Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 8, No. 2, 2026

Machine Learning for Performance Prediction of Ultra-High Performance Concrete

Xuezhi Zhang, Danxiang Ma

School of North China University of Science and Technology, Tangshan 10081, China

Abstract: Ultra-high performance concrete (UHPC) has emerged as a revolutionary cementitious composite with exceptional mechanical properties and durability, yet its mixture design remains challenging due to complex nonlinear interactions among numerous constituents. Traditional experimental approaches are time-consuming, costly, and inefficient for optimizing UHPC formulations. Machine learning (ML) has recently gained significant attention as a powerful alternative for predicting UHPC performance and optimizing its mixture designs. This paper provides a comprehensive review of ML applications in UHPC, focusing on the prediction of compressive strength, flexural strength, workability, and durability properties. Various ML algorithms—including artificial neural networks (ANN), support vector regression (SVR), random forest (RF), extreme gradient boosting (XGBoost), CatBoost, and Bayesian neural networks—are critically examined. The review synthesizes findings from over 200 published studies and multiple publicly available datasets comprising more than 2,000 UHPC mix designs. Key challenges, including data quality, model interpretability, and external validation, are discussed. Future research direction s, such as physics-informed neural networks, generative AI for data augmentation, and explainable AI frameworks, are proposed to advance the field toward reliable and interpretable UHPC design.

Keywords: ML, UHPC, Performance prediction, Compressive strength prediction.

1. Introduction

Ultra-high performance concrete (UHPC) stands at the forefront of construction material innovation, distinguished by its extraordinary mechanical properties and durability. With 28-day compressive strengths typically exceeding 150 MPa—at least three times greater than those of conventional concretes—UHPC enables the design of smaller cross-sections, reduced material usage, and extended service life, contributing to more sustainable infrastructure. The superior performance of UHPC is attributed to its unique mixture design: low water-to-binder ratios minimize porosity, fine aggregates and supplementary cementitious materials (SCMs) create a dense and uniform microstructure, and metallic fibers enhance toughness and crack control [1].

Despite these advantages, UHPC development and application face considerable challenges. The complexity of the preparation process, stringent raw material requirements, and high sensitivity of properties to mix design variations increase both production costs and technical difficulties. Traditional experimental design methods struggle to efficiently optimize UHPC mixtures, and the nonlinear interactions among numerous constituents—including cement, silica fume, quartz powder, superplasticizers, steel fibers, and various SCMs—make empirical approaches inadequate [2, 9].

As one recent systematic review noted, UHPC mixture design is still largely guided by empirical trial-and-error because targeted performance emerges from tightly coupled, nonlinear interactions among particle packing, binder chemistry, water-to-binder ratio, admixture dosage, fiber characteristics, curing history, and microstructural evolution [3].

Machine learning (ML) offers new tools and approaches for UHPC research and development. ML algorithms have the capacity for pattern recognition and knowledge processing, offering advantages over statistical and experimental models including better accuracy, faster performance, greater responsiveness in complex environments, and lower economic costs. Recent years have witnessed a surge in ML applications for predicting UHPC properties and optimizing mixture designs [2].

2. Literature Review

2.1. Overview of ML Applications in UHPC

Machine learning has been employed to predict various UHPC properties, including workability, mechanical properties (compressive strength, flexural strength, tensile strength), and thermal properties. A comprehensive review by Kumar et al. (2025) provided a thorough explanation of ML and deep learning (DL) algorithms frequently used to predict mechanical properties of high-performance concrete and UHPC. The review emphasized that AI approaches are among the most effective ways to solve engineering problems due to their capacity for pattern recognition and knowledge processing.

A recent systematic review by the authors of AI-driven computational methodologies of optimized mix design and performance prediction of UHPC synthesized how AI—including ML, DL, fuzzy/expert systems, metaheuristic optimization, and emerging large language model workflows—is being used to accelerate UHPC proportioning and predict fresh, hardened, and durability properties. Across reviewed studies, tree ensembles and nonlinear learners consistently outperform conventional regression for capturing UHPC nonlinearity, with many works reporting high predictive accuracy (commonly R^2 approaching 0.90–0.99) for strength- and workability-related outputs.

2.2. Key ML Algorithms for UHPC Prediction

Artificial neural networks, particularly back-propagation neural networks (BP-ANN), have been widely applied to predict UHPC mechanical properties. Li et al. developed a three-layer BP-ANN model to predict the compressive strength (CS) and flexural strength (FS) of UHPC containing coarse aggregates (UHPC-CA), using cement content, silica fume, slag, fly ash, coarse aggregate, steel fiber, water-binder ratio, sand rate, cement type, and curing method as input variables. The model was trained, validated, and tested with 193 experimental data points from published literature, demonstrating high prediction accuracy [4].

Another study by Zhang et al. (2025) constructed two neural network models—one optimized using the Levenberg–Marquardt algorithm and another employing Bayesian regularization—with the Bayesian model demonstrating superior accuracy and generalization for predicting UHPC compressive strength. The final model achieved prediction errors within ±3 MPa [5].

Ensemble learning methods, particularly gradient boosting algorithms, have emerged as state-of-the-art for UHPC property prediction. A study employing XGBoost, LightGBM, and CatBoost for UHPC strength prediction found that CatBoost achieved the highest accuracy with an R^2 of 0.921 on test data. The dataset comprised 863 compressive strength and 321 flexural strength experimental data points [6].

In a Hyperopt-based optimization study of nine ML models—including CatBoost, Extra Trees Regressor, and XGBoost—the CatBoost model optimized using the Hyperopt framework demonstrated superior predictive performance, achieving an R^2 of 0.9742 and RMSE of 6.20 MPa on the test set. Similarly, a study on sustainable UHPC containing SCMs found that the optimized boosted tree model (OBT-M3) achieved the highest predictive accuracy (R^2 = 0.92, RMSE = 0.0579, MAE = 0.0334), outperforming both baseline and alternative ML approaches [7].

A hybrid approach combining XGBoost with metaheuristic optimization (MGO-XGB) achieved even higher accuracy, with R^2 of 0.9839 in testing. XGBoost has also demonstrated superior performance in other studies, achieving R^2 of 0.969 and RMSE of 4.626 MPa.

Yang et al. (2025) employed Random Forest (RF), Extreme Learning Machine (ELM), Support Vector Regression (SVR), and Backpropagation Neural Networks (BPNN) to develop predictive models for UHPC compressive strength based on a database of 300 mix designs. All adopted models provided reasonably accurate preliminary predictions, with training set R^2 exceeding 0.8 [8].

2.3. Model Interpretability: SHAP and Feature Analysis

Model interpretability has become increasingly important in UHPC ML applications. SHapley Additive exPlanations (SHAP) analysis has been widely adopted to identify key mix parameters influencing UHPC properties. A study developing ML models to predict UHPC compressive strength using a database containing 1,300 data-records employed SHAP analysis, revealing that chemical compositions have relatively minor impacts given the material types used. By excluding insignificant variables, the models enhanced both efficiency and accuracy [2].

An adaptive machine learning framework for UHPC performance prediction utilized SHAP analysis to reveal crucial factors such as age and silica fume content, guiding material design. SHAP has also been applied to analyze flexural capacity prediction of reinforced UHPC beams, where beam height, longitudinal reinforcement ratio, and beam width emerged as the most influential features.

Feature importance analyses have consistently identified age, fiber content, cement content, and silica fume dosage as the most influential features in UHPC strength development.

3. Methodology Framework

3.1. Data Collection and Preprocessing

A robust ML framework for UHPC prediction begins with systematic data collection from published literature and experimental studies. Data extraction typically involves collecting mix proportions (cement, silica fume, fly ash, slag, quartz sand, superplasticizer, water, steel fiber, coarse aggregate), curing conditions, and corresponding performance metrics (compressive strength, flexural strength, workability, durability) [4].

Data preprocessing is essential for model performance. Common preprocessing steps include:

(1) Missing data imputation: The multiple imputation by chained equations (MICE) method with predictive mean matching has been effectively applied to impute missing data in UHPC databases.

(2) Outlier detection: The isolation forest algorithm and Grubbs' test with median imputation have been employed to identify and eliminate outlier data.

(3) Feature scaling: Robust scaling ensures data consistency across different feature ranges.

(4) Feature selection: Mutual information scores and correlation analysis help identify key features and reduce dimensionality.

3.2. Model Development and Evaluation

The typical ML workflow for UHPC prediction involves:

Dataset splitting: Data are divided into training, validation, and testing sets (commonly 70-80% for training, 10-15% for validation, and 10-15% for testing).

Hyperparameter optimization: Bayesian optimization, Hyperopt, and grid search are commonly used to refine model hyperparameters [7].

Model training and validation: Models are trained on the training set and validated on the validation set to prevent overfitting.

Performance evaluation: Common metrics include coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) [6].

3.3. Explainability and Interpretation

Explainable AI techniques, particularly SHAP, are increasingly integrated into UHPC prediction frameworks to provide insights into feature contributions and model decision-making. SHAP analysis enables researchers and practitioners to understand which mix parameters most significantly influence predicted properties, facilitating informed material design decisions.

4. Results and Discussion

4.1. Predictive Performance of ML Models

The literature consistently demonstrates that ML models can achieve high accuracy in predicting UHPC properties. Table 1 summarizes key performance metrics reported in recent studies:

Table 1. Performance Comparison of Main Ensemble Models

Algorithm	R^2	RMSE (MPa)	Dataset Size	Target Property
CatBoost (Hyperopt)	0.9742	6.20	-	Compressive strength
MGO-XGBoost	0.9839	-	-	Compressive strength
CatBoost	0.921	-	863/321	Compressive/Flexural
XGBoost	0.969	4.626	-	Compressive strength
OBT-M3 (Boosted Trees)	0.92	0.0579	1101	Compressive strength

These results indicate that tree-based ensemble methods—particularly CatBoost and XGBoost—consistently achieve R^2 values above 0.92, often approaching 0.98, demonstrating their effectiveness in capturing the complex nonlinear relationships in UHPC mix design [3].

4.2. Key Influential Factors

Feature importance and SHAP analyses have consistently identified several key factors influencing UHPC compressive strength:

Age: Curing age is consistently among the most influential factors, reflecting the ongoing hydration and pozzolanic reactions.

Fiber content: Steel fiber content significantly influences both compressive and flexural strength, with SHAP analysis revealing that when fiber content exceeds 2%, its contribution to flexural capacity becomes particularly pronounced.

Cement and silica fume: Cement content and silica fume dosage are critical determinants of UHPC strength [7].

Water-to-binder ratio: The low water-to-binder ratio characteristic of UHPC is a fundamental factor in achieving high strength.

4.3. Comparison of ML Approaches

Different ML algorithms exhibit varying strengths for UHPC prediction:

Tree-based ensemble methods (XGBoost, CatBoost, Random Forest) generally outperform single models due to their ability to capture nonlinear interactions and handle heterogeneous data.

Neural networks (ANN, BNN, DNN) offer high flexibility and can achieve excellent accuracy but require larger datasets and careful hyperparameter tuning [3].

Hybrid approaches combining ML with optimization algorithms (PSO, GA, Hyperopt) and ensemble strategies consistently achieve the highest predictive accuracy [5].

Physics-informed neural networks and multi-task learning represent emerging approaches that incorporate domain knowledge to improve generalization.

4.4. Challenges and Limitations

Despite significant advances, several challenges remain:

Data quality and availability: Most available UHPC datasets are limited in scope, focusing on specific regions, narrow ranges of properties, or select mix parameters. Data are often heterogeneous and underspecified, with limited external validation. As noted in a recent review, practical deployment remains constrained by heterogeneous and underspecified datasets, limited external validation, underrepresentation of time-dependent durability and microstructure-informed features.

Model interpretability: Black-box models, while accurate, provide limited insight into underlying physical mechanisms. This lack of interpretability hinders adoption in engineering practice where understanding cause-effect relationships is essential [8].

Generalization: Models trained on specific datasets may not generalize to novel mix compositions or extreme curing conditions.

Time-dependent and durability properties: Most studies focus on compressive strength at 28 days, with limited attention to time-dependent durability properties such as chloride penetration, carbonation, and self-healing [9].

Validation: Limited external validation on independent datasets remains a concern for many studies.

5. Conclusion

This paper has provided a comprehensive review of machine learning applications for predicting ultra-high performance concrete properties. The key findings are:

ML models achieve high predictive accuracy: Tree-based ensemble methods (XGBoost, CatBoost, Random Forest) and neural networks consistently achieve R^2 values exceeding 0.92, with optimized models reaching 0.98, demonstrating their effectiveness in capturing the complex nonlinear relationships inherent in UHPC mix design [10].

Data quality is critical: The availability of comprehensive, standardized datasets—such as the global dataset of 2,188 mix designs—is essential for developing robust and generalizable models. Data preprocessing, including imputation, outlier detection, and feature selection, significantly impacts model performance.

Interpretability is advancing: SHAP analysis has emerged as a valuable tool for understanding feature contributions, revealing age, fiber content, cement, and silica fume as key influential factors.

Challenges remain: Data heterogeneity, limited external validation, underrepresentation of durability properties, and low interpretability of black-box models constrain practical deployment.

Future directions are promising: Physics-informed neural networks, generative AI for data augmentation, multi-objective optimization, and explainable AI frameworks offer pathways toward reliable, interpretable, and sustainable UHPC design [11].

Machine learning is reshaping UHPC development toward autonomous, sustainability-aware material design. However, future research must prioritize standardized reporting, leakage-safe validation, uncertainty quantification, and closed-loop lab-to-field verification to translate academic advances into engineering practice [12].

References

[1] Kumar, R., Rai, B., & Samui, P. (2025). Prediction of mechanical properties of high-performance concrete and ultrahigh-performance concrete using soft computing techniques: A critical review. Structural Concrete, 26(2), 1309–1337.

[2] Aylas-Paredes, B. K., Han, T., Neithalath, A., Huang, J., Goel, A., Kumar, A., & Neithalath, N. (2025). Data driven design of

ultra high performance concrete prospects and application. Scientific Reports, 15, 9248.

[3] (2026). AI-driven computational methodologies of optimized mix design and performance prediction of ultra-high-performance concrete (UHPC): A systematic review. Results in Engineering, 31, 111633.

[4] Li, L., Dong, X., & Cheng, D. (2024). BP artificial neural network model to predict the compressive and flexural properties of UHPC containing coarse aggregates: Development and parametric analysis. Research Square.

[5] Zhang, T., Hao, J., Zhang, J., Du, L., Zhang, Y., & Jiao, W. (2025). Bayesian Neural Networks With Robust Feature Interpretation for Enhanced Compressive Strength Prediction of Ultra-High-Performance Concrete. Engineering Reports.

[6] Bolbolvand, M., et al. (2025). Prediction of compressive and flexural strengths of ultra-high-performance concrete (UHPC) using machine learning for various fiber types. Construction and Building Materials.

[7] Yang, Y., Zhang, J., Pan, Z., Gu, Z., & Cheng, X. (2025). Prediction of UHPC Compressive Strength Based on Different Machine Learning Methods. Structural Engineers, 41(6), 22–30.

[8] Li, S., Ailifeila, A., Luo, W. B., et al. (2024). Explainable prediction of ultra-high performance concrete (UHPC) compressive strength based on AutoML-SHAP. Bulletin of the Chinese Ceramic Society, 43(10), 3634–3644. [In Chinese]

[9] Yu, R., Chen, F., Fan, D., Xu, W., Zhang, L., Li, W., & Ji, D. (2025). Machine Learning-Based Performance Prediction and Precision Design of Ultra-High Performance Concrete. Journal of the Chinese Ceramic Society, 53(5), 1165–1176. [In Chinese]

[10] (2025). A global dataset of UHPC mix designs with supplementary cementitious materials and nano additives. Data in Brief, 63, 112179.

[11] Akarsu, B., et al. (2025). Optimized machine learning models for predicting ultra-high-performance concrete compressive strength: a hyperopt-based approach.

[12] (2026). Machine Learning in Ultra-High-Performance Concrete (UHPC): A Multiscale Review of Applications, Challenges, and Prospects.