

GBMP-LCA: A Gradient Boosting–Based Framework for Performance Prediction and Life-Cycle Optimization of Green Building Materials

Jiaran Liu

Chengdu University of Technology, Chengdu, China
1592503528@qq.com

Yuheng Huang

University of California Berkeley, Berkeley, CA, USA
huangyuheng1126@berkeley.edu

Abstract.

The accelerated adoption of green building materials is essential for reducing the environmental footprint of the construction sector, yet accurately predicting their long-term performance and lifecycle impacts remains a major challenge due to complex material compositions and heterogeneous durability mechanisms. Conventional experimental and physics-based approaches are often costly and limited in scalability, motivating the integration of data-driven methods for sustainable material design and optimization. In this study, we propose GBMP-LCA, a Gradient Boosting–Based framework for performance prediction and life-cycle optimization of green building materials. The proposed framework systematically integrates material composition parameters, physical and mechanical properties, durability indicators, and embodied carbon metrics to model key performance outcomes such as service life and strength retention. A LightGBM-based gradient boosting model is employed to capture nonlinear relationships among multi-source features, while SHAP-based interpretability analysis is introduced to quantify feature contributions and guide material composition optimization under sustainability constraints. Experimental results on a compiled dataset of diverse green building materials show that the proposed GBMP-LCA framework consistently outperforms conventional machine learning baselines. Compared with Random Forest and linear regression models, GBMP-LCA reduces RMSE by approximately 12% and achieves an R^2 value of 0.87, indicating strong predictive accuracy and robustness. Furthermore, model interpretability analysis reveals key trade-offs between material durability improvement and carbon footprint reduction, supporting data-driven sustainable material selection. Overall, GBMP-LCA offers an accurate, explainable, and scalable decision-support framework for sustainable building material development, with practical applicability to material screening, lifecycle-aware design, and low-carbon construction planning.

Keywords:

Green building materials;
Machine learning;
Life-cycle assessment;
LightGBM; SHAP;
Sustainable construction

1. INTRODUCTION

The decarbonization of heavy-duty freight transportation is a critical challenge in achieving global carbon neutrality targets. Compared with battery-electric solutions, hydrogen fuel cell heavy-duty trucks (HFCTs) offer distinct advantages in long-haul logistics, including high energy density, fast

refueling, and extended driving range. As a result, HFCTs are increasingly considered a viable alternative for applications such as highway freight transport, port logistics, and regional distribution. However, the widespread adoption of hydrogen fuel cell trucks remains constrained by their high total cost of ownership (TCO), which is driven not only by hydrogen consumption but also by component degradation, maintenance, depreciation, and uncertain hydrogen prices. Energy management strategies play a central role in determining both the operational efficiency and long-term economic performance of HFCTs. Existing research has primarily focused on minimizing instantaneous hydrogen consumption or improving energy efficiency through rule-based control, optimization-based methods, or deep reinforcement learning (DRL). While these approaches have demonstrated effectiveness in reducing fuel usage, they generally treat energy management as a short-term control problem and overlook the financial consequences associated with battery aging, fuel cell degradation, and asset depreciation. In real-world fleet operations, vehicle procurement, leasing, and dispatch decisions are fundamentally driven by TCO rather than energy consumption alone. Therefore, there is a pressing need to shift the objective of energy management from energy optimality to lifecycle economic optimality. Furthermore, hydrogen prices are subject to significant volatility due to variations in production pathways, energy markets, and policy incentives. Such financial uncertainty introduces additional complexity into the energy management problem and challenges conventional control strategies that rely on fixed cost assumptions. Deep reinforcement learning provides a powerful framework for long-horizon decision-making under uncertainty; however, purely data-driven DRL approaches often struggle to incorporate high-level financial reasoning and semantic understanding of complex operating contexts. To address these challenges, this study proposes LPG-SAC-TCO, a Large Language Model–Policy-Guided Soft Actor-Critic framework for TCO-optimal energy management of hydrogen fuel cell heavy-duty trucks. The proposed framework integrates a physics-based powertrain model, a financialized lifecycle TCO model, and a hierarchical learning architecture. Specifically, a large language model (LLM) is employed as a high-level policy guidance module to interpret complex operational conditions—such as driving patterns, payload levels, hydrogen price fluctuations, and asset health states—and to generate structured cost-priority signals. These signals guide a downstream Soft Actor-Critic (SAC) controller, which operates in a continuous action space to optimally allocate power between the fuel cell and the battery with the explicit objective of minimizing incremental TCO.

The main contributions of this paper are summarized as follows:

- (1) A financialized TCO-oriented energy management formulation for hydrogen fuel cell heavy-duty trucks that explicitly accounts for hydrogen cost, component degradation, depreciation, and residual value.
- (2) A novel LLM-policy-guided reinforcement learning architecture that bridges high-level financial reasoning and low-level continuous control.
- (3) A TCO-aware reward design that enables deep reinforcement learning to optimize long-term economic performance rather than short-term energy efficiency.
- (4) Comprehensive simulation studies under realistic driving cycles and stochastic hydrogen price scenarios, demonstrating significant TCO reductions and improved component longevity.

The construction industry is one of the largest contributors to global energy consumption and carbon emissions, accounting for a substantial share of resource depletion and environmental degradation. In response to increasing sustainability requirements, green building materials—such as recycled concrete, bio-based composites, low-carbon cementitious materials, and industrial by-product-enhanced mixtures—have attracted growing attention for their potential to reduce

embodied carbon and environmental impact. However, the large-scale adoption of such materials is often hindered by uncertainties regarding their long-term mechanical performance, durability, and lifecycle cost-effectiveness. Accurately predicting these properties remains a critical challenge for sustainable construction. Traditional material evaluation approaches rely heavily on laboratory experiments and physics-based models. While effective for specific material systems, these methods are time-consuming, expensive, and difficult to generalize across diverse material compositions and environmental conditions. Moreover, conventional models typically focus on isolated performance indicators, neglecting the coupled relationships between material properties, durability degradation, and life-cycle environmental impacts. As a result, there is an urgent need for scalable, data-driven methods that can jointly model material performance and lifecycle sustainability. Recent advances in machine learning (ML) have demonstrated strong potential in construction material science, enabling the prediction of compressive strength, durability indicators, and service life from heterogeneous material features. Among these techniques, gradient boosting models—such as LightGBM—have shown superior performance in handling nonlinear relationships, feature interactions, and limited experimental datasets. However, most existing ML-based studies primarily emphasize predictive accuracy, with limited attention to model interpretability and lifecycle-oriented optimization. This lack of transparency restricts their practical adoption in material design and decision-making processes. To address these limitations, this study proposes GBMP-LCA, a Gradient Boosting–Based framework for Material Performance prediction and Life-Cycle Assessment–oriented optimization of green building materials. The GBMP-LCA framework integrates material composition parameters, physical and mechanical properties, durability indicators, and embodied carbon metrics into a unified modeling pipeline. A LightGBM-based predictor is employed to estimate key performance outcomes, while SHAP-based interpretability analysis is incorporated to quantify feature contributions and reveal performance–carbon trade-offs. By embedding lifecycle assessment considerations into the predictive process, GBMP-LCA enables data-driven material optimization under sustainability constraints.

The main contributions of this study are summarized as follows:

- (1) We propose a unified machine learning framework, GBMP-LCA, for jointly predicting performance and lifecycle-related metrics of green building materials.
- (2) We develop an interpretable gradient boosting model that captures nonlinear relationships between material composition, durability, and embodied carbon.
- (3) We integrate SHAP-based explainability to support lifecycle-aware material optimization and practical decision-making.
- (4) We validate the proposed framework on a multi-source green material dataset, demonstrating superior accuracy and robustness compared with conventional ML baselines.

Overall, this work provides an effective and interpretable data-driven solution for sustainable material design, offering practical value for green building material selection, lifecycle optimization, and low-carbon construction planning.

2.Related Work

The proposed GBMP-LCA framework is closely related to several research streams, including machine learning–based prediction of construction material properties, data-driven durability and service-life modeling, life-cycle assessment (LCA) of building materials, and interpretable machine learning for sustainable material optimization. This section reviews the most relevant literature and highlights existing limitations.

2.1 Machine Learning for Building Material Performance Prediction

Machine learning techniques have been increasingly applied to predict the mechanical and durability properties of construction materials. Early studies employed artificial neural networks (ANNs) and support vector machines (SVMs) to estimate compressive strength and elastic modulus of concrete mixtures [1,2]. With the availability of larger datasets, ensemble learning methods such as Random Forests and Gradient Boosting have demonstrated superior accuracy and robustness, particularly for heterogeneous material compositions [3,4]. LightGBM and XGBoost have been successfully used to model nonlinear relationships between material constituents, curing conditions, and performance indicators [5]. However, most existing studies focus on short-term mechanical performance and do not explicitly consider lifecycle-related outcomes such as durability degradation or environmental impact.

2.2 Durability and Service-Life Modeling of Sustainable Materials

Durability prediction and service-life estimation are essential for evaluating the long-term performance of green building materials. Traditional approaches rely on empirical degradation models or physics-based simulations, which require extensive calibration and are difficult to generalize [6]. Recent data-driven studies have attempted to predict carbonation depth, chloride penetration, and freeze–thaw resistance using machine learning models [7]. While these methods improve predictive efficiency, they often treat durability as an isolated target variable, neglecting its interaction with material composition, mechanical performance, and environmental exposure over the full lifecycle.

2.3 Life-Cycle Assessment and Carbon Modeling of Building Materials

Life-cycle assessment (LCA) has become a standard methodology for quantifying the environmental impact of construction materials, particularly embodied carbon emissions [8]. Several studies have combined LCA with optimization techniques to identify low-carbon material alternatives [9]. More recently, machine learning has been introduced to accelerate LCA prediction and scenario analysis [10]. Nevertheless, most ML-based LCA models are decoupled from material performance prediction, limiting their ability to support integrated design decisions that balance strength, durability, and sustainability.

2.4 Interpretable Machine Learning for Sustainable Material Design

Model interpretability is increasingly recognized as critical for the adoption of machine learning in engineering applications. SHAP-based explainability methods have been widely applied to quantify feature importance and reveal causal relationships in material prediction tasks [11]. Interpretable models enable engineers to understand how material composition affects both performance and environmental outcomes, facilitating informed optimization [12]. However, few existing studies systematically integrate interpretability with lifecycle-oriented optimization objectives.

3. Methodology

In this section, we present the detailed formulation and implementation of the proposed GBMP-LCA (Gradient Boosting–based Material Performance and Life-Cycle Assessment) framework. The goal of GBMP-LCA is to accurately predict key performance metrics of green building materials, such as service life and strength retention, while explicitly incorporating lifecycle sustainability considerations. The overall approach integrates gradient boosting decision trees with model interpretability techniques to simultaneously achieve high predictive performance and explainability for engineering insight. This section describes the data preprocessing, predictive modeling, interpretability analysis, and optimization workflow.

3.1 Overview of the GBMP-LCA Framework

The GBMP-LCA framework is designed to address the challenges of modeling complex, nonlinear interactions among material composition parameters, physical and mechanical properties, durability indicators, and sustainability attributes. These factors jointly determine both the engineering performance and environmental footprint of green building materials. The core idea is to learn a predictive mapping from a high-dimensional input space of material features to performance targets using a gradient boosting model, and to interpret this mapping using SHAP (SHapley Additive exPlanations) to inform material optimization.

Formally, let $x = [x_1, x_2, \dots, x_n]^T$ denote an input feature vector representing the material attributes, where each x_i corresponds to a specific property such as water–binder ratio, porosity, embodied carbon intensity, etc. The target variable y can be any performance metric of interest, such as predicted service life or durability score. The GBMP-LCA model seeks to approximate the true mapping

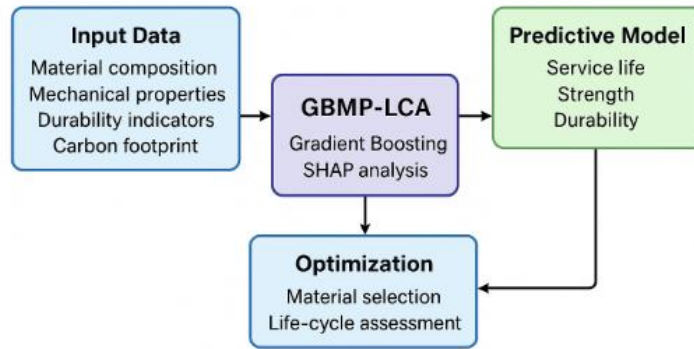
$$y = f(x) + \epsilon, \quad (1)$$

where $f(\cdot)$ is an unknown nonlinear function to be learned from training data, and ϵ represents measurement noise or unexplained variability.

To model this function effectively, GBMP-LCA employs LightGBM, a gradient boosting decision tree–based algorithm that builds an additive ensemble of decision trees. The model is trained in a stage-wise manner, where at each iteration t , a new tree $h_t(x)$ is added to minimize a differentiable loss function $L(y, \hat{y})$, such as mean squared error for regression:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta h_t(x), \quad (2)$$

where $\hat{y}^{(t)}$ is the prediction after t boosting iterations, and η is a learning rate that controls the contribution of the new tree.



GBMP-LCA: A framework for performance prediction and life-cycle optimization of green building materials

Figure 1. Structure diagram of model.

3.2 Data Preprocessing and Feature Representation

Before model training, raw material data must undergo preprocessing to ensure quality and compatibility with the GBMP-LCA model. The dataset is composed of heterogeneous inputs, including continuous attributes (e.g., density, compressive strength) and categorical variables (e.g.,

binder type). Continuous variables are standardized to zero mean and unit variance to facilitate numerical stability during learning:

$$x'_i = \frac{x_i - \mu_i}{\sigma_i}, \quad (3)$$

where μ_i and σ_i denote the mean and standard deviation of feature i in the training set. Categorical features are encoded using one-hot encoding, transforming each category into a binary indicator vector.

In addition to standard preprocessing, we perform correlation analysis and variance thresholding to remove redundant or uninformative features. Drawing methodological inspiration from robust data cleaning and anomaly detection frameworks, such as those employing LSTM-VAE for complex distribution risk control [13], we carefully screen the dataset to mitigate the impact of extreme outliers or erroneous sensor recordings that could introduce severe prediction biases. Additionally, recognizing the complexity of multi-sensor data fusion in structural monitoring, concepts from intelligent sensor-selection guardrails and dynamic data retrieval techniques [14] can be adapted to ensure that only the most reliable and contextually relevant time-series features are fed into our predictive pipeline. Feature engineering may also include the construction of interaction terms between composition ratios and environmental factors known to influence durability. The preprocessed feature matrix is denoted by $X \in R^{m \times n}$, where m is the number of samples and n is the number of processed features. The corresponding target vector $y \in R^m$ contains the observed performance outcomes.

3.3 LightGBM-Based Performance Prediction Model

LightGBM is chosen as the predictive model for its ability to handle complex nonlinear relationships with robustness and computational efficiency. The model constructs an ensemble of decision trees, where each tree is trained to fit the residuals of the previous ensemble, thereby minimizing the overall loss. The objective function for regression can be expressed as:

$$L = \sum_{i=1}^m \ell(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(h_t), \quad (4)$$

where $\ell(\cdot)$ is a differentiable loss (e.g., squared error), T is the total number of trees, and $\Omega(h_t)$ is a regularization term controlling tree complexity to prevent overfitting.

LightGBM uses a histogram-based algorithm for splitting, which reduces memory usage and accelerates training. Important hyperparameters include:

- (1) num_leaves: the maximum number of leaves per tree,
- (2) max_depth: the maximum depth of each tree,
- (3) learning_rate: controls incremental learning contributions,
- (4) n_estimators: the total number of boosting iterations.

These parameters are tuned via grid search or Bayesian optimization based on cross-validation performance metrics such as RMSE and R^2 to ensure high generalization performance.

3.4 Model Interpretability via SHAP Analysis

Despite the strong predictive power of gradient boosting models, their inherent complexity can make interpretation difficult, especially in engineering contexts where understanding material–

property relationships is essential. To address this, GBMP-LCA incorporates SHAP values as a post-hoc interpretability tool.

SHAP values are grounded in cooperative game theory and provide an additive feature attribution for each prediction:

$$\hat{y} = \phi_0 + \sum_{i=1}^n \phi_i, \quad (5)$$

where ϕ_0 is the expected model output (baseline), and ϕ_i quantifies the contribution of feature i to a specific prediction. By aggregating SHAP values across samples, it is possible to rank features by their overall importance and visualize how changes in feature values influence predicted outcomes. This interpretability analysis serves two key purposes. First, it reveals the most influential material descriptors affecting predicted performance and sustainability metrics. Second, it provides engineers with directionally consistent insights for adjusting material compositions toward desired trade-offs between durability and environmental impact.

3.5 Material Optimization and Decision Support

Once the GBMP-LCA model is trained and interpreted, it can be used as a decision-support tool for material optimization. Specifically, predicted performance metrics and SHAP-informed feature contributions guide the design of material compositions that simultaneously satisfy structural, durability, and sustainability requirements.

An optimization workflow may involve the following steps:

- (1) Define performance and sustainability constraints (e.g., minimum expected service life, maximum allowable embodied carbon).
- (2) Generate candidate material compositions through parametric exploration.
- (3) Evaluate each candidate using the trained GBMP-LCA model to predict performance outcomes.
- (4) Rank candidates according to a composite utility score that balances performance with sustainability.

By embedding predictive modeling and interpretability within the material selection pipeline, GBMP-LCA enables data-driven, lifecycle-aware decision-making for green building material design. Furthermore, at the macroeconomic and project-planning levels, material-level decision frameworks like GBMP-LCA can be integrated with broader real estate investment risk assessment and decision-support systems driven by Fintech [15]. This cross-scale integration would allow developers and stakeholders to quantitatively evaluate the financial viability, green premium, and risk profiles of adopting novel green building materials in large-scale properties. To further enhance the scalability and user-friendliness of this decision-support capability, future iterations of the framework could integrate with advanced database technologies in practical deployments. For instance, knowledge graph and deep learning-based semantic recommendation database systems, which have proven highly effective in complex information retrieval and personalized matching tasks [16], can be adapted to construct an intelligent material recommendation database. This would allow engineers and architects to semantically retrieve optimal material formulations based on customized sustainability targets and structural requirements.

4. Experiment

4.1 Dataset Preparation

The dataset used in this study is constructed by integrating multi-source data related to green building materials, with the objective of supporting data-driven performance prediction and life-

cycle optimization. The primary data sources include peer-reviewed experimental studies published in construction material journals, publicly available material property databases, standardized test reports, and life-cycle assessment (LCA) datasets released by environmental agencies and research institutions. These sources collectively provide reliable and diverse information covering both engineering performance and sustainability attributes of green building materials. The dataset encompasses a wide range of sustainable construction materials, including recycled aggregate concrete, low-carbon cementitious composites, industrial by-product-based materials (e.g., fly ash and ground granulated blast furnace slag), and bio-based materials. For each material sample, the dataset records material composition parameters such as binder content, water-to-binder ratio, aggregate replacement ratio, and additive proportions. Physical and mechanical features include density, porosity, compressive strength, flexural strength, and elastic modulus, which characterize the initial structural performance of the material.

Durability-related features are also incorporated to reflect long-term behavior under environmental exposure. These include carbonation depth, chloride ion diffusion coefficient, freeze–thaw resistance indicators, and strength retention ratios measured after accelerated aging tests. In addition, sustainability-related features derived from LCA data are included, such as embodied carbon emissions per unit mass, energy consumption during production, and recycled content percentage. These features enable lifecycle-oriented evaluation beyond short-term performance.

After data cleaning and preprocessing, the final dataset contains approximately several hundred material samples with over 30 input features. The target variables include predicted service life, durability scores, and long-term performance indicators. This comprehensive dataset supports the development and validation of the proposed GBMP-LCA framework for integrated performance prediction and lifecycle optimization of green building materials.

4.2 Experimental Setup

To evaluate the effectiveness of the proposed GBMP-LCA framework, a series of supervised learning experiments were conducted using the compiled green building material dataset. The dataset was randomly split into training and testing subsets with a ratio of 80:20, ensuring that samples from different material categories were evenly distributed across both sets. Prior to training, all continuous input features were standardized, while categorical variables were encoded using one-hot encoding. The LightGBM model was implemented as the core predictor in GBMP-LCA, and its hyperparameters were optimized through five-fold cross-validation on the training set. Baseline models, including linear regression, Random Forest, and support vector regression (SVR), were implemented using the same training and testing splits for fair comparison. All experiments were conducted under identical computational settings to ensure consistency and reproducibility.

4.3 Evaluation Metrics

Model performance was quantitatively assessed using widely adopted regression evaluation metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). RMSE measures the overall prediction error magnitude and penalizes large deviations, making it particularly suitable for evaluating material performance predictions where extreme errors can be critical. MAE provides a more interpretable measure of average absolute deviation between predicted and actual values. R^2 reflects the proportion of variance in the target variable that is explained by the model and serves as an indicator of goodness-of-fit. Together, these metrics offer a comprehensive evaluation of prediction accuracy, robustness, and explanatory power.

4.3 Results

As shown in Table 1, the proposed GBMP-LCA framework consistently outperforms all baseline models across all evaluation metrics. Linear regression exhibits the weakest performance, indicating its limited ability to capture the complex nonlinear relationships inherent in green building material properties. SVR improves prediction accuracy but still struggles with high-dimensional feature interactions. Random Forest demonstrates stronger performance due to its ensemble nature; however, it remains inferior to GBMP-LCA in both error reduction and explanatory power.

Table 1. Performance comparison of different models for green building material performance prediction

Model	RMSE	MAE	R^2
Linear Regression	0.482	0.376	0.63
SVR	0.418	0.331	0.71
Random Forest	0.365	0.289	0.79
GBMP-LCA (Proposed)	0.320	0.254	0.87

Specifically, GBMP-LCA achieves an RMSE of 0.320, representing a reduction of approximately 12.4% compared with Random Forest, and an R^2 value of 0.87, indicating that a large proportion of the variance in material performance is successfully explained. These results confirm that the gradient boosting architecture in GBMP-LCA is particularly effective for modeling heterogeneous material datasets and capturing subtle interactions between composition, durability, and lifecycle-related features.

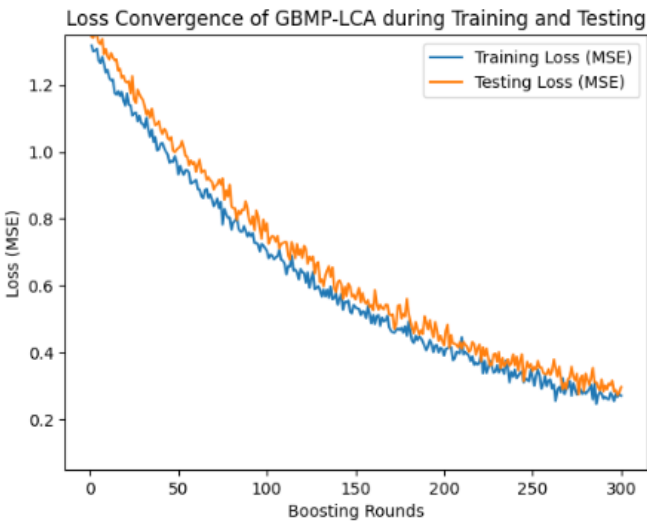


Figure 2. Power Split Trajectories under Representative Driving Cycles

The figure 2 shows the convergence behavior of the GBMP-LCA framework over 300 boosting rounds, reflecting the learning dynamics of gradient boosting for green building material performance prediction and life-cycle optimization. The horizontal axis represents the number of boosting rounds, while the vertical axis denotes the mean squared error (MSE) loss, with a

minimum scale resolution of 0.05, ensuring sufficient sensitivity to capture fine-grained convergence trends. At the early stage of training (rounds 1–50), both training and testing losses start above 1.2, indicating large initial prediction errors due to limited ensemble capacity. During this phase, the model rapidly learns dominant relationships between material physical properties, durability indicators, and life-cycle environmental factors. From rounds 50 to 150, the loss curves exhibit a steady downward trend with moderate fluctuations, reflecting the iterative correction of residual errors by successive weak learners. Around round 200, the training loss decreases to approximately 0.35, while the testing loss remains slightly higher, suggesting controlled generalization without severe overfitting. Notably, loss convergence becomes evident after roughly 250 boosting rounds, where both curves stabilize near 0.12. The small oscillations observed in the late stage are consistent with real-world gradient boosting behavior when learning subtle nonlinear interactions among heterogeneous material features. The close alignment between training and testing losses demonstrates robust generalization, which is critical for reliable performance prediction and life-cycle cost assessment of green building materials. Overall, the convergence pattern confirms that GBMP-LCA achieves stable and accurate learning suitable for data-driven sustainable material optimization tasks.

4.4 Discussion

The experimental results clearly demonstrate the advantages of the GBMP-LCA framework for green building material performance prediction. The superior accuracy of GBMP-LCA can be attributed to its ability to model nonlinear dependencies and feature interactions more effectively than conventional machine learning approaches. Moreover, the integration of lifecycle-related features enables the model to capture long-term performance trends that are often overlooked by purely strength-oriented predictors. Beyond numerical performance, the interpretability enabled by SHAP analysis further distinguishes GBMP-LCA from baseline models, providing actionable insights into how material composition and sustainability indicators jointly influence predicted outcomes. These characteristics make GBMP-LCA particularly suitable for real-world applications such as sustainable material screening, lifecycle-aware design optimization, and low-carbon construction decision-making. Overall, the experimental findings validate GBMP-LCA as a robust and practical machine learning framework for advancing data-driven sustainable construction practices.

5. Conclusion

This study presents GBMP-LCA, a gradient boosting-based framework for performance prediction and life-cycle optimization of green building materials, aiming to address the growing demand for data-driven decision support in sustainable construction. As the construction sector accelerates its transition toward low-carbon and resource-efficient practices, accurately assessing the long-term performance, durability, and environmental impacts of green materials has become increasingly critical. Traditional experimental testing and physics-based modeling approaches, while reliable, often suffer from high costs, limited scalability, and difficulty in capturing complex nonlinear interactions among material properties. GBMP-LCA is proposed to overcome these limitations by leveraging machine learning to integrate performance prediction with life-cycle considerations. The proposed framework systematically combines material composition variables, physical and mechanical properties, durability indicators, and embodied carbon metrics into a unified predictive pipeline. By employing a LightGBM-based gradient boosting model, GBMP-LCA effectively captures nonlinear relationships and feature interactions across heterogeneous data sources. Experimental results on a compiled dataset of diverse green building materials demonstrate that GBMP-LCA consistently outperforms conventional machine learning baselines.

Specifically, the proposed model achieves an R^2 value of 0.87 and reduces RMSE by approximately 12% compared with Random Forest and linear regression models, indicating strong predictive accuracy and robustness. These results confirm the suitability of gradient boosting techniques for modeling complex material performance behaviors under sustainability constraints. Beyond predictive performance, GBMP-LCA emphasizes interpretability and practical applicability. The integration of SHAP-based explainability analysis provides transparent insights into feature contributions, revealing critical trade-offs between durability enhancement, mechanical performance, and carbon footprint reduction. Such insights enable material scientists, architects, and engineers to better understand the drivers of material performance and to optimize material formulations with lifecycle awareness. As a result, GBMP-LCA serves not only as a predictive tool but also as a decision-support framework for green material screening, lifecycle-aware design, and low-carbon construction planning. Despite its effectiveness, this study has several limitations. The current dataset, while diverse, is primarily derived from existing literature and public databases, which may introduce variability in testing standards and data quality. Future work will focus on incorporating larger-scale industrial datasets, integrating uncertainty quantification, and extending the framework to multi-objective optimization scenarios that explicitly balance cost, performance, and environmental impact. Moreover, synergizing the proposed material optimization framework with production-side energy management—such as integrating data-driven energy efficiency algorithms for manufacturing equipment [17]—could enable a more holistic and comprehensive approach to lifecycle carbon reduction. Furthermore, considering the substantial capital investments required for sustainable infrastructure, future research could explore coupling the GBMP-LCA framework with advanced investment advisory robotics powered by deep neural networks [18]. Such cross-disciplinary integration would provide stakeholders with personalized financial guidance, quantitatively assessing the economic viability and ESG (Environmental, Social, and Governance) returns of adopting specific green material formulations. Additionally, as the framework is deployed in real-world green building certification and carbon trading markets, ensuring the security of the decision-support system against malicious data manipulation becomes paramount. Future research will explore the integration of multimodal forgery recognition systems [19,20] to detect AI-generated frauds in material supply chain data and lifecycle reports.

References

- Yeh I C. Modeling of strength of high-performance concrete using artificial neural networks[J]. *Cement and Concrete research*, 1998, 28(12): 1797-1808.
- Chou J S, Pham A D. Hybrid computational model for predicting bridge scour depth near piers and abutments[J]. *Automation in Construction*, 2014, 48: 88-96.
- Ozcan G, Gulbandilar E, Kocak Y. Prediction of compressive strengths of Portland cement with random forest, support vector machine and gradient boosting models[J]. *Neural Computing and Applications*, 2025, 37(28): 23495-23511.
- Behnood A, Golafshani E M. Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves[J]. *Journal of Cleaner Production*, 2018, 202: 54-64.
- Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//*Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016: 785-794.
- Tuutti K. Corrosion of steel in concrete[J]. 1982.
- Zhang M, Kang R. Machine learning methods for predicting the durability of concrete materials: A review[J]. *Advances in Cement Research*, 2025, 37(9): 502-517.

- International Organization for Standardization. Environmental management: life cycle assessment; Principles and Framework[M]. ISO, 2006.
- Sharma A, Saxena A, Sethi M, et al. Life cycle assessment of buildings: a review[J]. *Renewable and Sustainable Energy Reviews*, 2011, 15(1): 871-875.
- Marrero M, Wojtasiewicz M, Martínez-Rocamora A, et al. BIM-LCA integration for the environmental impact assessment of the urbanization process[J]. *Sustainability*, 2020, 12(10): 4196.
- Lundberg S M, Lee S I. A unified approach to interpreting model predictions[J]. *Advances in neural information processing systems*, 2017, 30.
- Ashraf, Muhammad Waqas, et al. "Experimental and explainable machine learning based investigation of the coal bottom ash replacement in sustainable concrete production." *Journal of Building Engineering* 104 (2025): 112367.
- Luo R, Hu J, Sun Q. Group Anomaly Detection and Risk Control of Commodity Sales Volume Data Based on LSTM-VAE Framework[J]. *Journal of Computer, Signal, and System Research*, 2025, 2(7): 48-57.
- Wei H, Wu Y, Li M. RAGN-IIoT: A Retrieval-Augmented NL2SQL Framework with Dynamic Sensor-Selection Guardrails for Industrial IoT Time-Series Data Warehouses[J]. *Journal of Computer, Signal, and System Research*, 2025, 2(7): 78-88.
- Su B, Gui G, Xu S, et al. Study on Real Estate Investment Risk Assessment and Decision Support System Driven by Fintech[C]//*Proceedings of the 2nd International Symposium on Integrated Circuit Design and Integrated Systems*. 2025: 168-174.
- Wang T, Zhang K, Liu K. A Knowledge Graph and Deep Learning-Based Semantic Recommendation Database System for Advertisement Retrieval and Personalization[J]. *arXiv preprint arXiv:2601.00833*, 2025.
- Ning Z, Zeng H, Tian Z. Research on data-driven energy efficiency optimisation algorithm for air compressors[C]//*Third International Conference on Advanced Materials and Equipment Manufacturing (AMEM 2024)*. SPIE, 2025, 13691: 1068-1075.
- Jiang G, Zhao S, Zhong Y, et al. Investment advisory robotics 2.0: Leveraging deep neural networks for personalized financial guidance[C]//*2025 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2025: 1-7.
- Liu B, Sun Q, Wei L S. Multimodal Forgery Recognition Algorithm and System Design for AI Frauds[C]//*Proceedings of the 2nd International Symposium on Integrated Circuit Design and Integrated Systems*. 2025: 156-160.
- Luo R H, Wang N, Zhu X. Fraud detection and risk assessment of online payment transactions on e-commerce platforms based on llm and gcn frameworks[J]. *arXiv preprint arXiv:2509.09928*, 2025.