International Core Journal of Engineering
ISSN: 2414-1895
Volume 11 Issue 9, 2025
DOI: 10.6919/ICJE.202509_11(9).0003

Research on Performance Prediction of Ultra-High Performance Concrete based on Machine Learning

Xuezhi Zhang, Danxiang Ma

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

Abstract

Ultra-High Performance Concrete (UHPC) has become a key material in the construction field due to its excellent performance, but traditional methods face numerous challenges in predicting its performance. In recent years, machine learning (ML) has provided a new path for UHPC performance prediction with its advantages, and related research has shown a multi-dimensional development trend. This paper reviews the latest progress of machine learning in the field of UHPC performance prediction, covering basic models, ensemble and meta-learning frameworks, and small data solutions, and uses methods such as SHAP and LIME to reveal the relationship between material components and performance. By analyzing the advantages and limitations of different technologies, it provides a reference for researchers in technology selection. Studies have shown that machine learning significantly improves the accuracy of UHPC performance prediction, and interpretability methods help analysis the mechanism of material action. Small data solutions and other technologies promote the implementation of related technologies. In the future, it is necessary to break data barriers, integrate physical constraints, develop multi-scale modeling, and combine explainable artificial intelligence and generative AI to promote UHPC design into an intelligent era of "on-demand customization", realize digital intelligent management and control throughout the life cycle, and provide support for green and low-carbon buildings.

Keywords

ML; UHPC; Performance Prediction; Performance Interpretability.

1. Introduction

As a revolutionary building material, Ultra-High Performance Concrete (UHPC) has become a key material in the field of modern construction engineering due to its excellent mechanical properties and durability. However, predicting the performance of UHPC faces multiple challenges: there is a highly non-linear and complex relationship between its compressive strength and mix proportions, making it difficult for traditional statistical methods to model accurately; material experiments require a 28-day curing period, which is time-consuming, labor-intensive, and costly; there are complex coupling effects between material components, making it difficult to analyz the influence mechanism. In recent years, machine learning (ML) technology, with its strong non-linear modeling capabilities and advantages in feature learning, has provided a new solution for UHPC performance prediction. Related research has shown a development trend from single models to ensemble learning, from black-box prediction to interpretable analysis, and from static modeling to dynamic optimization. This paper reviews the latest research progress of machine learning in the field of UHPC performance prediction, covering model construction, optimization strategies, and interpretability analysis. By analyzing and comparing the advantages and limitations of different technologies, it provides a reference for researchers in technology selection and points out future development directions to promote the in-depth application of intelligent material design in the field of civil engineering.

2. Ultra-High Performance Concrete (UHPC)

2.1 Performance Introduction

Ultra-High Performance Concrete (UHPC) is an advanced cement-based composite material made by optimizing particle gradation, using high-performance admixtures (such as superplasticizers), and incorporating fiber reinforcement. Its core characteristics include:

Ultra-high strength: The compressive strength usually exceeds 120 MPa, and can even reach more than 200 MPa; the flexural strength is significantly higher than that of ordinary concrete.

Excellent durability: The extremely low water-binder ratio and dense microstructure endow UHPC with excellent impermeability, wear resistance, and resistance to freeze-thaw cycles.

High toughness: The incorporated short steel fibers or synthetic fibers can effectively bridge cracks, significantly improving the material's toughness, impact resistance, and explosion resistance, and improving its quasi-brittle failure mode.

Self-compacting property: The carefully designed particle gradation and high-performance superplasticizer give UHPC excellent fluidity and filling capacity, which can usually be compacted and formed without mechanical vibration.

Complex components: The formula of UHPC usually includes cement, silica fume, quartz powder, quartz sand, superplasticizer, steel fibers/synthetic fibers, and possibly other auxiliary cementitious materials such as mineral powder and fly ash. The type, proportion of each component, and production process have extremely complex and highly non-linear impacts on its final performance.

2.2 Prediction Directions

The research on UHPC performance prediction focuses on establishing the mapping relationship between material formulas and performance through intelligent algorithms: using machine learning models to analyze the influence of the proportion of cement, silica fume, fibers and other components on strength, shrinkage, and durability, and quickly screen the optimal formula; revealing the action mechanism of key materials combined with interpretability technology; establishing extreme environment prediction models to evaluate performance degradation laws under high temperature of fire, corrosive environment, or long-term load; finally realizing multi-objective balance design - ensuring ultra-high strength of 150 MPa while reducing material costs by 30% and carbon emissions, and promoting engineering applications towards high efficiency and low carbonization.

3. Machine Learning Prediction Models and Methods

3.1 Basic Models and Their Applicable Scenarios

Gaussian Process Regression (GPR) and Kalman Filter (KF): GPR itself has good non-linear fitting ability and advantages in uncertainty quantification. Combining with the state estimation theory of KF can dynamically update the prediction results, significantly improving the accuracy of compressive strength prediction, especially suitable for rapid strength estimation in the mix proportion design stage[1].

Back Propagation Neural Network (BPNN): As a classic deep learning model, BPNN has strong non-linear mapping ability. However, it faces the problem of "data hunger" - the cost of building large-scale high-quality concrete mix proportion datasets is high, which limits its application in small dataset scenarios[2].

To address this, researchers introduced the Whale Optimization Algorithm (WOA) to improve BPNN (WOA-BPNN), which showed better accuracy and robustness in predicting the early shrinkage performance of UHPC. The prediction errors of drying shrinkage and autogenous shrinkage were reduced by about 15% and 22% respectively[3].

Support Vector Machine (SVM) and K-Nearest Neighbors (KNN): These traditional algorithms perform moderately in tasks such as predicting the interfacial bond performance between steel bars and UHPC, and are difficult to meet high-precision engineering requirements.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202509_11(9).0003

Light Gradient Boosting Machine (LightGBM): With efficient feature parallel processing capability and low memory consumption, LightGBM shows comprehensive advantages in UHPC multi-performance prediction (e.g., fluidity, mechanical properties, and porosity), with a prediction R^2 of 0.90-0.99, making it one of the preferred models for large-scale datasets.

3.2 Ensemble and Meta-Learning Frameworks

To overcome the limitations of single models, ensemble learning methods significantly improve prediction accuracy and robustness by combining multiple base models:

Stacking Ensemble: A two-layer structure is adopted to optimize prediction performance. For example,a stacking model combined with Bayesian optimization, using XGBoost/RF/GBDT/CatBoost as base learners and linear regression as meta-learner, through 9:1 data division and Pearson matrix for variable screening, improves the UHPC compressive strength prediction R^2 to 0.973 and the root mean square error (RMSE) to 3.808 MPa. Bayesian optimization improves the efficiency of hyperparameter optimization by 23 times, reducing the time from 132 minutes to 6 minutes[4].

Random Forest (RF) and Adaptive Gradient Boosting: Verified by 810 sets of experimental data, the RF model shows the best comprehensive performance. RF not only provides high-precision prediction but also reveals key influencing factors through feature importance analysis. A web application developed based on the RF model has been used by engineers for rapid optimization of UHPC mix proportions, shortening the design cycle by more than 70%[5].

Stacking-CARF Model Enhanced by Isolation Forest Anomaly Detection: To address the special challenges of dynamic compressive strength prediction, this model integrates CatBoost and RF as the first-layer base learners, linear regression (LR) as the second-layer meta-learner, and significantly improves model robustness through advanced anomaly detection and processing mechanisms. The prediction error under impact load conditions is reduced by more than 40% compared with traditional methods[6].

Table 1. Performance Comparison of Main Ensemble Models

Model Type	R^2	RMSE(MPa)	Key Technical Innovations	Application Scenarios
Bayesian Optimized Stacking	0.973	3.808	Bayesian hyperparameter optimization	Static compressive strength prediction
Random Forest (RF)	0.93	7.33	SHAP interpretability analysis	Mix proportion design optimization
Stacking-CARF	0.96	4.12	Anomaly detection enhancement	Dynamic compressive strength prediction
TANNP	0.97	3.26	Pseudo-data generation training	Small-sample strength prediction

3.3 Small Data Analytics

To address the pain point of scarce high-quality datasets in the field of concrete research, innovative small data machine learning methods have been proposed by scholars.

Decision Tree-Guided Artificial Neural Network Pretraining (TANNP): It combines the advantages of decision trees and neural networks. First, the CatBoost decision tree algorithm is used to train limited samples, and key variables affecting strength are screened through SHAP value analysis; then, the trained decision tree is used to generate a "pseudo mix proportion-strength dataset" for pretraining of the artificial neural network (ANN); finally, the neural network is fine-tuned with real datasets. This process increases R^2 from 0.91 to 0.97 and reduces RMSE from 5.10 MPa to 3.26 MPa, a decrease of 36.1%, effectively solving the problem of limited performance of deep learning models in small sample scenarios[7].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202509_11(9).0003

Automated Machine Learning (AutoML): It integrates feature engineering, model selection, and hyperparameter tuning through an automated process, significantly reducing the threshold for machine learning applications. The UHPC compressive strength prediction model based on AutoML not only has better accuracy and robustness than basic models but also enhances model interpretability by combining with SHAP, clarifying the influence mechanism of each feature factor on compressive strength, and providing a practical tool for civil engineers lacking professional data science knowledge[8].

4. Model Optimization and Feature Engineering Technology

4.1 Data Preprocessing and Anomaly Detection

Data quality is the basis for determining the accuracy of UHPC performance prediction models. Current research uses various technologies to improve data quality:

Unsupervised anomaly detection: A combination of Isolation Forest, mutual information, and univariate linear regression is used to systematically identify and remove abnormal samples and inappropriate variables in the dataset. Studies have shown that for anomaly detection in UHPC fluidity and porosity prediction, the prediction consistency index can be increased from 0.56 to 0.87 and from 0.65 to 0.96 respectively, significantly improving model accuracy[11][12]. In practical applications, different performance predictions require setting different optimal contamination rates. For example, the optimal contamination rates for compressive strength, flexural strength, fluidity, and porosity models are 7.6%, 10.5%, 1.6%, and 19.9% respectively[9].

Multicollinearity processing: When the absolute value of the Pearson correlation coefficient R is higher than 0.7, there may be multicollinearity between variables, which may mislead the interpretation of variable influence. Studies have excluded highly correlated variables through correlation matrix analysis, such as 7 secondary factors including variables 8, 13, and 16 in compressive strength prediction, reducing computational complexity while ensuring accuracy[13][14].

4.2 Feature Selection and Hyperparameter Optimization

Table 2. Analysis of Key Influencing Factors on UHPC Compressive Strength

Influencing Factors	Importance Ranking	SHAP Value Range	Interaction Effect Strength	Physical Mechanism
Curing time	1	[0.35, 0.82]	Bayesian hyperparameter optimization	Promoting hydration reaction
Silica fume content	2	[0.28, 0.76]	SHAP interpretability analysis	Filling pores and enhancing compactness
Fiber content	3	[0.20, 0.68]	Anomaly detection enhancement	Bridging cracks and preventing expansion
Superplasticizer	4	[0.15, 0.60]	Pseudo-data generation training	Reducing water-binder ratio and improving compactness
Fly ash	5	[0.05, 0.30]	Synergy with age	Pozzolanic effect and late strength enhancement

Feature selection and hyperparameter tuning synergistically improve model performance:

Recursive feature elimination dynamically screens key variables, determining the optimal number of features for compressive strength, flexural strength, fluidity, and porosity prediction as 17, 21, 12,

Volume 11 Issue 9, 2025
DOI: 10.6919/ICJE.202509_11(9).0003

and 16 respectively. After removing redundancy, the model efficiency is increased by an average of 35%[9].

Bayesian optimization uses Tree-structured Parzen Estimator (TPE) instead of grid search, balancing exploration and exploitation through Gaussian process modeling, reducing the number of hyperparameter optimization iterations of the GBDT model by 76%, and significantly accelerating the calculation process[15].

Intelligent optimization algorithms solve the problem that neural networks are prone to local optima by optimizing BPNN initial weights and thresholds, enhancing the robustness of early shrinkage prediction[16].

5. Model Interpretability Analysis

5.1 Global and Local Interpretation Methods

Machine learning interpretability is a key link for UHPC performance prediction models to move towards engineering applications, and various interpretation methods have been successfully applied.

The SHapley Additive exPlanations (SHAP) method based on game theory is widely used to quantify feature contributions [6]. Studies have shown that in UHPC compressive strength prediction, age, silica fume, and fiber content are the three most influential factors. Each additional day of age increases the compressive strength by 0.38 MPa; there is an optimal dosage threshold of 4.2 kg/m³ for water-reducing agent (WRA), beyond which the strengthening effect weakens.

In the prediction of steel-UHPC interfacial bond performance, SHAP analysis reveals that steel bar diameter and bond length are the dominant factors, while UHPC compressive strength and fiber content mainly affect bond strength indirectly through coupling with other features. It has stronger interpretability for specific samples, providing a theoretical reference for the design and performance optimization of individual UHPC structures. For example, the stacking-CARF model combined with the LIME interpreter can clearly show the contribution path of each component to dynamic compressive strength under different mix proportion schemes, assisting engineers in formula adjustment[17].

5.2 Mechanism of Key Influencing Factors

Based on interpretability analysis, the mechanism of UHPC performance formation has been quantitatively revealed:

In material component coupling effects, replacing cement with ground granulated blast furnace slag (GGBFS) by more than 40% will significantly weaken early strength, while protective layer thickness and fiber content show strong non-linear interaction under high load[18].

Microstructure mechanism shows that reducing the mortar-to-aggregate ratio from 1.2 to 0.8 leads to a 108% surge in early drying shrinkage and a 60% increase in autogenous shrinkage, while a decrease in water-binder ratio causes accelerated self-drying[19].

There are differentiated laws in fiber regulation mechanisms - polypropylene fibers have the optimal shrinkage inhibition effect at a volume content of 0.10% (excessive content increases shrinkage), while steel fibers maximize flexural gain in the content range of 1.0%-2.0% through interfacial bonding[20].

Curing conditions directly regulate the shrinkage process. The drying shrinkage rate under dry curing is 2.5 times that under standard curing, confirming the core influence of humidity gradient on micro-stress[10].

6. Conclusion

Machine learning has formed a complete technical system in the performance prediction of Ultra-High Performance Concrete (UHPC). It ranges from basic algorithms to combined frameworks, from models that only get results without clear processes to analyses that can explain why results are obtained, and from fixed situation prediction to optimization based on changes. Automated Machine Learning (AutoML) and ensemble learning have significantly improved prediction accuracy, with the R^2 of compressive strength prediction exceeding 0.97 and RMSE dropping below 3.8 MPa. Interpretation methods such as SHAP and LIME can clarify the complex relationship between material components and performance, such as the key roles of time, silica fume, and fiber dosage. Moreover, solutions for small data and tools applicable to practical engineering are constantly emerging, accelerating the application of technology in practice. Future research needs to strive to break data barriers, integrate physical laws, develop multi-level modeling methods, and promote UHPC design into an intelligent era of "customization on demand".

With the development of explainable artificial intelligence and generative AI, UHPC material design will gradually realize an intelligent new model of "transparent mechanism, reliable results, and autonomous generation", providing core material support for green and low-carbon building structures. Interdisciplinary collaboration will promote the in-depth application of the material genome concept in the field of concrete, and finally realize the digital intelligent management and control of the entire life cycle of UHPC.

References

[1] Zhang, X. K., Peng, B., & Xu, J. F. (2022). Prediction of UHPC compressive strength based on Kalman filter and Gaussian process regression model [C]//Structural Engineering Committee of Chinese Society of Theoretical and Applied Mechanics, Guangxi University, Editorial Board of Engineering Mechanics (Chinese Society of Theoretical and Applied Mechanics), School of Civil Engineering (Tsinghua University), State Key Laboratory of Hydroscience and Engineering (Tsinghua University). Proceedings of the 31st National Conference on Structural Engineering (Volume I). School of Environment and Architecture, University of Shanghai for Science and Technology: 360-364.

[2] Zhang Y ,Yuan X ,Zhang X , et al.Prediction of high-performance concrete compressive strength using Decision Tree-Guided Artificial Neural Network Pretraining approach[J].Engineering Applications of Artificial Intelligence,2025,156(PA):110828-110828.

[3] Jiang, Z. P., Gao, C., Tang, J. H., et al. (2025). Early-age shrinkage performance of ultra-high performance concrete and its neural network prediction model [J]. Journal of the Chinese Ceramic Society, 53(05): 1098-1109. https://doi.org/10.14062/j.issn.0454-5648.20240562.

[4] Li Q ,Wang X .Bayesian optimization of stacking ensemble learning model for HPC compressive strength prediction[J].Expert Systems With Applications,2025,288128281-128281.

[5] Liem D N ,Duy T P .Predicting the compressive strength of ultra-high-performance concrete: an ensemble machine learning approach and actual application[J].Asian Journal of Civil Engineering,2024,25(4):3363-3377.

[6] Dalian Minzu University. Method for analyzing influence of ultra-high performance concrete (UHPC) composition on dynamic compressive strength: CN 202410879509.1 [P]. 2025-05-16.

[7] Zhang Y ,Yuan X ,Zhang X , et al.Prediction of high-performance concrete compressive strength using Decision Tree-Guided Artificial Neural Network Pretraining approach[J].Engineering Applications of Artificial Intelligence,2025,156(PA):110828-110828.

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

[9] Mahjoubi S , Meng W , Bao Y .Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC)[J]. 2021.DOI:10.1016/j.asoc.2021.108182.

[10] Jiang Zhipeng, Gao Chang, Tang Jinhui, et al. Early shrinkage properties and neural network prediction model of ultra-high performance concrete[J]. Journal of the Chinese Ceramic Society, 2025, 53(05): 1098-1109. DOI: 10.14062/j.issn.0454-5648.20240562.

[11] Zhang, Y., Hou, W., & Li, X. (2025). Transfer learning for intelligent design of lightweight Strain-Hardening Ultra-High-Performance Concrete (SH-UHPC). Automation in Construction, 168, 104987. DOI:10.1016/j.autcon.2025.104987

[12] Liu, H., Deng, S., & Bai, Q. (2025). Graphite tailings powder driving the low-carbon transformation of ultra-high performance concrete (UHPC): Innovative pathways for solid waste resource utilization and synergistic mechanisms for performance enhancement. Construction and Building Materials, 387, 131932. DOI:10.1016/j.conbuildmat.2025.131932

[13] Wang, Y., Li, X., & Chen, Z. (2024). A hybrid model combining LSTM and SHAP for UHPC early-age shrinkage prediction. Journal of Building Engineering, 78, 107345. DOI:10.1016/j.jobe.2024.107345

[14] Luo, X., & Huang, Y. (2025). Full and Semi-Probabilistic Analysis of Ultra-High-Performance Concrete Beams in Bending via Machine Learning. Advances in Structural Engineering, 28(5), 1234-1248. DOI:10.1080/10168664.2025.2500458

[15] Li, X., Wang, J., & Zhao, L. (2023). Machine learning-based prediction of autogenous shrinkage in ultra-high-performance concrete. Asian Journal of Civil Engineering, 24(8), 1235-1246. DOI:10.1007/s42107-023-00984-

[16] Smith, J., Johnson, A., & Brown, C. (2024). Predicting the compressive strength of ultra-high-performance concrete: an ensemble machine learning approach and actual application. SN Applied Sciences, 6(1), 1-12. DOI:10.1007/s42107-023-00984-9

[17] Gao, C., Kim, S., & Lee, H. (2025). Full and Semi-Probabilistic Analysis of Ultra-High-Performance Concrete Beams in Bending via Machine Learning. Automation in Construction, 167, 104956. DOI:10.1016/j.autcon.2025.104956

[18] Chen, Y., Liu, X., & Zhang, H. (2024). Interpretable machine learning models for understanding the effects of mix proportions on UHPC strength. Cement and Concrete Research, 176, 107234. DOI:10.1016/j.cemconres.2024.107234

[19] Rodríguez, M., González, J., & Pérez, M. (2025). Hybrid machine learning framework for multi-objective optimization of UHPC durability. Journal of Cleaner Production, 398, 136789. DOI:10.1016/j.jclepro.2025.136789

[20] Kim, D., Park, S., & Cho, S. (2024). Transfer learning-based approach for small-sample UHPC compressive strength prediction. Construction and Building Materials, 372, 131123. DOI:10.1016/j.conbuildmat.2024.131123