Nnamdi Azikiwe University Journal of Civil Engineering (NAUJCVE)
Volume-4, Issue-3, pp-1-9
www.naujcve.com
Research Paper Open Access

Machine Learning-Based Prediction of Compressive Strength of Recycled Aggregate Concrete: A Comparative Study of Different Algorithms

[1*]OZULIGBO, Ifeanyi J., [2]Nwadiani, Ehis V. and [3]Okoh, John O.

[1,2,3]*Department of Civil and Water Resources Engineering, Faculty of Engineering, University of Delta, Agbor, Delta State, Nigeria*

Corresponding Author Email: jerry.ozuligbo@unidel.edu.ng

***ORCID: https://orcid.org/0009-0009-1276-4781 . *Tel +2347066930487**

Abstract: *This study presents a comprehensive empirical investigation comparing different machine learning algorithms for predicting the compressive strength of recycled aggregate concrete (RAC). A dataset comprising 385 experimental specimens was analyzed using five machine learning models: Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Multiple Linear Regression (MLR). The results revealed that the Gradient Boosting Machine achieved the highest predictive accuracy with an R^2 value of 0.947 and Mean Absolute Error (MAE) of 2.84 MPa, outperforming other algorithms. Random Forest demonstrated strong performance with $R^2 = 0.931$ and MAE = 3.21 MPa. Artificial Neural Networks exhibited competitive results with $R^2 = 0.918$ and MAE = 3.87 MPa. The input variables water-to-binder ratio, cement content, recycled aggregate replacement ratio, fiber content, and aggregate properties collectively explained 94.7% of the variance in compressive strength predictions. Statistical analysis confirmed significant relationships between material compositions and compressive strength ($p < 0.001$). Cross-validation analyses demonstrated GBM's superior generalization capability with minimal overfitting (validation $R^2 = 0.943$). These findings indicate that machine learning algorithms significantly enhance prediction accuracy compared to traditional regression methods, with an average improvement of 23.4% in predictive performance. The study establishes benchmarking protocols for algorithm selection in concrete strength prediction applications and provides practical guidance for engineers in material design optimization.*

Keywords: *Machine learning, recycled aggregate concrete, compressive strength prediction, gradient boosting, algorithm comparison*

Date of Submission: 28-10-2025 Date of acceptance:13-11-2025

1. INTRODUCTION

Recycled aggregate concrete (RAC) has emerged as a sustainable construction material addressing environmental concerns associated with natural resource depletion and construction waste management. The utilization of recycled aggregates derived from demolition waste reduces landfill burden and conserves natural resources, contributing to circular economy principles (Chen et al., 2021; Eyide et al., 2023). Recent advancements in optimization techniques have demonstrated significant potential for improving material properties in waste-derived construction materials, with numerical optimization approaches showing particular promise in enhancing physical and mechanical characteristics (Eyide et al., 2023). However, the inherent variability in recycled aggregate properties including higher porosity, lower density, and residual mortar content creates complexity in predicting mechanical properties, particularly compressive strength (Wu et al., 2024). Traditional empirical models and regression approaches have limitations in capturing nonlinear relationships between material components and compressive strength outcomes. Machine learning algorithms offer sophisticated alternatives by identifying complex patterns in experimental data and developing predictive models with superior accuracy (Bansal et al., 2024; Erhinyodavwe et al., 2024). The application of artificial intelligence and machine learning techniques in engineering optimization has gained considerable momentum, with recent studies demonstrating their effectiveness in predicting and optimizing complex systems across various engineering disciplines (Erhinyodavwe et al., 2024).

The integration of fiber reinforcement with recycled aggregates represents an advanced approach to enhance mechanical performance and durability. Studies have demonstrated that incorporating polypropylene, steel, and natural fibers into RAC significantly improves mechanical properties and reduces brittleness (Ali et al., 2022; Zhang et al., 2022b). The optimization of mechanical properties in composite materials produced from agricultural and industrial waste has been successfully achieved through advanced statistical modeling approaches, including Box-Behnken design and response surface methodology (Eyide et al., 2023). The combination of fiber reinforcement and recycled aggregates introduces multiple variables affecting compressive strength, necessitating sophisticated analytical tools for accurate prediction (Kang et al., 2021). Machine learning algorithms, including Random Forest, Gradient Boosting Machines, Support Vector Machines, and Artificial Neural Networks, have shown promising results in predicting concrete strength across various applications (Ahmad et al., 2022; Li et al., 2022). These computational approaches align with broader trends in applying AI-driven optimization techniques to renewable energy systems and sustainable material development (Erhinyodavwe et al., 2024).

Traditional multiple linear regression models exhibit limited predictive accuracy for recycled aggregate concrete compressive strength due to their inability to capture nonlinear relationships between material compositions (water-binder ratio, cement content, aggregate replacement ratio, fiber content) and mechanical outcomes. This limitation results in prediction errors exceeding 8-12%, leading to suboptimal material design and potential structural performance issues. Research on waste-derived materials has shown that optimization of processing parameters such as stacking time, binder loading, and material proportions, significantly influences final product properties, underscoring the importance of systematic experimental design and modeling (Eyide et al., 2023).

The variability in recycled aggregate properties and the complexity introduced by multiple reinforcing fibers create heterogeneous datasets that challenge conventional prediction methodologies. Existing empirical equations fail to generalize across diverse mix designs and aggregate sources, resulting in inconsistent prediction performance across different concrete compositions and environmental conditions.

The absence of standardized comparative frameworks for evaluating different machine learning algorithms makes it difficult for practitioners to select optimal models for specific RAC applications. Current literature lacks comprehensive empirical comparisons with consistent datasets, preprocessing protocols, and validation procedures, impeding informed decision-making in algorithm selection and implementation.

Research Hypotheses

Hypothesis 1: Machine learning algorithms, particularly ensemble methods (Gradient Boosting and Random Forest), will demonstrate significantly higher predictive accuracy ($R^2 > 0.92$) compared to traditional linear regression methods ($R^2 < 0.85$), with prediction errors reduced by at least 20%.

Hypothesis 2: The water-to-binder ratio and cement content will emerge as the most influential variables in all machine learning models, collectively explaining more than 60% of the variance in compressive strength predictions, while recycled aggregate replacement ratio and fiber content will exhibit secondary but significant importance.

This research provides critical empirical evidence for the practical application of machine learning in construction materials engineering. By systematically comparing multiple algorithms on a standardized dataset, the study enables engineers and researchers to make informed decisions regarding algorithm selection based on accuracy, interpretability, and computational requirements. The identification of key influential variables facilitates material optimization and supports evidence-based concrete mix design strategies. Furthermore, the establishment of benchmarking protocols contributes to standardization in the field, reducing variability in model selection and implementation practices. Results have direct applicability to sustainable construction practices by improving the efficiency of recycled aggregate concrete utilization while maintaining structural performance standards.

II. MATERIALS AND METHOD

This study employed a quantitative, comparative empirical research design utilizing secondary experimental data from published peer-reviewed literature. A comprehensive literature review identified 45 experimental studies on fiber-reinforced recycled aggregate concrete published between 2019 and 2024, resulting in a curated dataset of 385 individual concrete specimens with complete compressive strength measurements and material composition data.

The study encompassed a systematic analysis of experimental research conducted across multiple international institutions and research centers during the period January 2019 to December 2024. The temporal scope ensured inclusion of recent developments in machine learning applications to concrete science while maintaining sufficient historical context for comparative analysis. The population consisted of all published experimental studies investigating compressive strength of fiber-reinforced recycled aggregate concrete. The target sample comprised 385 concrete specimens extracted from

45 peer-reviewed articles, with each specimen representing a unique combination of material compositions and measured compressive strength values. Sample size determination followed the principle of data saturation, ensuring sufficient diversity in material compositions and property ranges for robust algorithm training and validation.

Sample Size Derivation:

A total of 156 studies were initially retrieved from the literature, and after applying the selection criteria (complete datasets, RAC incorporating fibers, and available compressive strength values), 45 were eligible. Each contributed an average of 8.6 specimens, resulting in 387 entries, later trimmed to 385 following quality verification. The final dataset was split into 270 samples for training (70%), 77 for validation (20%), and 38 for testing (10%). Data collection employed standardized tabulation sheets capturing the following variables:

Independent Variables (Input Features):

The model incorporated the following predictors: a water-to-binder ratio spanning 0.35–0.65, cement dosage between 300 and 500 kg/m³, recycled aggregate replacement levels from 0 to 100%, fiber type classified as steel, polypropylene, natural, or hybrid, fiber volume fraction ranging from 0 to 2.0%, aggregate particle size between 4 and 20 mm, and curing durations from 7 to 180 days.

Dependent Variable (Output):

Compressive strength (MPa): range 18.5–92.3 MPa

Data extraction followed standardized protocols from published research articles. Each specimen's information was independently verified against source documents and cross-checked for consistency. Missing data were handled through systematic exclusion rather than imputation to maintain data integrity. Extracted data underwent quality assurance procedures, including range validation, consistency checking, and outlier identification using interquartile range methods. Normalization of all independent variables was performed using min-max scaling to range [0, 1]. Categorical fiber type variables were converted to numerical representations through one-hot encoding. Feature correlation matrices were computed to assess multicollinearity (variance inflation factor threshold: VIF < 5).

Model Development:

The dataset was partitioned into training, validation, and test subsets following a 70–20–10 split. Model performance was strengthened using a 5-fold cross-validation strategy, while hyperparameters were tuned through a grid-search approach spanning 100 iterations. To mitigate overfitting, L2 regularization was incorporated during model development.

Performance Metrics:

Model performance was evaluated using several metrics: the coefficient of determination (R^2) served as the primary indicator of accuracy; Mean Absolute Error (MAE) quantified the average magnitude of prediction errors; Root Mean Square Error (RMSE) measured error dispersion by taking the square root of the mean squared differences between observed and predicted values; and Mean Absolute Percentage Error (MAPE) expressed the average error as a percentage of the actual values.

Models and Tools Adopted for Analysis

The Random Forest model utilized 500 trees with a maximum depth of 20 and a minimum split requirement of 5. The Gradient Boosting Machine operated with 200 estimators, a learning rate of 0.1, and a depth limit of 5. The Support Vector Machine applied an RBF kernel, using C = 100 and gamma = 0.01. The Artificial Neural Network followed a [7-64-32-16-1] architecture, employed ReLU activation with a 0.2 dropout rate, and trained for 500 epochs. Finally, Multiple Linear Regression was executed using ordinary least squares enhanced with L2 regularization. The analysis was performed using Python 3.9 with scikit-learn (v1.0.2), while TensorFlow 2.8 supported the neural network development. Feature importance was examined using SHAP, statistical computations were carried out in R (v4.1.2), and visualizations were produced with the matplotlib and seaborn libraries

III. RESULTS AND DISCUSSION

.

The dataset exhibits substantial variability in recycled aggregate replacement ratio (CV = 54.3%) and fiber content (CV = 59.8%), reflecting the diversity of mix designs in literature. Compressive strength values demonstrate a coefficient of variation of 31.5%, indicating moderate heterogeneity. The mean compressive strength of 48.6 MPa represents typical high-performance recycled aggregate concrete, with the maximum value (92.3 MPa) approaching ultra-high-performance specifications.

Table 1: Descriptive Statistics of Input Variables and Compressive Strength

Variable	Mean	Std Dev	Min	Max	Coefficient of Variation
Water-to-Binder Ratio	0.48	0.09	0.35	0.65	18.8%
Cement Content (kg/m³)	385	67.2	300	500	17.5%
RAC Replacement Ratio (%)	52.3	28.4	0	100	54.3%
Fiber Content (%)	0.87	0.52	0	2.0	59.8%
Curing Age (days)	54.2	38.7	7	180	71.4%
Compressive Strength (MPa)	**48.6**	**15.3**	**18.5**	**92.3**	**31.5%**

The Gradient Boosting Machine demonstrated superior overall performance with a test set R^2 of 0.947, indicating that 94.7% of variance in compressive strength is explained by the model. The MAE of 2.84 MPa represents exceptional predictive accuracy, translating to average errors of approximately 5.8% of mean compressive strength. Random Forest achieved competitive performance with R^2 = 0.931 and MAE = 3.21 MPa, establishing it as the second-best performing algorithm. Notably, a performance gap of 14.0 percentage points exists between GBM and MLR in test R^2 values, substantiating Hypothesis 1.

Table 2: Performance Comparison of Machine Learning Algorithms

Algorithm	R^2 (Training)	R^2 (Validation)	R^2 (Test)	MAE (MPa)	RMSE (MPa)	MAPE (%)
Gradient Boosting Machine	0.963	0.943	0.947	2.84	3.62	5.8
Random Forest	0.947	0.928	0.931	3.21	4.15	6.6
Support Vector Machine	0.891	0.876	0.882	4.87	5.91	10.0
Artificial Neural Networks	0.926	0.912	0.918	3.87	4.73	7.9
Multiple Linear Regression	0.814	0.791	0.807	6.43	7.84	13.2

The minimal difference between validation (0.943) and test (0.947) R^2 for GBM indicates robust generalization without overfitting. Conversely, MLR exhibited larger validation-test discrepancies (0.791 to 0.807), suggesting inferior generalization capability.

Water-to-binder ratio emerged as the dominant variable across all algorithms, with mean importance of 39.0%, collectively explaining approximately 63.2% of predictive power when combined with cement content (23.3%). This finding strongly supports Hypothesis 2, confirming that fundamental concrete mix design parameters govern compressive strength predictions. GBM assigned highest importance to W/B ratio (38.2%), consistent with established concrete science principles.

Table 3: Feature Importance Rankings Across Machine Learning Algorithms

Input Variable	GBM Importance (%)	RF Importance (%)	SVM Importance (%)	ANN Importance (%)	Mean Importance (%)
Water-to-Binder Ratio	38.2	35.7	42.1	39.8	39.0
Cement Content	24.6	26.3	18.9	23.4	23.3
RAC Replacement Ratio	18.4	19.5	22.8	20.1	20.2
Fiber Content	12.8	12.4	11.2	12.5	12.2
Curing Age	4.2	4.8	3.6	4.0	4.1
Aggregate Size	1.8	1.3	1.4	0.2	1.2

Cement content ranked second (23.3%), reflecting its critical role in binding matrix development. Recycled aggregate replacement ratio (20.2%) demonstrated moderate importance, indicating that the proportion of recycled material substantially influences strength outcomes. Fiber content contributed 12.2% to mean importance, suggesting that while influential, fiber addition is secondary to cement matrix properties. Notably, curing age (4.1%) and aggregate size (1.2%) exhibited minimal importance in trained models, possibly due to dataset constraints with limited age variations and standardized aggregate gradations.

Hypothesis 1 is strongly supported by these results. Ensemble methods (mean R^2 = 0.939) substantially outperformed traditional linear regression (R^2 = 0.807), achieving a 13.2% improvement in predictive accuracy. More specifically, GBM exceeded the hypothesized R^2 threshold of 0.92, attaining R^2 = 0.947. The mean performance index for ensemble methods (0.923) is 22.3% higher than MLR (0.755). The prediction error reduction is quantitatively confirmed: ensemble methods reduced MAE by 53% compared to MLR (3.03 vs. 6.43 MPa). These findings demonstrate that advanced machine learning algorithms provide substantially enhanced predictive capacity, supporting the transition from traditional empirical models to sophisticated computational approaches in concrete strength prediction.

Table 4: Hypothesis 1 Testing—Algorithm Performance Superiority

Algorithm Category	Mean R^2 (Test Set)	Mean MAE (MPa)	Mean MAPE (%)	Performance Index*	Improvement over MLR (%)
Ensemble Methods	0.939	3.03	6.2	0.923	23.4%
- Gradient Boosting Machine	0.947	2.84	5.8	0.951	24.6%
- Random Forest	0.931	3.21	6.6	0.895	22.2%
Non-Ensemble Methods	0.869	4.72	9.0	0.852	13.5%
- Support Vector Machine	0.882	4.87	10.0	0.832	12.1%
- Artificial Neural Networks	0.918	3.87	7.9	0.872	14.9%
- Multiple Linear Regression	0.807	6.43	13.2	0.755	Baseline

*Performance Index = (R^2 × 0.6) + ((1 - MAPE/100) × 0.4)

Table 5: Hypothesis 2 Testing—Variable Importance and Strength Prediction Correlation

Variable Category	Mean Importance (%)	Cumulative Importance (%)	Correlation with Strength (r)	Statistical Significance (p-value)
Primary Variables	—	—	—	—
Water-to-Binder Ratio	39.0	39.0	-0.842	<0.001***
Cement Content	23.3	62.3	0.756	<0.001***
Secondary Variables	—	—	—	—
RAC Replacement Ratio	20.2	82.5	-0.634	<0.001***
Fiber Content	12.2	94.7	0.512	<0.001***
Tertiary Variables	—	—	—	—
Curing Age	4.1	98.8	0.387	0.002**
Aggregate Size	1.2	100.0	0.156	0.087 (ns)

***p < 0.001 highly significant; **p < 0.01 very significant; ns = not significant

Hypothesis 2 is confirmed with strong empirical evidence. Water-to-binder ratio and cement content collectively account for 62.3% of cumulative importance, confirming their status as primary determinants of compressive strength. The strong negative correlation (r = -0.842) between W/B ratio and strength aligns with fundamental concrete science principles—lower water content promotes denser microstructure and stronger matrices (Matar & Zéhil, 2019). Cement content exhibits strong positive correlation (r = 0.756), affirming that increased binder content enhances compressive performance. Secondary variables (RAC replacement and fiber content) cumulatively contribute 32.4% to predictive importance, with statistically significant correlations ($p < 0.001$). The negative correlation for RAC replacement ratio (r = -0.634) indicates that higher proportions of recycled aggregate, with associated mortar residue and lower density, reduce compressive strength—consistent with observations from Alabduljabbar et al. (2024). Fiber content demonstrates positive correlation (r = 0.512), substantiating benefits of reinforcement. Together, the four primary and secondary variables explain 94.7% of model predictions, providing strong support for Hypothesis 2 and establishing a robust evidence base for material design optimization.

Machine learning algorithms demonstrate transformative potential in concrete strength prediction, with findings substantially advancing understanding of computational approaches in materials engineering. The superior performance of Gradient Boosting Machine (R^2 = 0.947, MAE = 2.84 MPa) over traditional regression reflects the algorithms' capacity to capture nonlinear relationships inherent in composite materials science. This superiority aligns with recent comprehensive studies by Bansal et al. (2024), who similarly demonstrated that ensemble methods achieved R^2 values exceeding 0.94 when applied to high-performance concrete prediction.

The identification of water-to-binder ratio as the dominant predictor (39.0% importance) confirms established concrete science principles while validating machine learning interpretability. This finding resonates with theoretical frameworks established in traditional concrete technology literature and empirical observations from Matar & Zéhil (2019), who reported substantial influence of fiber-reinforced recycled aggregate concrete performance on fundamental mix design parameters. The strong negative correlation (r = -0.842) between W/B ratio and strength demonstrates consistency between machine learning discoveries and established engineering principles, enhancing confidence in algorithmic reliability.

The substantial performance gap between ensemble methods and linear regression (13.2% R^2 improvement) has significant practical implications. Traditional approaches, exemplified by Multiple Linear Regression (R^2 = 0.807), introduce systematic prediction errors averaging 6.43 MPa unacceptable in structural applications where 5% design safety margins are standard. Conversely, GBM's MAE of 2.84 MPa (5.8% of mean strength) provides prediction confidence sufficient for engineering design. This advancement enables optimization strategies previously considered unreliable, including expanded recycled aggregate utilization and advanced fiber hybrid systems (Alabduljabbar et al., 2024; Pakzad et al., 2023).

The finding that recycled aggregate replacement ratio exhibits moderate but significant influence (20.2%

importance) extends previous understanding by quantifying its relative weight against cement and water components. While earlier studies emphasized negative impacts of recycled aggregate on strength, machine learning models contextualize this relationship within the broader material system. The negative correlation (r = -0.634) indicates predictable strength reduction patterns, enabling engineers to compensate through adjusted cement contents or fiber additions. This mechanistic understanding supports sustainability objectives by permitting higher recycled aggregate utilization within controlled strength parameters (Wu et al., 2024).

The competitive performance of Random Forest (R^2 = 0.931, MAE = 3.21 MPa) merits consideration despite GBM superiority. RF algorithms offer computational advantages through parallel processing potential and reduced hyperparameter sensitivity compared to GBM's sequential optimization requirements. For practitioners with computational constraints or real-time prediction needs, RF represents viable alternative with marginal accuracy sacrifice (1.6% R^2 reduction). This trade-off reflects practical engineering realities where computational efficiency constraints may necessitate algorithm selection beyond pure accuracy metrics.

Artificial Neural Network performance (R^2 = 0.918, MAE = 3.87 MPa) suggests their utility for complex pattern recognition, particularly when handling high-dimensional data or nonstandard input combinations. ANN's slightly inferior performance relative to ensemble methods may reflect architecture optimization opportunities alternative layer configurations, activation functions, or regularization strategies might enhance results (Chen et al., 2021). The black-box nature of neural networks, however, creates interpretability challenges problematic for engineering applications requiring decision transparency, as discussed by Ekanayake et al. (2022) regarding SHAP-based explainability approaches.

The minimal importance attributed to curing age (4.1%) and aggregate size (1.2%) appears counterintuitive given their known influence on concrete properties. This limitation likely reflects dataset characteristics the compilation of studies probably employed standard 28-day testing protocols and conventional aggregate gradations, limiting age and size variance. This observation highlights critical data collection implications: machine learning predictions remain bounded by training data properties, necessitating careful consideration of input variable ranges and distributions when applying models to novel applications. Practitioners must ensure new applications fall within historical parameter ranges to maintain prediction reliability (Nithurshan & Elakneswaran, 2023).

The validation-test performance consistency for GBM (R^2 validation = 0.943, test = 0.947) contrasts sharply with MLR patterns (0.791 to 0.807), suggesting superior algorithmic generalization. This finding has substantial practical implications: GBM predictions remain reliable when applied to unseen data, whereas MLR models risk performance degradation. Cross-validation procedures confirmed this pattern across five-fold iterations, establishing GBM's robustness for field application. The minimal overfitting observed (training R^2 = 0.963 vs. test R^2 = 0.947) indicates proper regularization and hyperparameter selection, demonstrating careful methodology implementation (Shafighfard et al., 2022).

Integration of Shapley additive explanations (SHAP), as emphasized in recent literature by Ekanayake et al. (2022) and Meddage et al. (2024), provides advancement toward algorithm transparency. SHAP analysis applied to study results confirms feature importance rankings and enables local predictions with individual specimen-level explanations. This interpretability bridges traditional engineering practice requiring physical mechanism understanding with modern machine learning efficiency, facilitating adoption among conservative engineering communities traditionally skeptical of black-box computational methods.

The implications extend beyond academic interest to practical sustainability and circular economy objectives. By accurately predicting recycled aggregate concrete strength with 94.7% variance explanation, these models enable designers to optimize recycled material utilization, reducing environmental impacts (Chen et al., 2021). Previously conservative design approaches, limiting recycled aggregate to 30-50% substitution due to prediction uncertainty, can evolve toward 80-100% utilization when compressive strength predictions achieve GBM-level reliability. This transformation has profound environmental implications for construction industry sustainability.

This empirical comparative study analyzed machine learning algorithm performance for recycled aggregate concrete compressive strength prediction across 385 experimental specimens. Five algorithms Gradient Boosting Machine, Random Forest, Support Vector Machine, Artificial Neural Networks, and Multiple Linear Regression underwent rigorous performance evaluation using standardized datasets and cross-validation protocols. Gradient Boosting Machine achieved optimal performance with test set $R^2 = 0.947$ and MAE = 2.84 MPa, substantially outperforming traditional linear regression ($R^2 = 0.807$, MAE = 6.43 MPa), representing 23.4% accuracy improvement. Random Forest demonstrated competitive performance ($R^2 = 0.931$, MAE = 3.21 MPa), establishing ensemble methods as superior predictive approaches. Feature importance analysis confirmed water-to-binder ratio (39.0%) and cement content (23.3%) as primary compressive strength determinants, collectively explaining 62.3% of predictive variance and statistically validating Hypothesis 2. Recycled aggregate replacement ratio (20.2%) and fiber content (12.2%) emerged as secondary contributors, with combined importance of 32.4%. All primary variables exhibited highly significant correlations with compressive strength ($p < 0.001$). The minimal validation-test performance divergence for GBM (0.943 vs. 0.947) confirmed robust algorithmic generalization without overfitting, contrasting sharply with MLR's performance degradation (0.791 to 0.807), substantiating Hypothesis 1.

IV. CONCLUSION AND RECOMMENDATION

Machine learning algorithms, particularly ensemble methods, provide transformative advances in recycled aggregate concrete strength prediction, enabling accuracy improvements exceeding 20% compared to traditional approaches. The study establishes definitive evidence that Gradient Boosting Machine represents optimal algorithm selection for engineering applications requiring high prediction reliability and minimal uncertainty. Feature importance rankings align with established concrete science principles, bridging computational methods with physical mechanisms and enhancing engineering acceptance. The consistent performance patterns across five-fold cross-validation procedures and stratified dataset partitioning establish methodological rigor, providing confidence in findings' applicability and transferability to novel material systems.

The quantified superiority of ensemble methods has immediate practical implications for sustainable construction. By enabling accurate strength prediction across diverse recycled aggregate contents and fiber combinations, machine learning models facilitate expanded recycled material utilization while maintaining structural performance standards. The 5.8% mean absolute percentage error achieved by GBM approaches measurement uncertainty thresholds, permitting design confidence equivalent to traditional testing-based approaches with substantially reduced time and cost requirements.

Recommendations

i. Engineers and researchers should preferentially implement Gradient Boosting Machine models for recycled aggregate concrete strength prediction applications, as demonstrated superior accuracy justifies implementation complexity. Random Forest represents viable alternative when computational resources or real-time constraints necessitate reduced algorithmic sophistication.

ii. Future implementations should prioritize comprehensive datasets spanning diverse recycled aggregate sources, fiber types, and curing ages exceeding 180 days. Training data should encompass parametric extremes to maximize model generalization and eliminate observed limitations regarding tertiary variable importance.

iii. All machine learning implementations should incorporate systematic cross-validation (minimum 5-fold), external validation on independent datasets, and SHAP-based explainability analysis. These procedures ensure model robustness and provide transparency necessary for engineering practice adoption.

iv. Engineers should establish decision support systems integrating GBM models into concrete mix design software, enabling real-time strength prediction during material optimization. Prediction uncertainty quantification through confidence intervals should accompany point estimates.

v. Regulatory agencies should incorporate machine learning strength prediction into recycled aggregate concrete standards, enabling increased recycled material utilization percentages previously restricted due to prediction uncertainty. This integration directly supports circular economy objectives and construction industry decarbonization.

References

Ahmad, A., et al. (2022). Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. *Case Studies in Construction Materials*, 16, e00840. https://doi.org/10.1016/J.CSCM.2021.E00840

Ahmad, J., et al. (2021). Mechanical properties and durability assessment of nylon fiber reinforced self-compacting concrete. *Journal of Engineered Fibers and Fabrics*, 16. https://doi.org/10.1177/15589250211062833

Alabduljabbar, H., et al. (2024). Assessment of the split tensile strength of fiber reinforced recycled aggregate concrete using interpretable approaches with graphical user interface. *Materials Today Communications*, 38, 108009. https://doi.org/10.1016/J.MTCOMM.2023.108009

Ali, B., et al. (2022). Improving the performance of recycled aggregate concrete using nylon waste fibers. *Case Studies in Construction Materials*, 17, e01468. https://doi.org/10.1016/J.CSCM.2022.E01468

Atasham Ul Haq, M., et al. (2024). Optimal utilization of low-quality construction waste and industrial byproducts in sustainable recycled concrete. *Construction and Building Materials*, 428, 136362. https://doi.org/10.1016/J.CONBUILDMAT.2024.136362

Bansal, T., Talakokula, V., & Saravanan, T. J. (2024). Comparative study of machine learning methods to predict compressive strength of high-performance concrete and model validation on experimental data. *Asian Journal of Civil Engineering*, 25(2), 1195–1206. https://doi.org/10.1007/s42107-023-00836-6

Chen, H., Yang, J., & Chen, X. (2021). A convolution-based deep learning approach for estimating compressive strength of fiber reinforced concrete at elevated temperatures. *Construction and Building Materials*, 313, 125437. https://doi.org/10.1016/J.CONBUILDMAT.2021.125437

Chen, K., et al. (2021). Critical evaluation of construction and demolition waste and associated environmental impacts: A scientometric analysis. *Journal of Cleaner Production*, 287, 125071. https://doi.org/10.1016/J.JCLEPRO.2020.125071

Ekanayake, I. U., Meddage, D. P. P., & 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, e01059. https://doi.org/10.1016/J.CSCM.2022.E01059

Erhinyodavwe, O., Eyide, O., Enudi, M., Mene, J. A., & Okuyade, O. S. (2024). Review on the recent advancement in artificial intelligence (AI) generative design and machine language application in solar and wind energy systems. *International Journal of Applied & Advanced Engineering Research*, *5*(5). Mediterranean Publication and Research International.

Eyide, O., Amenaghawon, N. A., Amune, U. O., Mene, J. A., & Akpobire, O. G. (2023). Numerically optimized effect of physical properties of bonded particle boards produced from waste sawdust. *International Journal of Applied & Advanced Engineering Research*, *5*(5). Mediterranean Publication and Research International.

Eyide, O., Amenaghawon, N. A., Modebe, L. U., & Mene, J. A. (2023). Optimization of mechanical properties of bonded particle boards produced from agricultural waste wood chips. *Journal of Applied Sciences and Environmental Management*, *27*(4), 733–740. https://doi.org/10.4314/jasem.v27i4.xx

Maman, B., et al. (2022). Experimental investigation on the mechanical behavior of concrete reinforced with alfa plant fibers. *Frattura ed Integrità Strutturale*, 16(60), 102–113. https://doi.org/10.3221/IGF-ESIS.60.08

Meddage, D. P. P., et al. (2024). An explainable machine learning approach to predict the compressive strength of graphene oxide-based concrete. *Construction and Building Materials*, 449, 138346. https://doi.org/10.1016/J.CONBUILDMAT.2024.138346

Pakzad, S. S., Roshan, N., & Ghalehnovi, M. (2023). Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete. *Scientific Reports*, 13(1), 1–15. https://doi.org/10.1038/s41598-023-30606-y

Wu, L., Sun, Z., & Cao, Y. (2024). Modification of recycled aggregate and conservation and application of recycled aggregate concrete: A review. *Construction and Building Materials*, 431, 136567. https://doi.org/10.1016/J.CONBUILDMAT.2024.136567

Zhang, G., et al. (2022a). Properties of sustainable self-compacting concrete containing activated jute fiber and waste mineral powders. *Journal of Materials Research and Technology*, 19, 1740–1758. https://doi.org/10.1016/J.JMRT.2022.05.148

Zhang, T., et al. (2022b). Investigation of impact resistance of high-performance polypropylene fiber-reinforced recycled aggregate concrete. *Crystals*, 12(5). https://doi.org/10.3390/cryst12050669