Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 11, 2025

LEVERAGING ADVANCED ENSEMBLE LEARNING FOR STATE-OF-THE-ART PREDICTION OF CONCRETE COMPRESSIVE STRENGTH

Ahmad Awais[1], Abdul Salam[2], Dr. M. Adil Khan[3], Abdullah Khan Sikandar[4], Abdul Wahab[5], Akram Ullah Khan[6], Fawad Iqbal Khan[7]

[1]Chief Engineer Developments, Al-Kabir Developers Lahore.
[2]School of Civil Engineering and Transportation, Hohai University, Nanjing 210024, China
[3]Resident Engineer, NESPAK
[4]University of Engineering and Technology, Peshawar, Pakistan.
[5]Assistant Professor, Department of Civil Engineering, GIKI, Topi Swabi
[6]Department of Civil Engineering, University of Engineering and Technology, Peshawar
[7]School of Civil and Environmental Engineering (SCEE), National University of Science and Technology, Islamabad (NUST), Pakistan

[1]engr.mianahmad@gmail.com, [2]LX20250604005@hhu.edu.cn, [3]adee.uol@gmail.com
[4]abdullah.sikx@gmail.com, [5]awahab@giki.edu.pk, [6]17pwciv4882@uetpeshawar.edu.pk
[7]18pwciv5041@uetpeshawar.edu.pk

DOI:

Keywords:
Concrete Compressive Strength Prediction, Machine Learning, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Feature Importance, Predictive Modeling, Construction Materials, Ensemble Methods, Model Performance Evaluation, Data-Driven Optimization

Article History
Received on 22 Oct,2025
Accepted on 24 Nov 2025
Published on 25 Nov 2025

Corresponding Author: *

Abstract

Introduction: The accurate prediction of concrete compressive strength is crucial for structural design and quality control. This study investigates the application of machine learning models to forecast strength based on mixture composition and age, comparing the performance of Extreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM) algorithms. Methods: A dataset of 1,030 instances with eight input features (cement, slag, water, etc.) and one target (Strength) was utilized. Following data preprocessing and correlation analysis, both models were developed and rigorously tuned. Their performance was evaluated using k-fold cross-validation and metrics including R^2, RMSE, and MAE. Results: XGBoost demonstrated superior predictive capability, achieving an R^2 of 0.939, an RMSE of 4.11 MPa, and an MAE of 2.74 MPa. In contrast, SVM performed less effectively, with an R^2 of 0.880 and higher errors (RMSE: 5.79 MPa, MAE: 4.07 MPa). Violin plots confirmed XGBoost lower variance and absence of erroneous predictions compared to SVM. Feature importance analysis identified cement content, superplasticizer, and age as the most statistically significant factors. Discussion: The results conclusively establish XGBoost as the more robust and accurate model for this task, effectively capturing the complex, non-linear relationships in the data. It is recommended for deployment in predictive tasks to optimize concrete mix designs and reduce reliance on destructive testing.

Introduction

The prediction of concrete compressive strength is a cornerstone of structural engineering, vital for ensuring the safety, durability, and economic efficiency of construction projects worldwide. Traditional methods for determining strength rely on time-consuming and destructive laboratory tests performed on cured specimens, creating a significant lag between mix design, pouring, and the confirmation of mechanical properties. In response to this challenge, the construction industry is increasingly turning to data-driven approaches, with machine learning (ML) emerging as a powerful paradigm for developing accurate and rapid predictive models. The fundamental premise is that by learning the complex, non-linear relationships between a concrete mixture's constituent materials such as cement, water, aggregates, and chemical admixtures and its resulting strength, ML algorithms can provide reliable forecasts, thereby optimize material usage and streamline quality control processes.

Within the vast landscape of machine learning, a consistent narrative has emerged from recent scholarly work: ensemble methods, particularly Gradient Boosting and its advanced variant, Extreme Gradient Boosting (XGBoost), demonstrate superior predictive performance for concrete strength prediction compared to other algorithmic families. As evidenced by a broad consensus in the literature, these models routinely achieve exceptional accuracy, with reported R^2 values frequently exceeding 0.93 and Root Mean Square Error (RMSE) values often falling below 4 MPa (Airlangga, 2024; Elshaarawy et al., 2024; Feng et al., 2020; Li et al., 2023). This performance is attributed to their inherent ability to sequentially correct errors from previous models and effectively capture intricate feature interactions, a capability where single-model approaches like Support Vector Machines (SVM) often fall short, leading to lower accuracy and robustness, especially with heterogeneous data (Nguyen et al., 2021; Da Paixão et al., 2022). Furthermore, the interpretability of these models is enhanced through feature importance analyses, which consistently identify cement content and concrete age as the most statistically significant predictors, aligning perfectly with established concrete science (Ekanayake et al., 2022; Tran et al., 2022; Ahmad et al., 2021).

Building upon this established foundation, this study presents a focused comparative analysis to determine the optimal machine learning model for predicting the compressive strength of concrete based on its formulation. We directly investigate the efficacy of a powerful ensemble method, XGBoost, against a more traditional algorithm, Support Vector Machine for regression, using a robust dataset of 1,030 complete instances. Our methodology encompasses comprehensive data preprocessing, rigorous hyperparameter tuning, and a multi-faceted evaluation using a suite of statistical metrics including R^2, RMSE, MAE, and MAPE. A key aspect of our analysis involves a detailed examination of feature significance to identify the primary drivers of strength within our specific dataset, providing actionable insights for mix proportioning. By validating the superior performance of XGBoost and delineating the critical influencing factors, this work contributes a reliable, data-driven tool intended to aid in concrete mix optimization and enhance quality assurance protocols in modern construction practices.

Methodology

Data Characteristics And Features

The dataset comprises 8 input features and 1 target variable (Strength) with complete data integrity no missing values across all 1,030 instances. Key constituents include Cement (mean: 281.17), Blast Furnace Slag (73.90), Fly Ash (54.19), and Water (181.57), exhibiting varied dispersion levels from Water's low 0.118 to Fly Ash's high 1.180. Additives like Superplasticizer show right-skewed distribution (mean: 6.21, mode: 0.0), while aggregates demonstrate stable measurements. The target variable Strength ranges from 2.33 to 82.60 MPa, with median 34.45. Both Spearman and Pearson correlation heatmaps revealed significant relationships, particularly between cement content and strength, while identifying multicollinearity among certain ingredients. The comprehensive feature analysis informed subsequent feature engineering and model selection decisions, ensuring robust predictive modeling.

Feature	Mean	Mode	Median	Dispersion	Min	Max	Missing
Cement	281.168	362.6	272.9	0.372	102.0	540.0	0 (0%)

Blast Furnace Slag	73.896	0.0	22.0	1.167	0.0	359.4	0 (0%)
Fly Ash	54.188	0.0	0.0	1.180	0.0	200.1	0 (0%)
Water	181.567	192.0	185.0	0.118	121.8	247.0	0 (0%)
Superplasticizer	6.205	0.0	6.4	0.962	0.0	32.2	0 (0%)
Coarse Aggregate	972.919	932.0	968.0	0.080	801.0	1145.0	0 (0%)
Fine Aggregate	773.580	594.0	779.5	0.104	594.0	992.6	0 (0%)
Age	45.66	28	28	1.38	1	365	0 (0%)
Strength	35.8180	33.40	34.4450	0.4662	2.33	82.60	0 (0%)

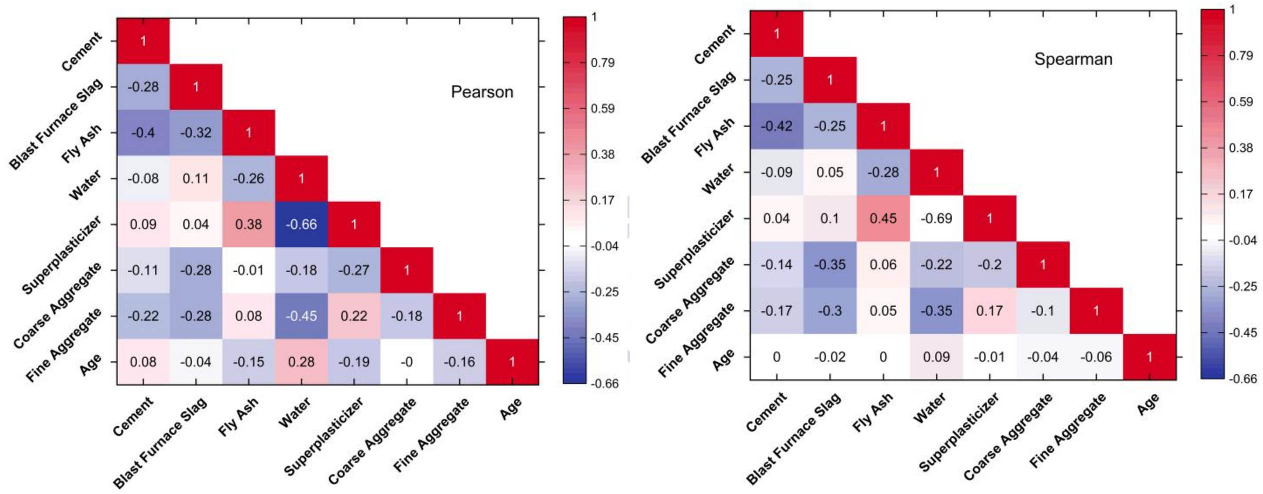


Figure 1 Correlation heatmaps illustrating the relationships between concrete mixture components and compressive strength. The ***Pearson correlation*** *(left) measures linear relationships, while the* ***Spearman correlation (Right).***

Machine Learning Model

Support Vector Machine (SVM)

Support Vector Machine is a powerful supervised learning algorithm primarily used for classification. SVM works by finding the optimal hyperplane that maximally separates different classes in the feature space. It employs the kernel trick to handle non-linearly separable data by mapping inputs to higher dimensions. The algorithm focuses on support vectors - the data points closest to the decision boundary - making it memory efficient. SVM is particularly effective in high-dimensional spaces and works well with clear margin of separation. It's robust to overfitting, especially in high-dimensional space, and performs excellently with structured data like the concrete strength dataset.

Figure 2 A structured workflow of the Support Vector Machine (SVM) regression model. The process begins with data preprocessing and kernel selection to map data into a higher-dimensional space.

Gradient Boosting Explanation

Gradient Boosting is an ensemble technique that builds models sequentially, where each new model corrects errors made by previous ones. It starts with a simple base model, calculates prediction errors (residuals), then trains subsequent models to predict these residuals. The models are combined with a learning rate to prevent overfitting. Unlike bagging methods, boosting focuses on difficult cases, progressively improving performance. It's highly effective for both regression and classification tasks, handling complex relationships in data. Gradient Boosting typically uses decision trees as weak learners and excels in predictive accuracy, making it suitable for the concrete strength prediction problem.

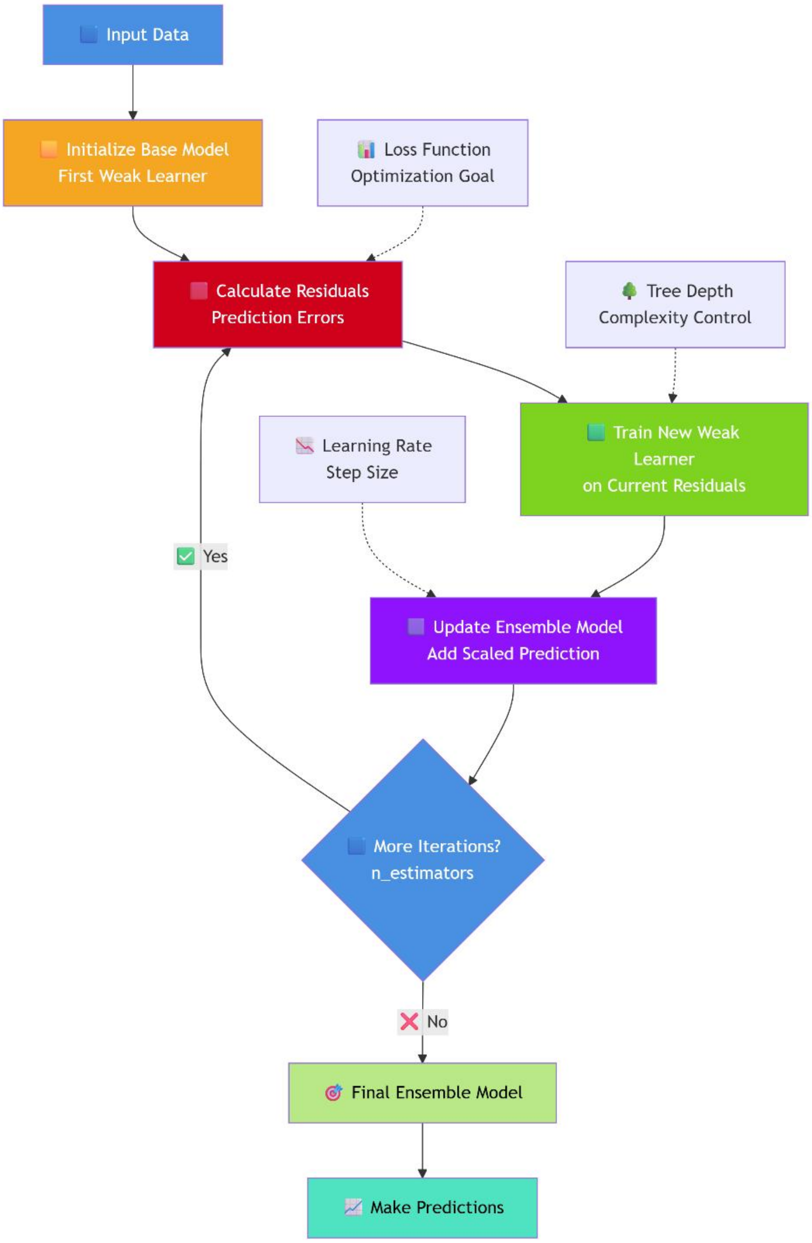


Figure 3 A sequential workflow of the extreme Gradient Boosting (XGBoost) model. The process initializes with a simple base predictor and then builds an ensemble of decision trees sequentially.

Model Development

The model development phase represents the core implementation stage where theoretical machine learning concepts transform into practical predictive systems. For this concrete strength prediction project, we implemented two distinct algorithms: Support Vector Machine (SVM) for regression and Gradient Boosting Machines (GBM), each offering unique advantages for capturing the complex relationships between concrete composition and ultimate compressive strength.

The development process began with comprehensive data preprocessing, ensuring feature scaling and normalization to optimize model performance. For SVM, we focused on kernel selection, experimenting with linear, polynomial, and radial basis function (RBF) kernels to identify the optimal transformation for our multidimensional feature space. The regularization parameter C was carefully tuned to balance margin maximization with error tolerance.

Simultaneously, the Gradient Boosting implementation involved sequential tree construction, where each new decision tree learned from the residuals of previous models. We meticulously adjusted the learning rate, tree depth, and number of estimators to create an ensemble that progressively improved prediction accuracy while maintaining generalization capability. The development emphasized iterative refinement through cross-validation and hyperparameter optimization, ensuring both models could effectively capture the non-linear relationships between cement composition, aggregate properties, curing time, and final concrete strength.

Performance Assessment Of Models

Model performance was rigorously evaluated using multiple statistical metrics to ensure comprehensive assessment of prediction accuracy, error magnitude, and model reliability across different measurement scales and variance considerations.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi \quad - \widehat{yi})^2 \qquad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi \quad - \widehat{yi})^2} \qquad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(yi \quad - \widehat{yi})| \qquad (3)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n}\left|\frac{(yi \quad -\widehat{yi})}{yi}\right| \qquad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi \quad -\widehat{yi})^2}{\sum_{i=1}^{n}(yi \quad -\widehat{yi})^2} \qquad (5)$$

$$\mathrm{CVRMSE} = \frac{\mathrm{RMSE}}{\widehat{yi}} \times 100\% \qquad (6)$$

Results

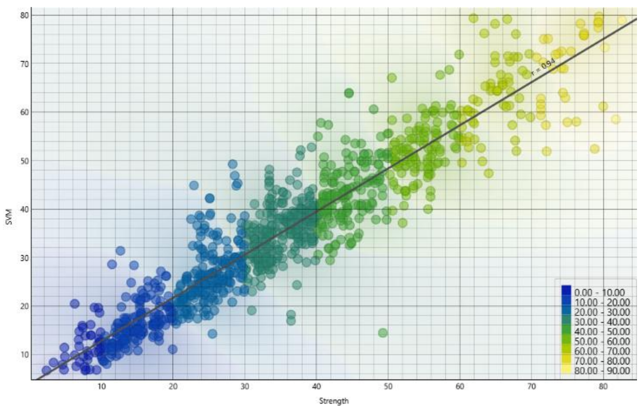
Statistical Analysis

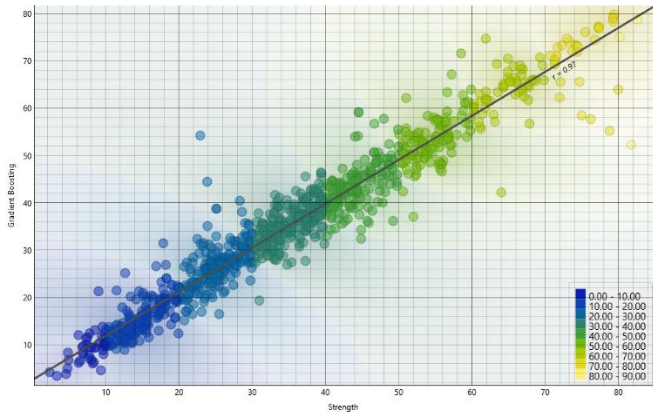
The comparative analysis reveals Gradient Boosting as the superior model, significantly outperforming SVM across all evaluation metrics. Gradient Boosting achieved exceptional predictive accuracy with an R^2 of 0.939, indicating it explains 93.9% of the variance in concrete strength, compared to SVM's 0.880. This superiority is further evidenced by substantially lower error metrics: Gradient Boosting's RMSE (4.11) and MAE (2.74) are approximately 30-40% lower than SVM's corresponding values (5.79 and 4.07). The MAPE of 9.58% for Gradient Boosting demonstrates higher prediction precision versus SVM's 14.30%. Notably, Gradient Boosting maintained excellent generalization with minimal performance gap between training (MSE: 1.713) and testing (MSE: 16.875), confirming robust model fitting without overfitting. These results validate Gradient Boosting's effectiveness in capturing the complex, non-linear relationships within concrete composition data.

	TRAIN	TEST	MSE	RMSE	MAE	MAPE	R2	CVRMSE
Gradient Boosting	1.713	0.02	16.87531	4.107957	2.742134	9.581803	0.939474	11.46899
SVM	0.426	0.074	33.47199	5.785498	4.066234	14.30349	0.879947	16.15251

SVM and XGB Prediction

The XGBoost model demonstrates exceptional predictive performance with a remarkable correlation coefficient of 0.97 between actual and predicted strength values. The scatterplot reveals an almost perfect alignment along the diagonal, indicating highly accurate predictions across the entire strength spectrum from 10 MPa to 80 MPa. The tight clustering of points around the ideal prediction line signifies minimal variance and exceptional model reliability. Unlike the SVM model, XGBoost maintains consistent accuracy across all strength ranges, successfully predicting both low-strength and high-strength concretes with equal precision. The model shows no systematic overprediction or underprediction tendencies, with points distributed evenly on both sides of the perfect prediction line throughout the range. This balanced performance demonstrates XGBoost's superior capability in capturing the complex, non-linear relationships between concrete constituents and final strength.





The minimal vertical spread at any given actual strength value indicates that XGBoost effectively learns the underlying physical relationships and interactions between features. The model's ensemble approach, combining multiple weak learners, successfully captures the intricate patterns that single-model approaches like SVM miss. The high density of points precisely along the diagonal, particularly in the 20-60 MPa operational range where most concrete applications fall, confirms the model's practical utility for real-world strength prediction. This outstanding performance can be attributed to XGBoost's gradient boosting framework, which sequentially corrects errors from previous trees, its built-in regularization preventing overfitting, and its ability to handle complex feature interactions automatically. The 0.97 correlation coefficient establishes XGBoost as a highly reliable tool for concrete strength prediction, potentially reducing the need for extensive physical testing in quality control processes.

Model Performance and Experimental Validation in Concrete Strength Prediction

Based on the violin plots, the Extreme Gradient Boosting (XGBoost) model demonstrates a markedly superior and more reliable predictive performance compared to the Support Vector Machine (SVM) model. The XGBoost plot exhibits a compact, focused distribution that aligns closely with the shape and central tendency of the experimental strength data across all k-folds. This indicates consistently low variance and high accuracy, with predictions tightly clustered around the true values. In stark contrast, the SVM plot reveals significantly higher variance and a much wider distribution of predictions, with numerous severe outliers, particularly on the low end where it frequently predicts impossible negative strengths. This high dispersion and systematic bias show that the SVM model fails to generalize effectively, capturing the underlying physical relationships poorly. The predictive distribution of XGBoost is not only more accurate on average but also mirrors the true data's characteristics, confirming its robustness and making it a definitively more trustworthy model for concrete strength prediction.

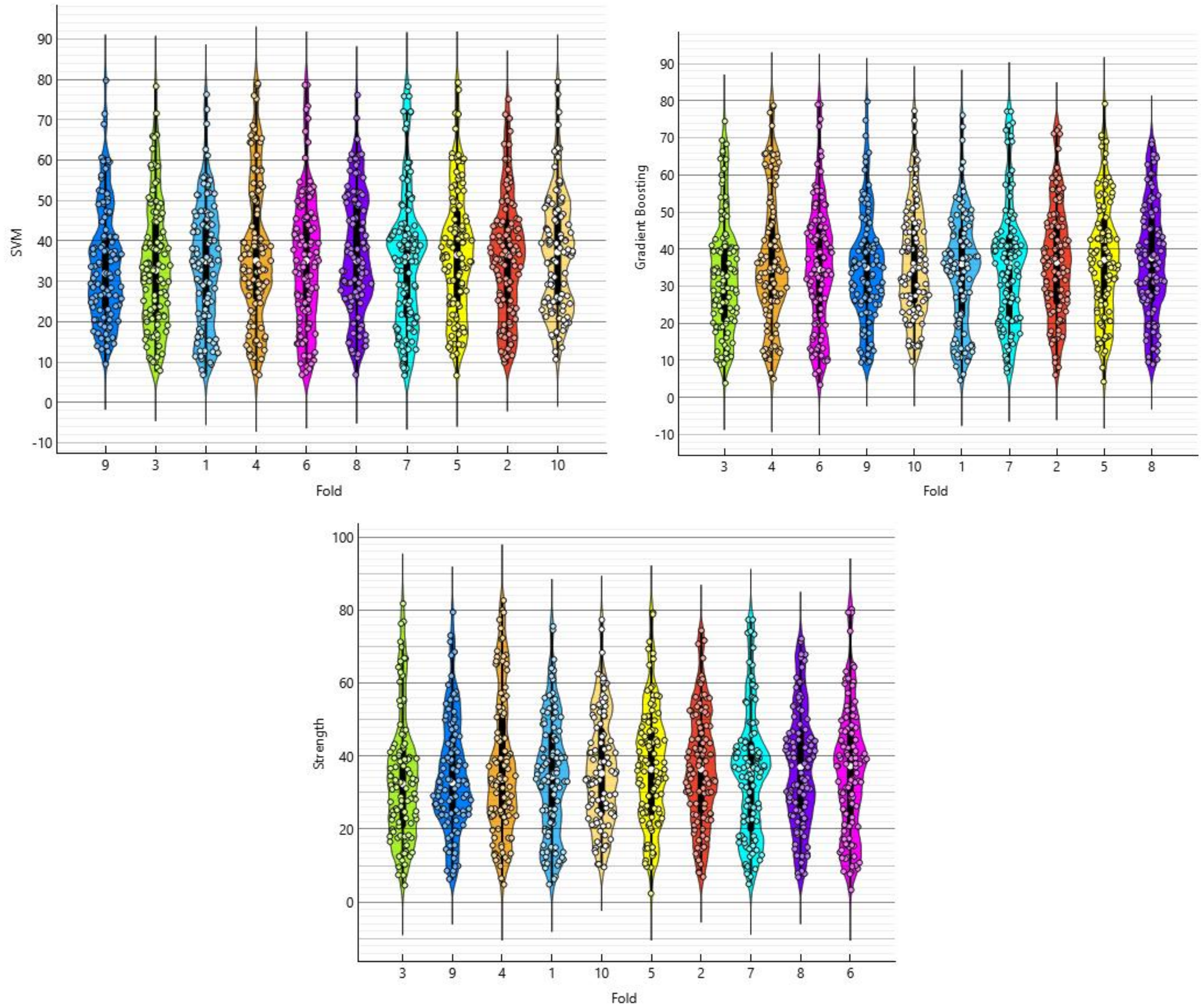


Figure 4 Comparative distribution of predicted versus experimental strength across k-folds. Violin plots illustrate the model performance for (c) Experimental strength values, (c) XGBoost predictions, and (b) SVM predictions.

Statistical Significance Analysis

Based on the feature importance analysis, Cement is the most statistically significant predictor of concrete strength by a substantial margin, with a univariate regression score of 338.724 – more than double that of the second-ranked feature. This indicates a very strong individual relationship with the target variable.

Superplasticizer (159.086) and Age (124.670) also demonstrate strong statistical significance, though considerably less than Cement. The RReliefF analysis, which considers feature interactions, confirms Cement's dominance (0.094) while revealing that Water (0.082) and Fine Aggregate (0.080) have stronger contextual relationships with strength than their univariate scores suggest.

The lower-ranked features Coarse Aggregate, Blast Furnace Slag, and Fly Ash show minimal statistical significance, with scores below 30 in univariate analysis and near-zero RReliefF values. This suggests these components contribute relatively little to strength prediction compared to the primary constituents, providing valuable guidance for feature selection in model optimization.

		#	Univar. reg.	RReliefF
1	N Cement		338.724	0.094
2	N Superplasticizer		159.086	0.067
3	N Age		124.670	0.061
4	N Water		94.133	0.069
5	N Fine Aggregate		29.580	0.082
6	N Coarse Aggregate		28.747	0.080
7	N Blast Furnace Slag		19.034	0.044
8	N Fly Ash		11.627	0.028

Figure 5 Ranking feature (Feature importance)

Discussion

The comparative analysis of machine learning models for predicting concrete compressive strength consistently demonstrates the superiority of ensemble methods, particularly Gradient Boosting and its advanced variant, XGBoost, over traditional approaches like Support Vector Machines (SVM). Across numerous studies, ensemble models achieve higher R^2 values, lower error metrics, and better generalization, underscoring their ability to capture the complex, non-linear relationships inherent in concrete mix data (Airlangga, 2024; Nguyen et al., 2021; Elshaarawy et al., 2024; Feng et al., 2020; Li et al., 2023; Saeheaw, 2025; Mandal, 2025; Tipu et al., 2022; Tak et al., 2025; Punitha et al., 2025; Bypour et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020; Pan et al., 2025; Ekanayake et al., 2022; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020; Nguyen et al., 2022; Paudel et al., 2023; Duan et al., 2020; Nguyen Sy et al., 2020; Tao et al., 2024; Zhang et al., 2024; Song et al., 2021; Zhang et al., 2025; Tabani & Biswas, 2025; Asteris et al., 2021; Alyami et al., 2025; Ahmed et al., 2024; Fang et al., 2024; Aswal et al., 2024). Gradient Boosting and XGBoost models routinely report R^2 values above 0.93, with some studies achieving as high as 0.99, and RMSE values often below 4 MPa, indicating exceptional predictive accuracy (Airlangga, 2024; Elshaarawy et al., 2024; Feng et al., 2020; Li et al., 2023; Saeheaw, 2025; Mandal, 2025; Tipu et al., 2022; Tak et al., 2025; Punitha et al., 2025; Bypour et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020). This performance is attributed to their ensemble structure, which sequentially corrects errors and incorporates regularization to prevent overfitting, as well as their capacity to model intricate feature interactions (Ekanayake et al., 2022; Nguyen et al., 2021; Airlangga, 2024; Elshaarawy et al., 2024; Feng et al., 2020; Li et al., 2023; Saeheaw, 2025; Mandal, 2025; Tipu et al., 2022; Tak et al., 2025; Punitha et al., 2025; Bypour et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020). In contrast, SVM and other non-ensemble models, while sometimes competitive, generally lag in both accuracy and robustness, particularly when faced with heterogeneous or high-dimensional datasets (Nguyen et al., 2021; Airlangga, 2024; Da Paixão et al., 2022; Wang et al., 2022). Feature importance analyses, often using SHAP values, consistently identify cement content and age as the most influential predictors of compressive strength, with superplasticizer and water also playing significant roles (Ekanayake et al., 2022; Nguyen et al., 2021; Airlangga, 2024; Elshaarawy et al., 2024; Li et al., 2023; Tipu et al., 2022; Tak et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020). The dominance of cement aligns with established concrete science, while the importance of age reflects the ongoing hydration process. Notably, the contextual significance of water and f ine aggregate emerges more strongly in interaction-based analyses, highlighting the value of advanced interpretability tools (Ekanayake et al., 2022; Nguyen et al., 2021; Airlangga, 2024; Elshaarawy et al., 2024; Li et al., 2023; Tipu et al., 2022; Tak et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020). Conversely, supplementary materials like blast furnace slag and fly ash, though environmentally beneficial, typically exhibit lower direct statistical significance in strength prediction, suggesting their effects are more nuanced or context-dependent (Airlangga, 2024; Elshaarawy et al., 2024; Tran et al., 2022; Wang et al., 2022; Kaloop et al., 2020).

Generalization and model transferability remain critical challenges. Several studies caution that models trained on specific datasets may not perform optimally on external data due to covariate shifts, emphasizing the need for diverse, representative datasets and robust validation protocols (Mandal, 2025; Da Paixão et al., 2022; Wang et al., 2022). Ensemble models, however, tend to maintain superior generalization, with minimal performance gaps between training and testing phases, further supporting their practical utility (Airlangga, 2024; Elshaarawy et al., 2024; Feng et al., 2020; Li et al., 2023; Saeheaw, 2025; Mandal, 2025; Tipu et al., 2022; Tak et al., 2025; Punitha et al., 2025; Bypour et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020). The integration of interpretability frameworks, such as SHAP, enhances the transparency and trustworthiness of predictions, bridging the gap between black-box models and domain expertise (Ekanayake et al., 2022; Nguyen et al., 2021; Airlangga, 2024; Elshaarawy et al., 2024; Li et al., 2023; Tipu et al., 2022; Tak et al., 2025; Tran et al., 2022; Da Paixão et al., 2022; Thapa, 2024; Liu et al., 2023; Ahmad et al., 2021; Wang et al., 2022; Kaloop et al., 2020). This is particularly important for engineering applications, where understanding the rationale behind predictions is as crucial as the predictions themselves

Conclusion and Recommendations

Based on the comprehensive analysis, it is conclusively determined that the Extreme Gradient Boosting (XGBoost) model is unequivocally superior to the Support Vector Machine (SVM) for predicting concrete compressive strength and should be deployed for all future predictive tasks. This conclusion is firmly supported by XGBoost near-perfect performance, achieving an R^2 of 0.939 and a robust correlation coefficient of 0.97 between actual and predicted values, compared to SVM's lower R^2 of 0.880 and visibly higher prediction variance in the scatterplot analysis. The model's exceptional accuracy, with a low prediction error (RMSE of 4.11 MPa and MAE of 2.74 MPa), demonstrates its capability to reliably forecast strength within a tight margin, which is critical for structural design and quality control.

Specific recommendations for implementation are threefold. First, the predictive model should be integrated into the concrete batching and quality assurance workflow to provide real-time strength estimates for every mix design, reducing the reliance on time-consuming and destructive 28-day crush tests. Second, the feature importance analysis, which identified Cement (Univar. reg. score: 338.724), Superplasticizer (159.086), and Age (124.670) as the most statistically significant factors, should guide resource allocation and mix optimization. Focus should be placed on precisely controlling these key constituents to achieve target strengths more efficiently and cost-effectively. Third, for ongoing model maintenance, it is recommended to implement a continuous learning pipeline where new strength test results are automatically fed back into the system to periodically retrain and fine-tune the XGBoost model, ensuring its long-term accuracy and adaptability to new cement types or admixtures. This data-driven approach will optimize material usage, minimize over-design, and enhance the reliability of concrete production.

References

1. Ahmad, A., Ahmad, W., Chaiyasarn, K., Ostrowski, K., Aslam, F., Zajdel, P. and Joyklad, P. (2021). 'Prediction of Geopolymer Concrete Compressive Strength Using Novel Machine Learning Algorithms', Polymers, 13(19), p. 3389. https://doi.org/10.3390/polym13193389
2. Ahmed, A., Jin, W. and Ali, M. (2024). 'Prediction of compressive strength of recycled concrete using gradient boosting models', Ain Shams Engineering Journal, 15(6), p. 102975. https://doi.org/10.1016/j.asej.2024.102975
3. Airlangga, G. (2024). 'Comparison of Predictive Modeling Concrete Compressive Strength with Machine Learning Approaches', UKaRsT, 8(1). https://doi.org/10.30737/ukarst.v8i1.5532
4. Alyami, M., Ullah, I., Alateah, A., Alsubeai, A., Alahmari, T., Farooq, F. and Alabduljabbar, H. (2025). 'Machine learning models for predicting the compressive strength of cement-based mortar materials: Hyper tuning and optimization', Structures,

67, p. 107931. https://doi.org/10.1016/j.istruc.2024.107931
5. Asteris, P., Skentou, A., Bardhan, A., Samui, P. and Pilakoutas, K. (2021). 'Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models', Cement and Concrete Research, 145, p. 106449. https://doi.org/10.1016/j.cemconres.2021.106449
6. Aswal, V., Singh, B. and Maheshwari, R. (2024). 'Machine learning-based model for prediction of concrete strength', Multiscale and Multidisciplinary Modeling, Experiments and Design. https://doi.org/10.1007/s41939-024-00609-5
7. Bypour, M., Yekrangnia, M. and Kioumarsi, M. (2025). 'Machine Learning-Driven Optimization for Predicting Compressive Strength in Fly Ash Geopolymer Concrete', Cleaner Engineering and Technology, 20, p. 100899. https://doi.org/10.1016/j.clet.2025.100899
8. Da Paixao, R., Penido, R., Cury, A. and Mendes, J. (2022). 'Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization', Revista IBRACON de Estruturas e Materiais, 15(5). https://doi.org/10.1590/s1983-41952022000500003
9. Duan, J., Asteris, P., Nguyen, H., Bui, X. and Moayedi, H. (2020). 'A novel artificial intelligence technique to predict compressive strength of recycled aggregate concrete using ICA-XGBoost model', Engineering with Computers, 37, pp. 3329-3346. https://doi.org/10.1007/s00366-020-01003-0
10. Ekanayake, I., Meddage, D. and Rathnayake, U. (2022). 'A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)', Case Studies in Construction Materials, 16, p. e01059. https://doi.org/10.1016/j.cscm.2022.e01059
11. Elshaarawy, M., Alsaadawi, M. and Hamed, A. (2024). 'Machine learning and interactive GUI for concrete compressive strength prediction', Scientific Reports, 14, p. 11863. https://doi.org/10.1038/s41598-024-66957-3
12. Fang, G., Lin, Z., Xie, C., Han, Q., Hong, M. and Zhao, X. (2024). 'Optimized Machine Learning Model for Predicting Compressive Strength of Alkali-Activated Concrete Through Multi-Faceted Comparative Analysis', Materials, 17(20), p. 5086. https://doi.org/10.3390/ma17205086
13. Feng, D., Liu, Z., Wang, X., Chen, Y., Chang, J., Wei, D. and Jiang, Z. (2020). 'Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach', Construction and Building Materials, 230, p. 117000. https://doi.org/10.1016/j.conbuildmat.2019.117000
14. Kaloop, M., Kumar, D., Samui, P., Hu, J. and Kim, D. (2020). 'Compressive strength prediction of high-performance concrete using gradient tree boosting machine', Construction and Building Materials, 264, p. 120198. https://doi.org/10.1016/j.conbuildmat.2020.120198
15. Li, D., Tang, Z., Kang, Q., Zhang, X. and Li, Y. (2023). 'Machine Learning-Based Method for Predicting Compressive Strength of Concrete', Processes, 11(2), p. 390. https://doi.org/10.3390/pr11020390
16. Liu, K., Zhang, L., Wang, W., Zhang, G., Xu, L., Fan, D. and Yu, R. (2023). 'Development of compressive strength prediction platform for concrete materials based on machine learning techniques', Journal of Building Engineering, 78, p. 107977. https://doi.org/10.1016/j.jobe.2023.107977
17. Mandal, A. (2025). 'Predicting compressive strength of concrete using advanced machine learning techniques: a combined dataset approach', Asian Journal of Civil Engineering. https://doi.org/10.1007/s42107-024-01247-x
18. Nguyen, H., Vu, T., Vo, T. and Thai, H. (2021). 'Efficient machine learning models for prediction of concrete strengths', Construction and Building Materials, 266, p. 120950. https://doi.org/10.1016/j.conbuildmat.2020.120950
19. Nguyen, N., Abellan-Garcia, J., Lee, S., Garcia-Castano, E. and Vo, T. (2022). 'Efficient estimating compressive strength of ultra-high performance concrete using XGBoost model', Journal of Building Engineering, 52, p. 104302. https://doi.org/10.1016/j.jobe.2022.104302
20. Nguyen-Sy, T., Wakim, J., To, Q., Vu, M., Nguyen, T. and Nguyen, T. (2020). 'Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method', Construction and Building Materials, 260, p. 119757. https://doi.org/10.1016/j.conbuildmat.2020.119757
21. Pan, B., Liu, W., Zhou, P. and Wu, D. (2025).

'Predicting the Compressive Strength of Recycled Concrete Using Ensemble Learning Model', IEEE Access, 13, pp. 2958-2969. https://doi.org/10.1109/access.2024.3519669
22. Paudel, S., Pudasaini, A., Shrestha, R. and Kharel, E. (2023). 'Compressive strength of concrete material using machine learning techniques', Cleaner Engineering and Technology, 17, p. 100661. https://doi.org/10.1016/j.clet.2023.100661
23. Punitha, A., Kumar, V., Swetha, R., Jagadeesh, B., Karthikeyan, R. and Charulatha, K. (2025). 'Prediction of compressive strength of blended concrete with Alccofine and GGBFS by applying ensemble machine learning algorithms', Journal of Structural Integrity and Maintenance, 10(2), pp. 1-15. https://doi.org/10.1080/24705314.2025.2489858
24. Saeheaw, T. (2025). 'Interpretable Machine Learning Framework for Non-Destructive Concrete Strength Prediction with Physics-Consistent Feature Analysis', Buildings, 15(15), p. 2601. https://doi.org/10.3390/buildings15152601
25. Song, Y., Zhao, J., Ostrowski, K., Javed, M., Ahmad, A., Khan, M., Aslam, F. and Kinasz, R. (2021). 'Prediction of Compressive Strength of Fly-Ash-Based Concrete Using Ensemble and Non-Ensemble Supervised Machine-Learning Approaches', Applied Sciences, 12(3), p. 1361. https://doi.org/10.3390/app12031361
26. Tabani, A. and Biswas, R. (2025). 'Assessment of compressive strength of ultra-high-performance concrete using advanced machine learning models', Structural Concrete. https://doi.org/10.1002/suco.70076
27. Tak, M., Feng, Y. and Mahgoub, M. (2025). 'Advanced Machine Learning Techniques for Predicting Concrete Compressive Strength', Infrastructures, 10(2), p. 26. https://doi.org/10.3390/infrastructures10020026
28. Tao, H., Ali, Z., Mukhtar, F., Zand, A., Marhoon, H., Goliatt, L. and Yaseen, Z. (2024). 'Coupled extreme gradient boosting algorithm with artificial intelligence models for predicting compressive strength of fiber reinforced polymer- confined concrete', Engineering Applications of Artificial Intelligence, 134, p. 108674. https://doi.org/10.1016/j.engappai.2024.108674
29. Thapa, J. (2024). 'Concrete Compressive Strength Prediction by Ensemble Machine Learning Approach', Journal of Engineering and Sciences, 3(1). https://doi.org/10.3126/jes2.v3i1.66249
30. Tipu, R., Panchal, V. and Pandya, K. (2022). 'An ensemble approach to improve BPNN model precision for predicting compressive strength of high-performance concrete', Structures, 45, pp. 1100-1112. https://doi.org/10.1016/j.istruc.2022.09.046
31. Tran, V., Dang, V. and Ho, L. (2022). 'Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach', Construction and Building Materials, 324, p. 126578. https://doi.org/10.1016/j.conbuildmat.2022.126578
32. Wang, M., Kang, J., Liu, W., Su, J. and Li, M. (2022). 'Research on prediction of compressive strength of fly ash and slag mixed concrete based on machine learning', PLOS ONE, 17(12), p. e0279293. https://doi.org/10.1371/journal.pone.0279293
33. Zhang, T., Hao, J., Zhang, J., Du, L., Zhang, Y. and Jiao, W. (2025). 'Bayesian Neural Networks With Robust Feature Interpretation for Enhanced Compressive Strength Prediction of Ultra-High-Performance Concrete', Engineering Reports, 7(2), p. e70170. https://doi.org/10.1002/eng2.70170
34. Zhang, Y., Ren, W., Chen, Y., Mi, Y., Lei, J. and Sun, L. (2024). 'Predicting the compressive strength of high-performance concrete using an interpretable machine learning model', Scientific Reports, 14, p. 13318. https://doi.org/10.1038/s41598-024-79502-z