IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Comparative analysis of concrete strength prediction analysis using Machine learning

Kalpesh Wani[1], Prashant Shimpi[2]

[1]PG-Student, Professor ,Department of Computer Engineering :0009-0009-6580-2075 Godavari Collage of Engineering Jalgaon ,Maharashtra,india 425001

[2]Assistant Professor, Department of Computer Engineering ::0009-0004-9588-4298 Godavari collage of Engineering Jalgaon ,Maharashtra,india,425001

Abstract *– This paper presents a critical review and comparative evaluation of machine learning methods to predict the compressive strength of concrete with particular emphasis on mixes that incorporate supplementary cementitious materials. The analysis includes ensemble algorithms (e.g., Random Forest, CatBoost, XGBoost, AdaBoost, Gradient Boosting), regression algorithms (Support Vector Regression, Linear Regression), and neural networks (Artificial Neural Networks, Extreme Learning Machines) by synthesizing the results of the latest studies. CatBoost was the most effective one, with the highest R 2 values of up to 0.94 in several studies. The most significant parameters found are concrete age, water-to-binder ratio, and cement content. Other important gaps in current literature that are identified during the review are inconsistency in datasets, lack of external model validation, lack of interest in hybrid models, and real-world implementation barriers. These results provide a basis on which universal and predictable forecasting models can be developed to more sustainably build concrete constructions.*

Keywords*-RandomForest,,CatBoost,,LinearRegression,,NeuralNetwork*

INTRODUCTION

The most commonly used construction material globally is concrete, as only water is consumed more than concrete. This is due to its superior versatility, cost efficiency and reliable structural functionality. An important property of concrete is the capacity to transition between a liquid, working mixture to a solid, load bearing material, and has thus enabled the erection of numerous infrastructures, such as residential homes, multifaceted bridges, and skyscrapers. However, there are significant environmental negativities associated with the prevalence of concrete. The production of Portland cement, the most common binder, consumes a lot of energy and has a calcification process that produces large amounts of carbon dioxide. Cement industry contributes to the anthropogenic emissions of CO_2 in the world, by itself, ranging between 7 to 8 percent, which poses a significant burden on the climate change mitigation efforts. These are the environmental issues and the loss of the natural resources that have led to a shift towards more sustainable construction practices.

In response to these challenges, the use of supplementary cementitious materials (SCMs) has become an effective solution. A fly ash, ground granulated blast-furnace slag, and silica fume can be used to substitute some of the Portland cement, and reduce the carbon footprint of concrete manufacture. Fly ash, which is a by-product of coal burning, offers some benefits, such as, higher workability, lower hydration heat, higher long-term strength, and higher durability characteristics, such as, greater resistance to permeability and sulfate attack. Secondary hydration reaction in blended fly ash results in denser microstructure and better transition zone between interfaces which translates to concrete with a long service life and reduced maintenance requirements. Moreover, the use of fly ash may also reduce cement consumption by about 10 percent without affecting similar strength and provide economic and environmental benefits.

In spite of these advantages, the compressive strength of concrete and mixtures, in particular, with SCMs, is still a challenge to predict properly. The most important mechanical property in terms of structural design and quality control is compressive strength. Nevertheless, the process of development of strength is not simple and is controlled by non-linear relationships between mixture components, proportions, curing regimes, and age. Traditional laboratory tests are slow, labor intensive and can only conduct a small sample. The multi-variable relationship that characterizes concrete behavior is frequently too complex to be explained by empirical formula or statistical regression. With the rise of machine learning (ML), predictive modeling in most types of engineering has changed. Many recent works have used ML methods to estimate concrete compressive strength, especially in mixes with industrial byproducts such as fly ash. Ensemble (e.g., AdaBoost, Random Forest, XGBoost, CatBoost) and non-ensemble models have demonstrated higher predictive power than traditional regression models when using such inputs as cement content, water content, curing age, and aggregate proportions. The use of ML to create sustainable, low-carbon materials has been supported by research; for instance, Rand Forest models with just-in-time learning have increased the accuracy of prediction with life-cycle information. Researchers have also found that CatBoost, K-Nearest Neighbors (KNN), and Support Vector Regression (SVR) are effective, with CatBoost having lower error rates and large correlation coefficients. Research on high-performance concrete (HPC) has determined that XGBoost has been one of the most accurate among the models tested.

The increasing literature is evidence of increased sophistication in ML applications. Artificial Neural Networks (ANN), random forest, and Long Short-Term Memory (LSTM) networks are some of the techniques that allow the analysis of complex variable interactions effectively. The development of classical methods of statistics to modern deep learning and ensemble models is thoroughly recorded, and genetic algorithms are becoming more popular to optimize the efficiency of models and hyperparameter optimization. New activity also emphasizes the incorporation of explainable AI systems like SHAP to increase the interpretability of a model and the creation of graphical user interfaces to increase practicality. In addition, the supervised learning models such as Decision Trees, bagging ensembles and hybrid models have been effective in predicting fly ash-based concrete and comparative analyses between hybrid ML models and optimized methods are also scarce. Machine learning algorithms are also very good at detecting non-linear tendencies and intricate patterns in high-dimensional data, which is why they are best suited to forecast concrete compressive strength. Through feature learning, the ML models can develop mapping functions between input variables and the resulting strength and make predictions quickly and inexpensively to optimize the mix design and reduce project timelines. Nevertheless, even though there are many separate studies, the systematic review that summarizes the results of various algorithms, data sets and experimental conditions is lacking. The relative performance of different ML methods, the similarity of rankings of feature importance, and the generalizability of results remain scattered and inconclusive since most of the studies investigate the specific algorithms, specific concrete types, or specific datasets.

MACHINE LEARNING FOR CEMENT STRENGTH PREDICTION

Machine learning has emerged as a revolutionary method of predicting the compressive strength of concrete, and it has conquered the shortcomings of the traditional experimental methods, which are not only time-consuming and costly but also often inaccurate. With the help of the previous experimental data, the ML algorithms are outstanding at describing the non-linear relationships between mix components and the resultant strength. Such models are fed with input variables such as cement, water to binder ratio, content of aggregates, content of supplementary cementitious materials (fly ash and slag), and dosage of superplasticizer and time of curing to produce compressive strength predictions. Ensemble methods such as Random Forest, Gradient Boosting, and CatBoost are one of the techniques that have demonstrated excellent results through the aggregation of multiple learners to reduce both prediction bias and variance. Moreover, interpretability methods such as SHAP analysis add to the model transparency, as they identify the most significant parameters, which helps engineers to refine mix designs at a higher efficiency.

LITERATURE REVIEW

Other recent studies have explored the use of machine learning (ML) methods to predict compressive strength of concrete, especially mixtures using industrial byproducts including fly ash. Paudel and colleagues [1] state that ensemble models (e.g., AdaBoost, Random Forest, XGBoost) and non-ensemble methods have a higher predictive accuracy than traditional regression. The variables that are normally considered in these models are cement dosage, content of water, curing time and aggregate properties. However, the authors observe that sensitivity analysis and optimization of input variables have been under-researched, and stronger and more generalizable predictive models are required [1]. According to a study by Zviazhynski et al. (2025), machine learning is important in predicting concrete strength, particularly in sustainable and low-carbon material. The authors used the models of Random Forest with just-in-time learning to improve the accuracy of the prediction on the basis of life-cycle data. Their results demonstrate the importance of hyperparameter fine-tuning using both accuracy and uncertainty indicators. Nevertheless, they also find out that there are chronic problems, such as small training datasets, and insufficient validation of long-term strength forecasts. Addition of fresh-state parameters like slump has been demonstrated to enhance predictive stability with the exception of greater applicability to diverse concrete mixtures still being a desired outcome [2].

As shown by Beskopylny and co-authors (2022), machine learning methods are effective in forecasting concrete compressive strength with the help of assembled experimental data. The predictive accuracy was the best in algorithms like CatBoost, K-Nearest Neighbors (KNN), and Support Vector Regression (SVR). CatBoost was the best of all these with lower error rates and higher correlation coefficients. However, the paper also recognizes the current challenges associated with managing heterogeneous data, reduced feature generalization, and lack of exploration of hybrid or adaptive models that would be used in real-time construction projects [3]. Liu and others (2022) concentrated on the use of machine learning to forecast compressive strength of high-performance concrete (HPC). Models such as XGBoost, random forest, and the SVR also showed good predictive performance using experimental data with XGBoost recording the highest accuracy and lowest levels of errors. The study highlights the importance of selecting a model that improves prediction. Nevertheless, limitations like the use of small datasets and lack of validation in different real world construction scenarios are identified as limitations [4].According to Batra et al. (2023), a crucial aspect of the structural quality and performance in civil engineering is to predict the compressive strength of concrete. The conventional experimental methods are time consuming and cumbersome because of various controlling variables such as the water content, temperature and the type of concrete. Recent studies suggest using machine learning methods, including Artificial Neural Networks (ANN), Random Forest, and Long Short-Term Memory (LSTM) networks, to enhance the accuracy of prediction. Such techniques allow the analysis of multi-variable relationships effectively, but there are still difficulties in the choice of models, optimization of features, and the work with various real-world data [5].

Recent developments in the prediction of concrete compressive strength (CCS) signify a shift in the traditional usage of statistics to the development of advanced machine learning (ML) models. The findings of Satyanarayana et al. (2024) and other up-to-date works are that researchers tend to use deep learning and ensemble models to improve the accuracy of prediction, especially when it comes to concrete mixes with supplemental cementitious materials. Random Forest Regression and AdaBoost algorithms have proven to be very strong. Moreover, genetic algorithm and other optimization methods are being applied to optimize the model performance and hyperparameter optimization. However, there is a lack of comparative evaluations of hybrid ML models and optimized methods, which highlights the necessity of combined predictive-optimization schemes [6]. This change in paradigm to using sophisticated ML models is supported by further studies on the prediction of CCS. Random Forest, Gradient Boosting, and XGBoost, also known as ensemble techniques, have shown themselves to be effective in large datasets, as well as deep learning techniques. Predictive accuracy is enhanced even more by hybrid and meta-heuristic models. Model interpretability has also been enhanced by using explainable AI tools, including SHAP. In spite of these achievements, there are still several issues in data standardization, model transparency, integration with structural engineering practice, and real-time field application [7].

Other articles highlight the importance of ML models with interactive tools to predict CCS. It is demonstrated that ensemble methods, in particular, Gradient Boosting (CatBoost), are more accurate and achieve lower error rates than more traditional models of regression and neural networks. The use of explainable AI methods such as SHAP adds to the increased interpretability, whereas practical usability can be improved through the incorporation of graphical user interfaces (GUIs). Nevertheless, the problems regarding the generalization of models, the diversity of the datasets, and the implementation in the actual construction setting are not addressed [8]. In a similar study, Omotayo et al. (2024) tested six ML algorithms with 526 data points of ordinary Portland cement concrete without pozzolans. Their results pointed to the better behavior of the boosting algorithms, with CatBoost having the best accuracy (R 2 = 0.94). The ratio of the water-binder turned out to be the strongest input factor, and its importance to the development of strength cannot be overestimated [9]. Finally, the recent research on fly ash-based concrete has demonstrated increased use of ML techniques. The use of supervised learning models, which are Artificial Neural Networks, Decision Trees and ensemble models such as bagging have shown better predictive performance due to mix design parameters and curing conditions. Comparative analyses show that ensemble models tend to perform better as compared to individual ones. However, there are still constraints including variability of datasets, applicability in the real world, and absence of deep learning and optimization strategies integration [10]. Machine learning has emerged as a viable method to forecast the compressive strength of concrete, and surpass the limitations of conventional laboratory-based methods. Five tree-based regression algorithms were tested on 1,030 samples of mix constituents of different mixes including fly ash and slag (Selvi et al., 2025). The findings showed that CatBoost was more effective than the other models. SHAP-based interpretability analysis evidenced that the concrete age and cement content had the strongest effect on strength, and the importance of temporal and compositional variables in predictive modeling could not be underestimated [11]. Likewise, machine learning has been effectively used to predict the compressive strength of fly ash concrete to offer cost-effective alternatives to the conventional testing methods. A comparative study of extreme learning machine, random forest, standard support vector regression (SVR), and a grid search-optimized SVR model was done by Jiang et al. (2025) on a sample of 270 samples. Their results revealed that the hybrid SVR model with optimization was the best in terms of predictive accuracy. The most significant were found to be age and water content with other variables such as superplasticizer and fly ash content having relatively lesser significance [12].

Limitations of Existing Research

Despite the merits of the available literature, some common weaknesses can be traced throughout the literature. First, the characteristics of the data sets are very heterogeneous in terms of sample size, composition, and quality, with the values varying between 270 and more than 1,700 data points. This heterogeneity makes it difficult to compare the results of direct cross-studies and compromises the generalizability of models. The use of limited training data and lack of long-term strength predictions validation are also persistent problems. Second, most studies are conducted on small-scale specific categories, including ordinary Portland cement concrete in the absence of pozzolans or mixtures that contain only one type of supplementary cementitious material (SCM). This leaves a knowledge gap on understanding the performance of models in a wider range of mix designs. Third, sensitivity analysis and optimization of input variables have been relatively understudied, which indicates the necessity to design stronger and more generalized predictive frameworks. Fourth, the persistent challenges are managing heterogeneous datasets, poor extrapolation of features, and in-depth investigation of hybrid or adaptive models that can be used in real-time construction settings. Fifth, there is no external validation with other datasets, which poses the question of possible overfitting and the actual generalizability of obtained results. Sixth, despite the explainable artificial intelligence (XAI) tools like SHAP, there are still difficulties in the fields of data standardization, transparency, and connection to the structural engineering practice and field performance. Lastly, the practical deployability and computational efficiency of complex ensemble models in practical construction environments is a understudied topic, as is the integration of these predictive models in the current quality control processes. The main aim of the current paper is to address these limitations by applying a systematic review and comparative analysis.

COMPARATIVE ANALYSIS

Table 1: Comparative Analysis

Authors & Their Work	Methodology Used	Research Gap
Satish Paudel et al. (2023) – Prediction of compressive strength using ML[1].	Ensemble (AdaBoost, Random Forest, XGBoost) and non-ensemble models (MLP, Linear Regression)	Limited generalization across datasets and lack of real-time validation
Bahdan Zviazhynski et al. (2025) – Predicting concrete strength using ML and fresh-state measurements [2].	Random Forest with just-in-time learning and uncertainty-based optimization	Limited dataset size and lack of generalization across diverse concrete types
Beskopylny et al. (2022) Concrete strength prediction using ML methods [3].	CatBoost, KNN, SVR applied on experimental datasets	Limited generalization across different construction conditions
Liu (2022) High-performance concrete strength prediction using ML [4].	XGBoost, Random Forest, SVR applied on experimental HPC dataset	Limited dataset size and lack of real-world validation
Batra et al. (2023) Survey on concrete strength prediction using ML [5].	Review of ANN, Random Forest, LSTM, and fuzzy logic techniques	Lack of experimental validation and performance comparison.
Satyanarayana et al. (2024) Prediction of compressive strength of blended concrete [6].	Ensemble ML models (Random Forest, AdaBoost) with Genetic Algorithm optimization	Limited comparison with other deep learning models and real-time validation
Sengupta et al. (2026) State-of-the-art review on CCS prediction [7].	Review of ML models: SVR, Decision Trees, Random Forest, Gradient Boosting, XGBoost, Deep Learning, Meta-heuristic models	Lack of standardized datasets and limited real-time implementation
Elshaarawy et al. (2023) ML and GUI for CS prediction [8].	Ensemble ML models (CatBoost), regression, neural networks, fuzzy inference, SHAP analysis, GUI development	Limited generalization across diverse datasets and real-world construction scenarios
Omotayo et al. (2024) – Assessment of ML methods for concrete compressive strength prediction [9].	Linear Regression, Random Forest, Decision Trees, Gradient Boost, Support Vector Machine, CatBoost; evaluated using MSE, RMSE, MAE, R2*R*2	Limited to concrete without pozzolanic content; does not explore model generalizability to pozzolanic mixes or real-time field applications
Mahajan & Bhagat (2022) ML prediction of fly ash concrete strength [10].	ANN, Decision Trees, Bagging ensemble, statistical validation (RMSE, R^2, k-fold)	Limited use of deep learning and optimization techniques; lack of large-scale datasets
Selvi et al. (2025) Predicting concrete strength using regression-based machine learning techniques [11].	Random Forest, boosting variants (including CatBoost); evaluated using R2*R*2, RMSE, MAE; SHAP analysis for feature importance	Does not compare with non-tree-based algorithms such as SVM or neural networks; limited to single dataset without external validation

Jiang, Li, & Zhou (2025) Compressive strength prediction of fly ash concrete using machine learning techniques [12].	Extreme learning machine, random forest, SVR, grid search-optimized SVR; evaluated using five indices; feature importance analysis	Limited to fly ash concrete with relatively small dataset (270 groups); does not explore deep learning approaches or generalization to other supplementary cementations materials.

CONCLUSION

A synthesis of twelve selected studies on machine learning applications for concrete compressive strength prediction is presented in the table. The employed methodologies fall into several categories: ensemble learning methods such as AdaBoost, Random Forest, XGBoost, and CatBoost; non-ensemble techniques including MLP, SVR, KNN, and Linear Regression; and optimization strategies like genetic algorithms and grid search. Additionally, some studies integrate SHAP-based interpretability analysis and graphical user interface (GUI) development. Across the literature, recurring research gaps emerge, including limited and heterogeneous datasets, insufficient external validation and real-time field testing, inadequate generalization across different concrete mixtures (for instance, non-pozzolanic versus SCM-based formulations), limited comparison with deep learning or hybrid architectures, and a lack of standardized data benchmarks. Collectively, these shortcomings impede the creation of robust and practically deployable predictive frameworks.

REFERENCES

1. Paudel, S., Pudasaini, A., Shrestha, R. K., & Kharel, E. (2023). Compressive strength of concrete material using machine learning techniques. *Cleaner Engineering and Technology*. Elsevier.
2. Zviazhynski, B., White, C., Lees, J. M., & Conduit, G. J. (2025). Predicting concrete strength using machine learning and fresh-state measurements. *Data-Centric Engineering*, *6*, 37.
3. Beskopylny, A., Stel'makh, S., Shcherban, E. I., Mailyan, L., Meshkin, B., Razveva, I., Cherniutin, A., & Beskopylny, N. (2022). Concrete strength prediction using machine learning methods CatBoost, K-nearest neighbors, support vector regression. *Applied Sciences*, *12*(11), 10664
4. Liu, Y. (2022). High-performance concrete strength prediction based on machine learning. *Computational Intelligence and Neuroscience*, *2022*, 5802217
5. Batra, I., D, R., H. D., & Kumar, V. (2023). Concrete strength prediction using machine learning algorithms: Survey. *International Journal of Research in Engineering and Science (IJRES)*, *11*(1), 122–127.
6. Satyanarayana, G. V. V., Kumar, C. V., Karthikeyan, R. M., Punitha, A., Ivanovich, N., Joshi, A., & Hussein, L. (2024). Machine learning prediction and optimization of compressive strength for blended concrete by applying ANN and genetic algorithm. *Cogent Engineering*, *11*(1), 2376914.
7. Sengupta, H., Patra, O., Roy, J. S., & Chakraborty, A. (2026). Concrete compressive strength prediction using machine learning: A state-of-the-art review. *International Journal of Engineering Research & Technology (IJERT)*, *15*(1).
8. Elshaarawy, M. K., Alsaadawi, M. A., & Hamed, A. K. (2023). Machine learning and interactive GUI for concrete compressive strength prediction. Scientific Reports.
9. Omotayo, O. O., Arum, C., & Ikumapayi, C. M. (2024). Assessment of machine learning methods for concrete compressive strength prediction. Journal of Soft Computing in Civil Engineering, *8*(4), 116–140.
10. Mahajan, L., & Bhagat, S. (2022). Machine learning approaches for predicting compressive strength of concrete with fly ash admixture. *Research on Engineering Structures & Materials*.
11. Selvi, S., Murthy, S. E., Govindan, B., Rani, D. L., & Suresh, M. (2025). Predicting concrete strength using regression-based machine learning techniques. Journal of Soft Computing in Civil Engineering, *9*(4), 1–15.
12. Jiang, Y., Li, H., & Zhou, Y. (2025). Compressive strength prediction of fly ash concrete using machine learning techniques. *Buildings*, *15*(2), 1–15.