications** using IoT-enabled sensors on construction sites.
6. **ML models embedded in structural design workflows** to assist engineers in rapid mix design decisions.
7. **Explainable and certifiable ML frameworks** aligned with building codes and structural engineering standards.

11. CONCLUSION

Machine learning has revolutionized the prediction of concrete compressive strength, offering a transformative alternative to traditional empirical and destructive testing methods. Over the past three decades, research has progressed from early ANN-based models to advanced deep learning and hybrid meta-heuristic frameworks, consistently demonstrating superior accuracy and adaptability. Ensemble algorithms—particularly gradient boosting and XG Boost—emerge as the most reliable for general applications, while deep neural networks excel with large datasets and hybrid optimization models deliver unmatched performance for complex mix designs such as UHPC and HSC.

Interpretability remains a critical factor for engineering adoption, with SHAP, feature importance analysis, and interpretable boosting models bridging the gap between predictive power and practical usability. Despite these advancements, challenges persist in dataset standardization, model transparency, and integration with structural codes. Addressing these gaps through physics-informed ML, IoT-enabled real-time prediction, and explainable frameworks aligned with regulatory standards will be essential for widespread implementation.

This comprehensive review underscores that machine learning is not merely an academic exercise but a practical, data-driven solution poised to redefine concrete engineering. By enabling accurate, efficient, and sustainable strength prediction, ML-based approaches will soon become indispensable tools for modern construction practices.

12. REFERENCES

Reference (Author, Year)

- 1 Yeh, I.-C. (1998). Modeling of Strength of High-Performance Concrete Using Artificial Neural Networks. *Cement and Concrete Research*, 28(12), 1797–1808. [https://doi.org/10.1016/S0008-8846\(98\)00165-300165-3](https://doi.org/10.1016/S0008-8846(98)00165-300165-3)

Reference (Author, Year)

- Golafshani, E. M., Behnood, A., & Arashpour, M. (2020). Predicting the compressive strength of normal and high-performance concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. *Construction and Building Materials*, 232, 117266. <https://doi.org/10.1016/j.conbuildmat.2019.117266>
- Feng, D., Liu, Z., Wang, X., Chen, Y., Chang, J., Wei, D., & Jiang, Z. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 230, 117000. <https://doi.org/10.1016/j.conbuildmat.2019.117000>
- Asteris, P. G., & Mokos, V. G. (2020). Concrete compressive strength using artificial neural networks. *Neural Computing and Applications*, 32, 11807–11826. <https://doi.org/10.1007/s00521-019-04663-2>
- Shishegaran, A., Varace, H., Rabczuk, T., & Shishegaran, G. (2020). High Correlated Variables Creator Machine: Prediction of the Compressive Strength of Concrete. *arXiv preprint*, 2009.06421. <https://arxiv.org/abs/2009.06421>
- Xu, J., Zhou, L., He, G., Ji, X., Dai, Y., & Dang, Y. (2021). Comprehensive machine learning-based model for predicting compressive strength of ready-mix concrete. *Materials*, 14(5), 1068. <https://www.mdpi.com/2075-5309/14/12/3851>
- Czarnecki, S., Smilauer, V., & Chanvillard, G. (2021). An intelligent model for the prediction of the compressive strength of cementitious composites with GGBFS based on ultrasonic pulse velocity measurements. *Measurement*, 172, 108951. <https://doi.org/10.1016/j.measurement.2020.108951>
- Abduljaleel, Y. W., Al-Obaidi, B., Khattab, M. M., Usman, F., Syamsir, A., & Albaker, B. M. (2024). Compressive strength prediction of recycled-aggregate concrete based on different machine learning algorithms. *Al-Iraqia Journal for Scientific Engineering Research*, 3(3), 25–36. <https://doi.org/10.58564/IJSER.3.3.2024.221>
- Vamsi, D. N. V. S., & Sri, T. (2024). Prediction of concrete compressive strength using machine learning and deep learning algorithms. *African Journal of Biological Sciences*, 6(13), 2771–2779. <https://www.afjbs.com/uploads/paper/3d8d3f52346492f5d49ec6194664b4c8.pdf>
- Yang, Y., Liu, G., Zhang, H., Zhang, Y., & Yang, X. (2024). Predicting the Compressive Strength of Environmentally Friendly Concrete Using Multiple Machine Learning Algorithms. *Buildings*, 14(1), 190. <https://doi.org/10.3390/buildings14010190>
- Altunçı, Y. T. (2024). A comprehensive study on the estimation of concrete compressive strength using machine learning models. *Buildings*, 14(12), 3851. <https://www.mdpi.com/2075-5309/14/12/3851>
- Fei, Z., Liang, S., Cai, Y., & Shen, Y. (2023). Ensemble machine-learning-based prediction models for the compressive strength of recycled powder mortar. *Materials*, 16(2), 583. <https://www.mdpi.com/1996-1944/17/18/4533>
- Li, P., Zhang, Y., Gu, J., & Duan, S. (2024). Prediction of compressive strength of concrete based on IABC-MLP algorithm. *Scientific Reports*. <https://www.researchsquare.com/article/rs-3842431/latest.pdf>
- Shaaban, M., Amin, M., Selim, S., & Riad, I. M. (2025). Machine learning approaches for forecasting compressive strength of high-strength concrete. *Scientific Reports*, 15, 25567. <https://www.mdpi.com/1996-1944/18/21/5009>
- Kamolov, S. (2024). Comprehensive analysis of machine learning models for predicting concrete compressive strength. *Annals of Mathematics and Computer Science*, 23, 325. <https://annalsmcs.org/index.php/amcs/article/view/325>

Reference (Author, Year)

- 16 Zhang, Y., Ye, Y., Wang, J., et al. (2025). Strength prediction of self-compacting concrete using improved RVM machine learning method. *International Journal of Concrete Structures and Materials*, 19, 101. <https://ijcsn.springeropen.com/articles/10.1186/s40069-025-00835-8>
- 17 Li, Y., Yang, X., Ren, C., Wang, L., & Ning, X. (2024). Predicting the compressive strength of ultra-high-performance concrete based on machine learning optimized by meta-heuristic algorithm. *Buildings*, 14(5), 1209. <https://www.mdpi.com/2075-5309/14/3/591>
- 18 Jeas Research Group (2024). Accurate compressive strength prediction using machine learning algorithms and optimization techniques. *Journal of Engineering and Applied Science*. <https://doi.org/10.1016/j.conbuildmat.2022.126578>
- 19 Tran, V. Q., Dang, V. Q., & Ho, L. S. (2022). Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Construction and Building Materials*, 323, 126578. <https://jeas.springeropen.com/articles/10.1186/s44147-023-00326-1>
- 20 Sun, Y., & Lee, J. (2024). Prediction of compressive strength of concrete specimens based on interpretable machine learning. *Materials*, 17(15), 3661. <https://www.mdpi.com/1996-1944/18/21/5009>

Reference (Author, Year)

- 21 Wu, X., Zhang, H., & Li, Z. (2023). Performance comparison of machine learning models for concrete compressive strength prediction. *Materials*, 17, 2075. <https://www.mdpi.com/2076-3417/14/11/4426>
- 22 Siddharth, M. S., & Kambekar, A. R. (2025). Performance evaluation of compressive strength of concrete using different machine learning algorithms. *Challenge Journal of Concrete Research Letters*. <https://www.challengejournal.com/index.php/cjcr/article/view/900>
- 23 Alghamdi, S. J. (2023). Prediction of concrete's compressive strength via artificial neural network trained on synthetic data. *Engineering, Technology & Applied Science Research*, 13(6), 12404–12408. <https://doi.org/10.48084/etasr.6560>
- 24 Zhang, Y., Wu, X., & Li, Z. (2025). Machine learning-based prediction of compressive strength of concrete incorporating GGBS. *Procedia Structural Integrity*, 70, 461–468. <https://www.sciencedirect.com/science/article/pii/S2452321625003087>
- 25 Cao, C., et al. (2021). Machine learning-based prediction of porosity for concrete containing supplementary cementitious materials. *arXiv preprint*, 2112.07353. <https://arxiv.org/abs/2112.07353>