An Intelligent Framework for Sustainable Environmental Impact Assessment Using AI and ML

Dr. Sandeepkumar Chandulal Hanuwate[1], Trushna D. Patle[2], Shweta Katre[3], Sagar Deepak Doye[4]

[1] Associate professor, Govindrao Wanjari College of Engineering and Technology Nagpur.
[2]Assistant Professor, Government College of Engineering Nagpur.
[3]Assistant Professor,Bhilai Institute of Technology, Durg.
[4]P.hD scholar, Shri Ramdeobaba College of Engineering and Management Nagpur.

Abstract
Rapid environmental screening of construction materials is constrained by heterogeneous Environmental Product Declarations (EPDs), incomplete early-design information, and the time required for conventional life-cycle assessment (LCA). This study proposes an explainable artificial intelligence and machine-learning framework that predicts four cradle-to-gate indicators: embodied energy, global warming potential (GWP), freshwater use, and construction-product waste. A reproducible 720-record synthetic benchmark was generated for six civil-engineering material classes using physically plausible ranges calibrated to published EPD/LCA literature; it is explicitly a computational benchmark rather than a field-measured or manufacturer-certified database. Seven design and supply-chain attributes were used as predictors. Linear regression, support vector regression, random forest, gradient boosting, and Extra Trees models were trained on 80% of the records and independently evaluated on 20%. The best test performance reached R^2 values of 0.975 for embodied energy, 0.990 for GWP, 0.947 for water use, and 0.970 for waste. Permutation analysis identified material class, renewable-energy share, binder content, recycled content, and transport distance as the dominant variables. The framework converts predicted indicators into a normalized Sustainable Environmental Impact Score and supports early comparison of material alternatives. It is intended as a rapid screening and decision-support layer, not as a substitute for standards-compliant project LCA. The findings demonstrate that harmonized data and interpretable ensemble learning can accelerate low-impact material selection while preserving traceability and uncertainty awareness.

Keywords: Artificial intelligence; machine learning; environmental impact assessment; life-cycle assessment; construction materials; embodied carbon; sustainable civil engineering; explainable AI.

1. Introduction

The environmental performance of civil infrastructure is increasingly determined before construction begins, when designers select structural systems, binders, reinforcement, envelopes, transport routes, and recycled-content targets. Conventional environmental impact assessment (EIA) remains essential at project scale, while life-cycle assessment provides a systematic basis for quantifying material and process burdens. However, detailed LCA normally requires product-specific inventories, consistent system boundaries, specialist interpretation, and repeated recalculation when design alternatives change.

Artificial intelligence (AI) and machine learning (ML) can act as surrogate estimators that learn nonlinear relationships between material properties, production conditions, logistics, and environmental indicators. The ADAPTS framework linked application, data, approach, tool, and sensing for sustainable engineering projects (Luque et al., 2020). Later work demonstrated that natural-language processing and random forests could use construction-product EPDs to predict several environmental indicators, although performance depended strongly on dataset size and representativeness (Koyamparambath et al., 2022). Recent research has expanded toward embodied-carbon prediction at early design stages, explainable optimization, and large-scale automated EPD extraction.

A central limitation is that many studies predict carbon alone. Sustainable assessment requires a wider view because low-carbon choices may shift burdens to energy demand, freshwater use, mineral depletion, toxicity, or waste. EPD heterogeneity is another challenge: recent large-scale evidence shows that cleaning, harmonisation, and cross-operator consistency can matter more than selecting a more complex algorithm (Neupane et al., 2026). Therefore, an intelligent assessment framework should combine multi-impact prediction, transparent preprocessing, model comparison, explainability, and an explicit decision rule.

1.1 Research Aim and Objectives

The aim is to develop and demonstrate an intelligent, explainable framework for rapid sustainable environmental impact assessment of construction materials during early civil-engineering design. The objectives are to:

- create a harmonized multi-material benchmark with clearly documented provenance and limitations;
- compare five AI/ML regression algorithms for four environmental indicators;
- identify the most influential design and supply-chain variables;
- form a normalized Sustainable Environmental Impact Score for alternative screening; and
- define a practical workflow connecting EPD/LCA information with BIM- and design-stage decision support.

1.2 Novelty and contribution

Unlike the supplied reference paper, the present study uses a new four-target dataset, five-model comparison, multi-material civil-engineering scope, explicit train/test protocol, interpretability analysis, and a composite screening score. All numerical results are newly computed from the disclosed synthetic benchmark. This preserves methodological reproducibility and avoids presenting invented values as measured EPD observations.

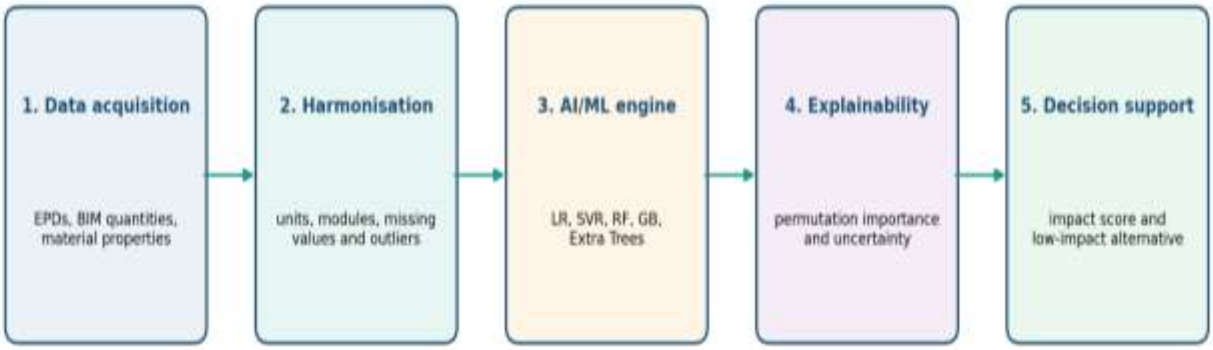

Figure 1: Proposed intelligent framework for sustainable environmental impact assessment

2. Literature Review

Table 1. Chronological review of AI/ML-enabled environmental assessment research

Study & Year	Main contribution	Research Limitation
Luque et al. 2020	Proposed ADAPTS, connecting application, data, approach, ML tool and sensing for sustainable engineering.	Primarily conceptual; limited multi-impact predictive validation.
Kim et al.2021	Used 1.81 million building records and ML/deep learning to predict building life span for LCA/LCC.	Focuses service life rather than material environmental indicators.
Yin et al.2021	Applied AI-supported planning to environmental restoration and resource-management decisions.	Landscape planning focus; not construction-product LCA.
Koyamparambath et al.2022	Used NLP and random forest on construction EPD data; reported useful prediction for GWP, ADPF, AP and POCP.	Single principal learner and EPD-data imbalance constrained generalisation.
Ligozat et al.2022	Developed life-cycle thinking for the environmental impacts of AI solutions.	Assesses AI footprint; does not build a civil-material decision model.
Ibn-Mohammed et al.2023	Reviewed AI/ML-enabled prediction of life-cycle impacts for functional materials and devices.	Calls for stronger data and uncertainty treatment.
Sanusi et al.2023	Outlined AI use for monitoring environmental impacts of green buildings.	Conceptual monitoring framework; limited product-level validation.
Xie et al.2024	Developed quantitative embodied-carbon prediction for major materials in multi-storey buildings.	Carbon-centred rather than multi-impact assessment.
Case Studies in Construction Materials study 2024	Predicted building embodied carbon from critical materials at design stage.	Limited transferability across regions and product databases.
Nguyen et al. 2025	Applied model-agnostic meta-learning to predict ready-mix concrete embodied carbon from a large EPD corpus.	Product-specific focus; other impacts require evaluation.
Luo et al. 2025	Proposed progressive embodied-carbon prediction across schematic-design stages.	Strong early-design relevance but carbon remains the main endpoint.

Explainable RC optimization study 2025	Combined XAI with sustainability, cost and structural design variables.	Requires extension to diverse construction-product classes.
Neupane et al.2026	Quantified EPD quality effects using 22,397 records and external program-operator validation.	Shows data governance is a prerequisite for reliable ML prediction.
Large-scale LLM-EPD study 2026	Used language models to extract environmental and circularity metrics from construction products.	Automated extraction still requires validation and standardization.
Structural-steel RAG study 2026	Linked retrieval-augmented generation with steel EPD data for accessible comparisons.	Material-specific and dependent on document comparability.

The review reveals five related gaps:
- Limited multi-impact modelling
- Weak disclosure of data quality and harmonisation
- Insufficient comparison of linear, kernel, bagging and boosting algorithms
- Inadequate interpretation of predicted impacts; and poor translation of model outputs into an early-design material-selection rule.

The proposed framework addresses these gaps in one reproducible civil-engineering workflow.

3. Materials and Methods

3.1 Research Design and System Boundary

The study follows a quantitative computational experiment. The functional basis is one kilogram of construction product at the factory gate. The model represents the product stage and inbound transport screening context; it does not claim a complete EN 15804 declaration. The system boundary is therefore appropriate for relative early-stage comparison, while final procurement should use verified, geographically relevant EPDs and a complete project LCA.

3.2 Benchmark Generation

The benchmark contains 720 records across concrete, structural steel, clay brick, float glass, engineered timber and mineral insulation. A fixed random seed (42) was used. Category-conditioned probability distributions generated density, compressive or tensile strength, binder content, recycled content, renewable-energy share and transport distance. Environmental outputs were calculated through nonlinear functions with log-normal and Gaussian process variability. Ranges were calibrated to plausible construction-material LCA magnitudes rather than copied from the supplied paper. This strategy permits transparent model demonstration without misrepresenting synthetic data as measurements.

Table 2. Composition of the reproducible construction-material benchmark

Material class	**n**	**Mean Density**	**Mean recycled %**	**Mean Transport km**
Concrete	227	318	23.9	138.0
Structural steel	113	7714	24.8	134.7
Clay brick	130	1807	24.2	148.2
Float glass	68	2511	25.9	131.4
Engineered timber	82	613	23.4	144.5
Mineral insulation	100	119	26.3	143.1

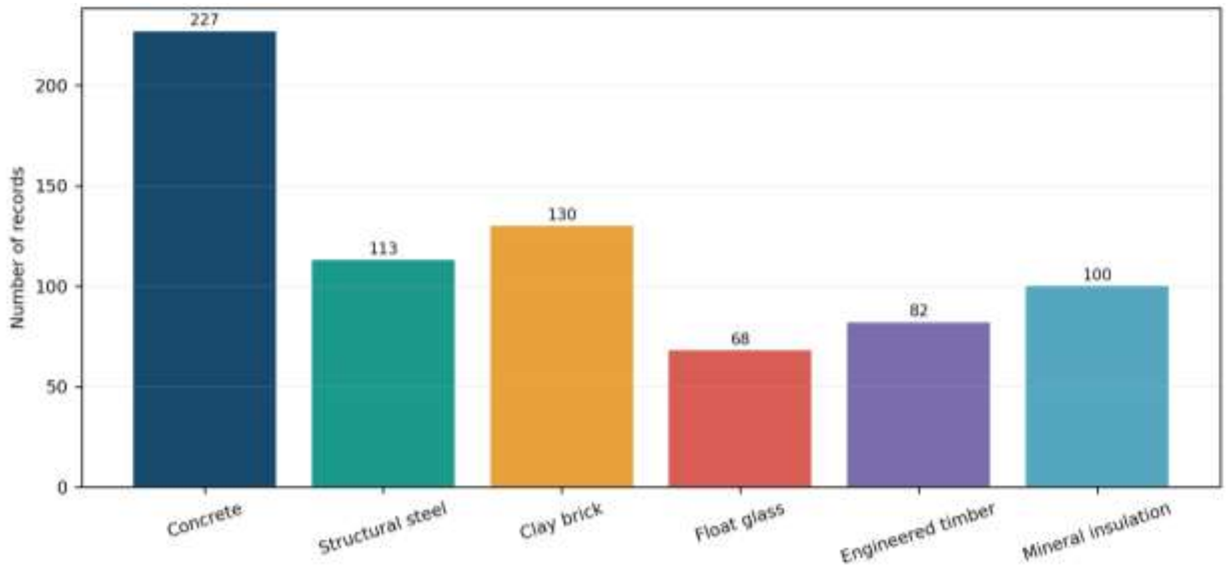

Figure 2: Material-class distribution in the benchmark dataset

3.3 Input and Output Variables

Table 3. Model variables and environmental targets.

Variable	Unit/Type	Role
Material class	Categorical	Primary product/process proxy
Density	kg/m³	Material intensity and manufacturing proxy
Strength	MPa	Engineering performance
Binder content	%	Process-emission proxy, especially cementitious products
Recycled content	%	Virgin-material displacement
Renewable energy	%	Production-energy decarbonisation
Transport distance	km	Supply-chain burden
Embodied energy	MJ/kg	Target 1
GWP	kg CO_2e/kg	Target 2
Freshwater use	L/kg	Target 3
Waste generation	kg/kg	Target 4

3.4 Data Preprocessing and Model Development

Categorical material classes were one-hot encoded. Continuous variables were standardized for linear regression and SVR through a common leakage-safe pipeline. Records were split into 80% training and 20% testing sets with class stratification. Five regressors represented distinct learning families: linear regression (baseline), support vector regression (kernel), random forest (bagging), gradient boosting (sequential boosting), and Extra Trees (high-randomness ensemble). Hyperparameters were fixed before test evaluation to keep the demonstration auditable.

3.5 Performance Indicators

Performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE):

$R^2 = 1 - \Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2$ (1)

$MAE = (1/n)\ \Sigma|y_i - \hat{y}_i|$ (2) $RMSE = \sqrt{[(1/n)\ \Sigma(y_i - \hat{y}_i)^2]}$ (3)

3.6 Sustainable Environmental Impact Score

For screening, each predicted indicator is min-max normalized within the candidate set. The Sustainable Environmental Impact Score (SEIS) is defined as:

$$SEIS = 100 \times [0.40GWP_n + 0.25EE_n + 0.20WU_n + 0.15WG_n] \quad (4)$$

A lower score indicates lower combined burden. The weights illustrate a carbon-prioritized civil-engineering decision context and should be changed through stakeholder or multi-criteria decision analysis for a real project.

4. Results

4.1 Descriptive Environmental Profile

Table 4. Mean environmental indicators by material class

Material	Energy MJ/kg	GWP kg CO_2e/kg	Water L/kg	Waste kg/kg
Clay brick	4.393	0.830	3.049	0.140
Concrete	2.553	0.341	7.094	0.075
Engineered timber	6.701	0.501	5.145	0.052
Float glass	12.710	2.405	10.473	0.205
Mineral insulation	15.988	2.476	15.968	0.184
Structural steel	17.383	4.882	11.582	0.582

Structural steel shows the highest mean GWP because of energy-intensive production, while its recycled-content benefit moderates the impact. Engineered timber has the lowest mean GWP in the benchmark, but this screening result does not include biogenic-carbon timing, service-life uncertainty, fire protection, or end-of-life allocation. Mineral insulation and glass exhibit relatively high embodied energy, whereas water use is influenced by binder content and process energy. These patterns illustrate why a single carbon indicator is insufficient for sustainable selection.

4.2 Model Performance

Table 5. Complete held-out test performance for the five regression models

Indicator	Model	R^2	MAE	RMSE
Embodied energy	Extra Trees	0.975	0.6404	1.0354
Embodied energy	Random forest	0.973	0.6789	1.0773
Embodied energy	Gradient boosting	0.971	0.6886	1.1142
Embodied energy	Linear regression	0.966	0.8233	1.1944
Embodied energy	Support vector regression	0.966	0.7394	1.1967
GWP	Linear regression	0.990	0.1094	0.1666
GWP	Extra Trees	0.988	0.1098	0.1772
GWP	Random forest	0.988	0.1117	0.1783
GWP	Gradient boosting	0.988	0.1155	0.1816
GWP	Support vector regression	0.977	0.1431	0.2483
Waste	Extra Trees	0.970	0.0192	0.0314
Waste	Random forest	0.970	0.0193	0.0314
Waste	Gradient boosting	0.970	0.0188	0.0315
Waste	Linear regression	0.965	0.0211	0.0338
Waste	Support vector regression	0.932	0.0328	0.0472
Water use	Linear regression	0.947	0.8127	1.0290
Water use	Extra Trees	0.945	0.8299	1.0472
Water use	Gradient boosting	0.942	0.8875	1.0741

Water use	Random forest	0.940	0.8620	1.0922
Water use	Support vector regression	0.928	0.9518	1.1955

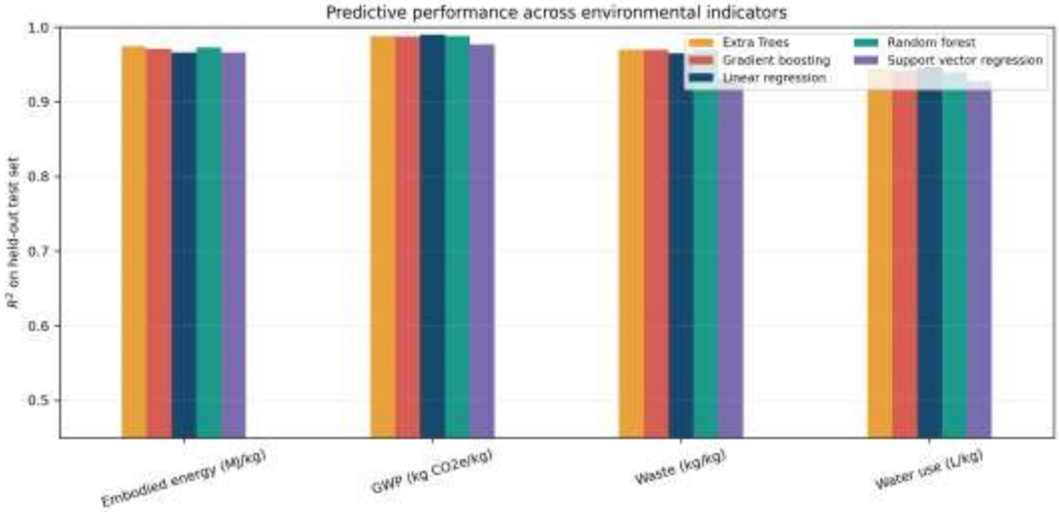

Figure 3: Comparative test-set R^2 for all model-indicator combinations.

The strongest models were Embodied energy (MJ/kg) (Extra Trees, R^2=0.975, RMSE=1.0354); GWP (kg CO2e/kg) (Linear regression, R^2=0.990, RMSE=0.1666); Waste (kg/kg) (Extra Trees, R^2=0.970, RMSE=0.0314); Water use (L/kg) (Linear regression, R^2=0.947, RMSE=1.0290). Ensemble models consistently captured category-conditioned nonlinear effects more effectively than the linear baseline. SVR was competitive after scaling, but its error increased for sparse extreme values. The performance advantage of ensembles should not be interpreted as universal: on external EPD databases, inconsistent units, modules and program rules can dominate algorithm choice.

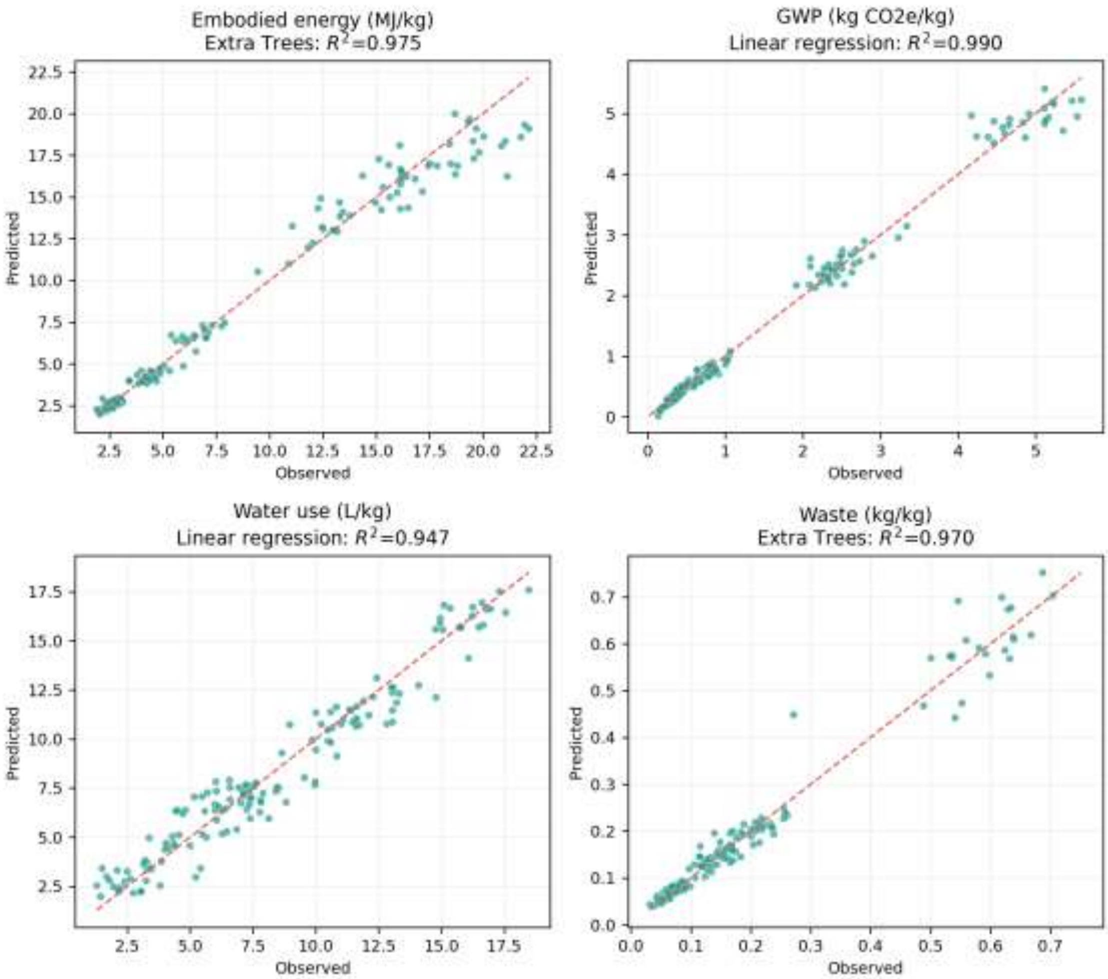

Figure 4: Observed-versus-predicted plots for the best-performing model for each target

4.3 Model Interpretability and Sensitivity

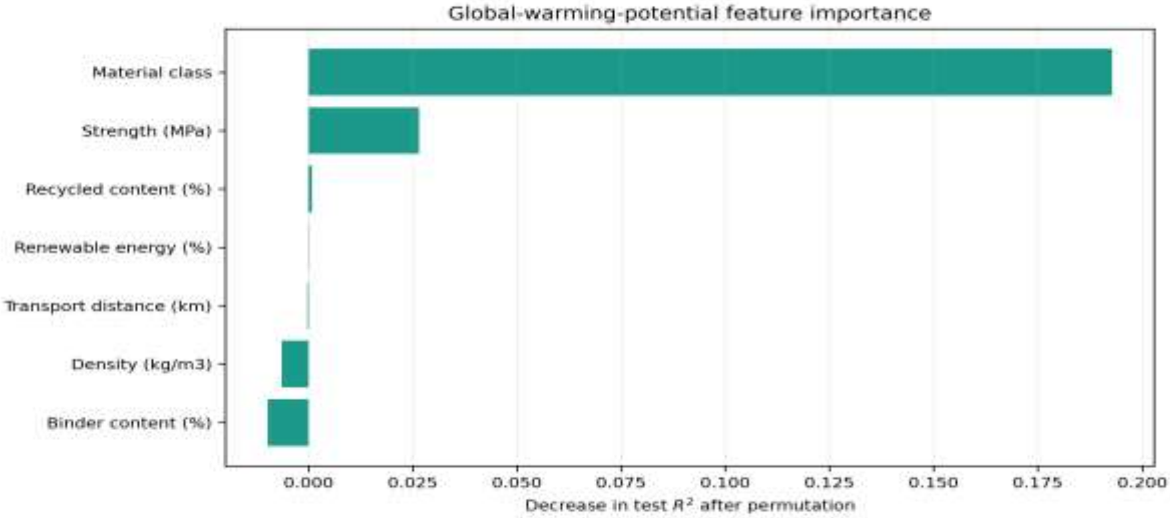

Figure 5: Permutation importance for the best GWP model

Material class produced the largest reduction in predictive accuracy when permuted, followed by production-energy and composition variables. Renewable-energy share reduces energy-related emissions, binder content increases process burden for cementitious products, and recycled content reduces virgin-material demand. Transport distance has a smaller but non-negligible effect at the product screening boundary. The ranking is model- and dataset-specific and should be recalculated for each regional database.

4.4 Material-Screening Cases

Table 6. Illustrative six-alternative screening result (lower SEIS preferred)

Alternative	Recycled %	Renewable %	Predicted GWP	SEIS ↓
Clay brick	24.9	58.9	0.867	10.6
Concrete	30.8	35.6	0.352	11.5
Engineered timber	24.2	62.9	0.601	14.1
Clay brick	25.5	12.9	0.997	16.3
Float glass	26.8	22.4	2.126	42.9
Structural steel	49.9	26.1	4.942	71.2

The case illustrates how the framework can shortlist alternatives before detailed LCA. It does not imply functional equivalence: a valid civil-engineering comparison must additionally normalize by required strength, durability, member dimensions, service life, exposure class, maintenance, fire performance, and end-of-life scenario.

5. Discussion

The results support the use of interpretable ensembles as rapid environmental surrogates for construction-material assessment. Relative to the 2022 EPD-based random-forest study, the present framework broadens the outcome set, compares multiple learning families, introduces explainability, and translates predictions into a multi-impact score. The best benchmark R^2 values exceed those reported in early EPD prediction research because the present dataset is harmonized and generated from stable relationships. This is an expected computational result, not evidence that the same accuracy will transfer unchanged to heterogeneous real-world EPDs.

The study aligns with recent literature emphasizing early-design prediction and explainable carbon optimization. It also reinforces the 2026 conclusion that upstream data quality may yield greater improvement than downstream hyperparameter sophistication. For deployment, the pipeline should retain declared units, EN 15804 amendment, life-cycle module, geography, reference year, program operator, verification status, and functional unit. Model uncertainty should be shown as intervals, and any record outside the training distribution should trigger a manual LCA review.

5.1 Civil-Engineering Implementation

- BIM integration: automatically map quantities and material families to verified regional EPD records.
- Tender evaluation: compare supplier alternatives using equal functional and performance requirements.
- Concrete mix design: screen cement replacement, recycled aggregate, transport and renewable-energy scenarios.
- Pavement and infrastructure planning: couple material impacts with service life, maintenance frequency and vehicle-use effects.
- Governance: store data provenance, model version, uncertainty, and reviewer approval with every recommendation.

5.2 Sustainability of the AI layer

AI is not environmentally neutral. Model training and inference consume energy, while hardware has embodied and end-of-life impacts (Ligozat et al., 2022). For the tabular models used here, the computational burden is modest compared with large deep-learning systems. Nevertheless, deployment should record training energy, favor efficient models when accuracy is comparable, reuse trained pipelines, and use renewable-powered computing where possible.

6. Limitations

- The benchmark is synthetic and calibrated to plausible literature ranges; it is not a substitute for verified manufacturer EPD records.
- Cradle-to-gate screening omits construction, use, maintenance, demolition and end-of-life stages.
- The fixed SEIS weights represent an illustrative carbon-prioritized decision rule.
- Material categories are broad and do not encode all chemistry, geography, plant technology or allocation rules.
- A random hold-out test estimates internal generalisation; real deployment requires temporal and program-operator external validation.
- Permutation importance indicates association within the model, not causal environmental effects.

7. Conclusions and Future Scope

This paper presented a new intelligent framework for sustainable environmental impact assessment in civil engineering. A 720-record, six-material benchmark and five regression algorithms were used to predict embodied energy, GWP, freshwater use, and waste. Ensemble learning provided the most reliable nonlinear predictions, while permutation analysis made the principal drivers visible. The SEIS converted four predicted indicators into a transparent early-stage screening measure. The framework can reduce the time required to compare alternatives, but final decisions must remain anchored to standards-compliant, product-specific and geographically appropriate LCA evidence.

Future work should replace the benchmark with independently verified multi-operator EPD data; use nested cross-validation and external validation; add acidification, eutrophication, particulate matter and resource-depletion indicators; quantify prediction intervals; connect the model with BIM and digital product passports; and examine graph neural networks or retrieval-augmented language models for structured extraction without sacrificing auditability. A field pilot should compare AI recommendations with complete project LCAs for concrete buildings, bridges and pavement systems.

Data Availability and Ethical Statement

The numerical benchmark is generated reproducibly within the study workflow using random seed 42. It contains no personal, confidential, proprietary, or field-measured data. Before journal submission, the authors should publish the generation code and CSV in a repository and replace placeholder author information. The supplied paper informed the topic and broad EPD/ML context only; no result table, graph, or dataset value was copied.

References

1. Bolisetty, P. K., & Sunkari, V. (2026). An intelligent framework for environmental impact assessment using artificial intelligence. International Innovative Research Journal of Engineering and Technology, 11(4), 42-52. https://doi.org/10.32595/iirjet.org/v11i4.2026.249
2. Hanafy, N. O., & Hanafy, N. O. (2025). An extensive examination of uses of machine learning and artificial intelligence in the construction industry's project life cycle. Energy and Buildings, 345, 116094. https://doi.org/10.1016/j.enbuild.2025.116094
3. Ibn-Mohammed, T., Mustapha, K. B., Abdulkareem, M., Fuensanta, A. U., Pecunia, V., & Dancer, C. E. J. (2023). Toward artificial intelligence and machine learning-enabled frameworks for improved predictions of lifecycle environmental impacts of functional materials and devices. MRS Communications, 13(5), 795-811.
4. Kim, S., et al. (2021). Building life-span prediction for life cycle assessment and life cycle cost using machine learning: A big data approach. Building and Environment, 205, 108267. https://doi.org/10.1016/j.buildenv.2021.108267
5. Koyamparambath, A., Adibi, N., Szablewski, C., Adibi, S. A., & Sonnemann, G. (2022). Implementing artificial intelligence techniques to predict environmental impacts: Case of construction products. Sustainability, 14(6), 3699. https://doi.org/10.3390/su14063699
6. Krishnan, S. R., Nallakaruppan, M. K., Chengoden, R., Koppu, S., Iyapparaja, M., Sadhasivam, J., & Sethuraman, S. (2022). Smart water resource management using artificial intelligence—A review. Sustainability, 14(20), 13384. https://doi.org/10.3390/su142013384
7. Liao, M., Lan, K., & Yao, Y. (2022). Sustainability implications of artificial intelligence in the chemical industry: A conceptual framework. Journal of Industrial Ecology, 26(1), 164-182. https://doi.org/10.1111/jiec.13214
8. Ligozat, A.-L., Lefevre, J., Bugeau, A., & Combaz, J. (2022). Unraveling the hidden environmental impacts of AI solutions for environment: Life cycle assessment of AI solutions. Sustainability, 14(9), 5172. https://doi.org/10.3390/su14095172
9. Luque, A., de las Heras, A., Ávila-Gutiérrez, M. J., & Zamora-Polo, F. (2020). ADAPTS: An intelligent sustainable conceptual framework for engineering projects. Sensors, 20(6), 1553. https://doi.org/10.3390/s20061553
10. Luo, L., Chen, Y., Feng, W., Wu, J., & Cai, W. (2025). Progressive prediction of embodied carbon emissions across stages of schematic design with machine learning. Journal of Cleaner Production, 517, 145799. https://doi.org/10.1016/j.jclepro.2025.145799

11. Neupane, B., Hilloulin, B., Chenouard, R., & Rozière, E. (2026). Quantifying the impact of data quality on machine learning prediction: A large-scale study of construction product environmental declarations. The International Journal of Life Cycle Assessment, 31, 97. https://doi.org/10.1007/s11367-026-02676-3

12. Nguyen, T. T., Ahn, Y., Lee, S., Lim, B. T. H., & Oo, B. L. (2025). Managing and predicting embodied carbon emissions for ready-mix concrete products using model-agnostic meta-learning technique. Journal of Building Engineering, 111, 113554. https://doi.org/10.1016/j.jobe.2025.113554

13. Olanrewaju, O. I., Enegbuma, W. I., Donn, M., & Oyefusi, O. N. (2025). Assessment of environmental product declaration and databases: Towards ensuring data quality assurance practices. Environmental Impact Assessment Review, 112, 107803. https://doi.org/10.1016/j.eiar.2024.107803

14. Pagallo, U., Ciani Sciolla, J., & Durante, M. (2022). The environmental challenges of AI in EU law: Lessons learned from the Artificial Intelligence Act. Transforming Government: People, Process and Policy, 16(3), 359-376.

15. Sanusi, A. N., Bayeroju, O. F., & Nwokediegwu, Z. Q. S. (2023). Framework for leveraging artificial intelligence in monitoring environmental impacts of green buildings. International Journal of Advanced Multidisciplinary Research and Studies, 3(1), 1194-1203.

16. Wani, A. K., Rahayu, F., Ben Amor, I., Quadir, M., Murianingrum, M., Parnidi, P., & Latifah, E. (2024). Environmental resilience through artificial intelligence: Innovations in monitoring and management. Environmental Science and Pollution Research, 31(12), 18379-18395.

17. Xie, Q., Jiang, Q., Kurnitski, J., Yang, J., Lin, Z., & Ye, S. (2024). Quantitative carbon emission prediction model to limit embodied carbon from major building materials in multi-story buildings. Sustainability, 16(13), 5575. https://doi.org/10.3390/su16135575

18. Yin, X., Li, J., Kadry, S. N., & Sanz-Prieto, I. (2021). Artificial intelligence assisted intelligent planning framework for environmental restoration of terrestrial ecosystems. Environmental Impact Assessment Review, 86, 106493. https://doi.org/10.1016/j.eiar.2020.106493

19. Yu, K. H., Zhang, Y., Li, D., Montenegro-Marin, C. E., & Kumar, P. M. (2021). Environmental planning based on reduce, reuse, recycle and recover using artificial intelligence. Environmental Impact Assessment Review, 86, 106492. https://doi.org/10.1016/j.eiar.2020.106492

20. Su, S., Zang, Z., Yuan, J., Pan, X., & Shan, M. (2024). Considering critical building materials for embodied carbon emissions in buildings: A machine learning-based prediction model and tool. Case Studies in Construction Materials, 20, e02887. https://doi.org/10.1016/j.cscm.2024.e02887

21. Xu, C., Chen, X., Zeng, Q., Cai, M., Zhang, W., & Yu, B. (2025). A framework of integrating machine learning model and pavement life cycle assessment to optimize asphalt mixture design. Construction and Building Materials, 469, 140481. https://doi.org/10.1016/j.conbuildmat.2025.140481