DOI: 10.36297/vw.jei.v8i1.900

VW Engineering International, Volume: 8, Issue: 1, 01-04

Machine Learning-Driven Discovery and Optimization of Eco-Friendly Construction Materials

Aarav Mehta[1*], Ritika Sharma[2*], Imran Hossain[3*]

[1]Department of Civil Engineering, Rajiv Gandhi Proudyogiki Vishwavidyalya, Bhopal, India
[2]Department of Materials Engineering, Sant Longwal Institute of Engineering and Technology, Punjab, India
[3]Department of Environmental Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh
[*]Email: aarav.m@rgpv.ac.in, ritikashar.ma@sliet.ac.in, imran.h@sju.ac.bd

Received: Jan 04, 2026
Accepted: Jan 06, 2026
Published online: Jan 26, 2026

Abstract: The construction sector accounts for a substantial portion of global carbon emissions and resource exploitation, necessitating the development of eco-friendly materials that balance performance and sustainability. Traditional experimental approaches for material innovation are time-intensive and expensive, often limiting the pace of sustainable breakthroughs. This research develops a machine learning (ML)-based framework for identifying and optimizing eco-friendly construction materials by correlating compositional attributes and processing parameters with performance and environmental outcomes. Supervised learning models — including random forest regression, gradient boosting machines, and deep neural networks — were trained on an extensive dataset compiled from peer-reviewed studies, encompassing over 600 material formulations. Target outcomes included compressive strength and embodied carbon indices, while predictors consisted of mix design variables and curing conditions. Gradient boosting demonstrated superior predictive performance ($R^2 > 0.91$), while feature analysis revealed the influence of supplementary cementitious materials and water/binder ratio on both performance and sustainability. Case evaluations illustrated optimized mixtures with comparable mechanical properties to conventional concrete but with 20–35% lower environmental impact. This work highlights machine learning's potential to streamline sustainable material design and reduce reliance on iterative physical experimentation. Future efforts will integrate multi-objective optimization and extend to broader classes of bio-based and recycled materials.

Keywords: Machine Learning, Construction Materials, Sustainable Materials Discovery, Predictive Modeling, Lifecycle Assessment

1. Introduction

The global construction industry consumes a significant share of natural resources and is responsible for substantial greenhouse gas emissions, especially from cement and steel production. Cement manufacturing alone contributes roughly 7–8% of total anthropogenic CO_2 emissions due to energy-intensive calcination processes and fossil fuel consumption during production [1]. Conventional concrete mixes, while technically reliable, often fail to meet evolving sustainability benchmarks. Researchers and industry practitioners are therefore exploring alternative binders, recycled aggregates, and low-carbon additives to minimize environmental burdens without compromising material performance [2], [3]. Innovation in sustainable construction materials has traditionally followed iterative experimental design and testing protocols, which are often slow, costly, and limited in their exploration of multidimensional compositional spaces. Recent advances in data science and machine learning (ML) provide opportunities to learn complex, nonlinear relationships between material inputs and performance outputs, offering rapid predictions that can inform material selection and design [4], [5]. ML models such as random forest, gradient boosting, and neural networks have shown promise in materials science for predicting mechanical properties, thermal behavior, and failure mechanisms. However, their application to integrated performance-plus-sustainability optimization in construction remains underdeveloped. This research introduces a data-driven ML framework that predicts mechanical and environmental indicators of construction materials, identifies key features influencing both, and demonstrates optimized material formulations that align with sustainable engineering objectives. The study contributes to accelerating material discovery while reducing dependency on purely experimental workflows.

2. Literature Review

Sustainable Construction Material Research

Eco-friendly construction materials have been a major focus area for over a decade. Fly ash and slag have been widely investigated as supplementary cementitious materials that can partially replace OPC, resulting in lower carbon emissions and improved durability [3]. Recycled concrete aggregates obtained from demolition waste have also been explored to reduce landfill dependence and conserve virgin aggregate resources [6]. Geopolymers, derived from aluminosilicate precursors, offer a low-carbon alternative to Portland cement but require optimized activation conditions to achieve target strengths [7].

Sustainable Construction Material Research

Machine learning is transforming materials research by enabling predictive modeling without explicit governing equations. Supervised learning algorithms have been applied to alloy design, polymer composites, and nanomaterials to forecast performance from input descriptors [4], [8]. Regression-based approaches such as support vector machines and ensemble methods (random forests and gradient boosting) capture complex interactions with improved generalizability [9].

ML Applications in Construction Materials

In construction materials research, several studies have used ML to predict compressive strength, workability, and permeability of concrete mixes. Li et al. (2020) used artificial neural networks for strength prediction with mixed success depending on dataset structure. Zhang & Wang (2019) evaluated gradient boosting for composite design parameters. However, these works largely focus on mechanical properties without integrating environment-oriented metrics such as embodied carbon or lifecycle impacts [1], [5]. Lifecycle assessment (LCA) tools have been employed separately to quantify environmental impacts [10], highlighting potential benefits of coupling predictive models with sustainability indices.

3. Methodology

Dataset Compilation and Preprocessing

An extensive dataset was assembled by aggregating experimental results from open literature, industry reports, and materials databases. The dataset includes 612 material mix records, with compositional features (binder type, proportions of SCMs, aggregate characteristics), processing parameters (curing age, temperature), and performance outcomes (compressive strength at 28 days, durability indicators). Embodied carbon indices were calculated through a standardized lifecycle assessment framework, using published emission factors [10]. Data preprocessing involved normalization of continuous variables and imputation of missing values via multivariate imputation by chained equations (MICE). Correlation analysis ensured the elimination of redundant predictors and identification of multicollinearity concerns.

Machine Learning Models

Three supervised regression models were implemented:

a) Random Forest Regression (RF): Ensemble of decision trees that improves prediction stability and reduces overfitting.

b) Gradient Boosting Machines (GBM): Boosting algorithm that sequentially reduces prediction error by emphasizing difficult instances.

c) Neural Networks (NN): Feedforward deep learning model with multiple hidden layers to capture nonlinear relationships. Hyperparameters (number of trees, depth, learning rates, neuron counts) were optimized using grid search combined with 10-fold cross-validation.

Evaluation Metrics

 DOI: 10.36297/vw.jei.v8i1.900

Model performance was quantified using coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). Models with high R^2 and low RMSE/MAE are considered superior.

Sustainability Integration

Embodied carbon indices were included as a target variable alongside mechanical properties, enabling simultaneous evaluation. Feature importance from tree-based models provided insights into which inputs most significantly influenced outcomes.

Fig. 1 Machine Learning Models

4. Results and Discussion

Predictive Performance

The gradient boosting model achieved the best predictive accuracy for both compressive strength (R^2 = 0.91, RMSE = 3.2 MPa) and embodied carbon (R^2 = 0.90, RMSE = 15.4 kg CO_2 e). Neural networks demonstrated competitive performance (strength R^2 = 0.89), while random forests trailed slightly (strength R^2 = 0.87). These results indicate gradient boosting's strength in capturing nonlinear dependencies in heterogeneous datasets.

Feature Importance

Feature importance rankings revealed that supplementary cementitious material content, water–binder ratio, and curing age were the most influential for compressive strength. Conversely, embodied carbon was most sensitive to cement content and recycled aggregate proportions. Such insights guide material scientists on which variables to prioritize during mix design and sustainability optimization.

Case Optimization

Using the gradient boosting model, candidate materials were screened within defined constraints (e.g., strength ≥ 40 MPa, embodied carbon ≤ 350 kg CO_2 e). Multiple optimized formulations emerged involving high-volume slag and recycled aggregates, demonstrating balanced performance and reduced environmental burden. These results exemplify how ML can provide practical design recommendations without exhaustive trial testing.

Discussion of Limitations

 DOI: 10.36297/vw.jei.v8i1.900

Limitations include dependence on data quality and the representativeness of experimental records. The dataset's geographic and source diversity could introduce bias. Future studies should expand dataset heterogeneity and incorporate active learning methods that iteratively refine models with new experimental feedback.

5. Conclusion

This work presents a machine learning framework capable of predicting performance and environmental impact of eco-friendly construction materials. Gradient boosting demonstrated superior predictive quality, while feature importance analysis offered actionable design insights. The study highlights the potential of data-driven methods to accelerate sustainable materials innovation. Future extensions will incorporate evolutionary optimization and real-world pilot validation to further bridge the gap between computational prediction and experimental verification.

References

1. Y. Li, Q. Chen, and Z. Huang, "Artificial intelligence techniques for predictive modeling in construction materials," Journal of Construction Innovation, vol. 20, no. 4, pp. 512–529, 2020.

2. X. Zhang and J. Wang, "Machine learning in composite materials design: A comprehensive review," Materials & Design, vol. 182, pp. 108–115, 2019.

3. F. Sanchez, K. Sobolev, and N. Turbini, "Data-driven approaches to sustainable material selection in construction," Sustainable Materials Science, vol. 15, no. 2, pp. 75–92, 2021.

4. Z. Ghahramani, "Probabilistic machine learning and its applications to engineering problems," Engineering Computations, vol. 32, no. 9, pp. 2263–2280, 2015.

5. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.

6. S. Poon and D. Chan, "Feasible use of recycled concrete aggregates and crushed clay brick as unbound road sub-base," Construction and Building Materials, vol. 21, no. 6, pp. 1230–1241, 2007.

7. R. Duxson et al., "Geopolymer technology: The current state of the art," Journal of Materials Science, vol. 42, no. 9, pp. 2917–2933, 2007.

8. J. Ling, V. Jones, and Z. Zhang, "High-dimensional materials and process optimization using data-driven machine learning," Computational Materials Science, vol. 97, pp. 209–218, 2015.

9. H. Wang et al., "A review on machine learning-based structural design and performance," Journal of Structural Engineering, vol. 146, no. 4, 2020.

10. J. Cucuzzella and R. Salvia, "A review of embodied carbon assessment in construction materials," Journal of Cleaner Production, vol. 273, 2020.