https://doi.org/10.1051/e3sconf/202671101009

Sustainable Material Properties Driven by the Data Using Mathematical and Machine Learning Optimization Techniques

Moti Ranjan Tandi[1*], Subrata Majee[2], and Karan Khati[3]

[1]Department of CS & IT, Kalinga University, Raipur, India.
[2]Department of Civil Engineering, Kalinga University, Raipur, India.
[3]New Delhi Institute of Management, New Delhi, India.

Abstract. Sustainable material development is becoming more data-driven with mathematical modeling and machine learning incorporating properties such as durability, strength and environmental impact into the equation. This research is devoted to a hybrid approach to predicting the design of environmentally responsible materials using optimization techniques. Conventional methods often include trial and error experiments or linear modeling, which take time, are not as accurate and do not account for complex and nonlinear relationships between material parameters. They usually don't allow optimization of more than one conflicting objective (e.g. strength vs. carbon footprint) at the same time. To solve these problems, this paper proposes a Multi-Objective Optimization framework (Genetic Algorithms (GA) - Machine Learning (ML) Regression Models). The ML model is used to learn from past datasets of materials to predict performance indicators and the GA is used to explore the design space to simultaneously find the compositions of materials that meet the design criteria of sustainability and performance. This methodology is applied for optimize the eco-concrete mix design in construction industry, with regard to reduce embodied CO_2 emissions with the same mechanical strength requirements. The proposed system significantly improves environmental and structural metrics, offering a data-driven pathway to sustainable material innovation.

1 Introduction

A. Motivation for Sustainable Material Design

With rapid industrialization and urbanization, there has been a surge in demand for construction and manufacturing materials. The extraction of raw materials for these processes, as well as the traditional processes themselves, is also highly reliant on raw materials consumed, resulting in wasted energy and resources, including those that are not renewable [1]. This concern has shifted the focus of the university and industry towards sustainable materials, which minimize environmental impact and carbon footprint, are efficient in resource use, and promote a circular economy. Sustainable materials are regarded as producing or having minimal waste and carbon footprint while meeting or exceeding design performance and performance metrics, such as strength, durability, and thermal stability [2]. This trajectory towards sustainable design, construction, and manufacturing of materials will depend on a proper methodological approach that can predict and optimize multiple material properties based on environmental criteria [10] [3].

B. Limitations of Conventional Material Optimization

Typical optimization methods are primarily based on trial-and-error experimentation, empirical correlations, and empirical linear modeling. Traditional optimization methods have time and resource limitations, primarily based on nonlinear, multidimensional relationships between composition and properties [14]. For instance, altering one constituent in concrete can simultaneously affect all of its essential properties, like workability, strength, and CO_2 emissions. Generally, conventional optimization techniques are focused on either one or two objectives. The primary feature of all existing optimization methods (e.g., linear or nonlinear) is that optimization is discrete. Thus, the issue of one-objective optimization or the two-goal optimization problem (i.e., cost vs performance) [5] occurs when the basis and/or the Environmental Trade-offs are left out. In recent years, sustainability has cast a cloud of uncertainty over what performance means, as it must be constantly assessed in terms of a continually changing regulatory landscape and lifecycle performance expectations [7].

C. Role of Data-Driven Methods in Material Innovation

Combining data-driven technologies, particularly machine learning (ML) and mathematical optimization techniques, provides an innovative approach to developing sustainable materials [15]. These data-driven technologies enable to explore design spaces that would otherwise be unlikely to investigate and can reveal patterns in high-dimensional data that would be too cumbersome to characterize analytically [4]. Supervised

* Corresponding author: ku.motiranjantandi@kalingauniversity.ac.in

ML algorithms, such as regression trees or ensemble models, can be trained to predict target material properties (e.g., compressive strength, thermal conductivity, CO_2 emissions) from material compositions and processing conditions based on previously executed experimental data [6]. This GA-ML hybrid will enable to create a closed-loop system; the ML subroutine will direct the optimization algorithm, while the optimizer varies the material compositions to approach performance and sustainability targets [9][8].

2 Literature Review

The rising need of materials that are sustainable in construction industry has compelled a lot of studies on how to maximize the material properties in terms of strength, durability and environmental effects of the material. The necessity to minimize the environmental impact of the construction industry (in the most pressing way) has triggered this change, especially the inventions related to concrete and other construction materials. Conventionally, material optimization has been pegged on trial-and-error experimentation and empirical correlations, which are not able to reflect the nonlinear and complex relationship between material compositions and performance [12]. Such a range of conflicting objectives and the requirement to address the problem using multi-objective optimization methods, including performance and environmental impact, have prompted the use of data-driven strategies, including machine learning (ML) and mathematical optimization tools.

2.1 Sustainable Material Development and Optimization

The material design should be sustainable and consider a wide range of variables such as efficiency of resources, durability, and lower environmental impact. Conventional techniques, despite being successful in a number of contexts, are not capable of dealing with the multidimensional nature of contemporary material design issues [13]. Indicatively, when the composition of ingredients in concrete is changed not only the strength, but also workability, cost and carbon emission are changed. Machine learning models and algorithms in particular, supervised learning algorithms such as the Random Forest Regression (RFR) showed that they could predict the property of materials with a high level of accuracy using experimental data instead of relying on predetermined performance characteristics of the materials involved [1] [2]. The capability saves a lot of time and resources consumed in developing the material as opposed to the conventional approach.

2.2 Machine Learning in Sustainable material optimization

The recent research shows the promise of ML methods in predicting and optimization of sustainable material properties. The application of supervised models especially regression trees and ensemble models have been made in predicting such critical properties as compressive strength and thermal conductivity depending on the material composition [3] [4]. Highly precise predictions of experimental data may be achieved with these techniques on large volumes of experimental data, enabling the quick identification of promising material compositions. A major advantage of optimization with MLs is a less computational time and experiment cost required to investigate alternative material formulations 5 [11].

2.3 Multi-purpose optimization and Sustainability Trade-Offs.

Multi-objective optimization techniques with special emphasis on Genetic Algorithms (GA) have been found useful in solving trade-offs among multi-objective conflicting aims, namely, strength, durability, and carbon emissions. Materials optimization that combines the environmental objectives with performance indicators can be optimized with the help of the integration of ML models and GAs. The hybrid method can more effectively search through large and complicated spaces of designs, and it has been effectively used to optimize concrete mix designs in eco-concrete [6] [7]. Moreover, life cycle analysis (LCA) can be used to include long-term sustainability in optimization routines, and this offers the decision-maker total data on the material impact in the long-term.

2.4. Gaps in Existing Research and Future Research

Although the recent studies have demonstrated promising outcomes in the application of the data-driven approaches towards the sustainable material design, a number of challenges exist. Different and larger datasets are required to train more robust models, especially when a new material, e.g., polymers or composites is being used. The latter might also be investigated in future research by applying the methods of deep learning, which may be able to reproduce more intricate relations in material properties and enhance prediction accuracy. The ability to extend the range of optimization algorithms to the larger environmental and economic contexts will also play a pivotal role in the successful implementation of the optimization methods in the real-world use scenarios in the future [10].

3 Proposed Framework

3.1 Overview of the Hybrid GA-ML Approach

The proposed framework comprises an ML regression model and a Multi-Objective GA designed to discover and optimize sustainable material compositions. The hybrid approach takes advantage of both the predictive power of the ML model and the sensitivity of the search of the GA to handle multiple conflicting objectives at the same time, such as compressive strength, durability, cost, and embodied carbon footprint.

Multi-Objective Genetic Algorithm (NSGA-II) is used for optimisation of compositions of eco-concrete mixes in the proposed framework. NSGA-II is programmed to successfully compute Pareto-optimal solutions based on mechanisms of non-dominated sorting and crowding distance, and is thus suitable to problems that are conflicting such as compressive strength and CO_2 emissions. The algorithm that is used generates an initial set of candidate solutions, generating code for it using some performance indicators, and implementing selection, crossover, and mutation operations. Its main strength is to remain diverse, offer a range of optimal solutions to a range of trade-offs. Use of NSGA-II instead of other methods such as SPEA2 is because it has the best balance of convergence and heterogeneity during the optimization of material.

The operation of this system is done in two phases. At first, the ML model is trained using the existing experimental or simulated data sets to learn the non-linear relations between the input features of materials (ex. proportions of raw materials) and the output properties. Once the model is highly accurate and precise in predicting, then this model is integrated in the GA loop where it serves as a surrogate evaluator, which helps to quickly predict material performance. The GA will then move on to evolve material design options through generations, which will improve designs using natural genetic processes, such as selection, crossover and mutation, converging towards optimal solutions for a sustainable performance target.

3.2 Machine Learning Regression for Property Prediction

Machine Learning is a critical component within the framework as it provides a data-driven surrogate model. A model structure that can be employed in this paper, depending on the complexity of the data and its availability, is the Random Forest Regression (RFR). These models are trained on training data so as to map the relationship between the features of material composition and performance indicators. The cross-validation will be used to validate the model, and the analysis of the model will be conducted based on R 2 score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) measures to establish whether the model will be generalized successfully.

3.3 Multi-Objective Genetic Algorithm for Optimal Composition

The GA component represents the solution space that is optimal in terms of material compositions, both in terms of performance and sustainability. The GA begins with a randomly created population of candidates (each being a unique material formulation). At each iteration of the GA, each candidate is evaluated using the ML model to estimate the materials' properties. The GA selects the best individuals based on fitness criteria, including multiple objectives (for instance, high strength, low cost, and low CO_2 emissions).

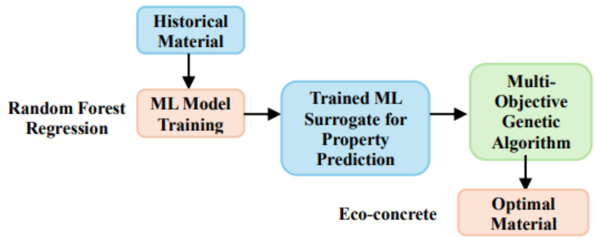

Figure 1. Block diagram of the hybrid GA – ML Framework

Figure 1 shows the hybrid GA-ML framework. It starts by taking historical or simulated information about material datasets in order to generate a regression-based ML model. The trained ML model then serves as a fast and computationally inexpensive predictor of material properties. The surrogate model is embedded in a multi-objective GA that explores different potential materials in composition. The GA model refines candidate solutions iteratively based on the predicted performance of each candidate. The GA offers a database of the best sustainable material designs. The key benefits of this hybrid are that this method will greatly minimize experimentation and keep the material discovery fast.

4 Case Study: Eco-Concrete Mix Optimization

4.1 Dataset Description and Feature Selection

The case study aims at optimizing the eco-concrete mixes to be used in the construction process of sustainable buildings. The dataset has 1,030 concrete mix records from experimental study technical reports and publicly available databases. It has input variables (cement, fly ash, slag, total water, coarse aggregate, fine aggregate, admixtures, and curing time) and output variables (compressive strength (MPa), & CO_2 emission per m^3 (kg CO_2-eq). Correlation analysis and mutual information ranking were used to select the most important input parameters with respect to compressive strength and carbon footprint.

4.2 Training the Regression Model

A RFR model was modeled to the dataset in order to predict the compressive strength and CO_2 emissions according to the chosen features of the mix. The data was divided into 80 % training and 20 % testing data. The RFR model was adjusted with the help of cross-validation and a grid search of the hyperparameter values such as the number of estimators and trees depth. The values of R^2 of strength prediction were $R^2 = 0.93$ and R 2 = 0.89 of CO_2 prediction indicated that the predictions were accurate.

5 Results and Discussion

5.1 Performance Metrics of the Regression Model

The RFR model showed high predictive performance for compressive strength and carbon emissions. The RFR model was developed using 80% of the dataset, while the remaining 20% was used to validate its performance. Table 1 contains the evaluation metrics. The score of 0.93 for strength prediction is an R^2 score that shows a very accurate model, and even R^2 of 0.89 for carbon emission prediction seems to show a good model generalization. The relatively small values for MAE and RMSE indicate that the model accurately captures the inherent, complex relationships found in the dataset.

Table 1. Regression Model Performance

Property	**R^2 Score**	**MAE (Mean Abs. Error)**	**RMSE (Root Mean Squ. Error)**
Compressive Strength (MPa)	0.93	2.1 MPa	3.4 MPa
CO_2 Emissions (kg/m³)	0.89	5.8 kg/m³	9.2 kg/m³

5.2 Optimization Trade-offs: Strength vs Carbon Footprint

The GA-RFR model led to the Pareto front depiction of a trade-off between maximum compressive strength and minimum of CO_2 emissions. The solutions indicated that modest reductions in strength (5–10%) could achieve significant reductions in emissions (40–50%). As such, decision-makers can pick mixed designs that align with their sustainability goals, depending on the project.

5.3 Comparison with Baseline Methods

Table 2. Comparison with Baseline Methods

Method	**Strength Accuracy**	**CO_2 Reduction vs. Baseline**	**Multi-Objective Capability**
Conventional Design	High	Low (~5%)	No
Strength-only Optimization	Very High	Moderate (~20%)	No
Proposed GA-ML Framework	High	High (~45%)	Yes

A hybrid method was proposed, which was evaluated against mixed design methods and a single-objective optimization (strength only). The results in Table 2 above demonstrate that the hybrid approach yielded a significantly more balanced performance than the other strategies. Normally design methods tend to prioritise strength over environmental consequences. A hybrid approach of GA-ML can be a good balance between structural performance and sustainability.

6 Conclusion and Future Work

Proposed hybrid optimization architecture, where the Multi-Objective Genetic Algorithm (NSGA-II) was combined with the use of Machine Learning regression models, is efficient in terms of optimizing sustainable material compositions, in particular, with eco-concrete. This framework has shown that it could predict some important material properties (compressive strength and CO_2 emission) and at the same time search through the design space to find the best solutions which between sustainability and performance. The findings indicate that the GA-ML framework can provide a high level of gain on the material design in the eco-efficient way, which allows reducing the carbon footprint without the need to decrease the necessary structural performance. The framework has the potential to save the experimental costs and time considerably by using data-driven predictions and evolutionary search processes, which is a more efficient way to achieve sustainable material innovation.

Nevertheless, the framework has some limitations, which should be tackled. It depends on the quality and availability of the experimental data which influences the accuracy of the prediction of the ML model and any gaps in the data or biases may influence the reliability of results. Also, the framework is eco-concrete-oriented, though the further research and adaption of the framework to other materials like polymers, composite or other construction materials are necessary. Future studies are required to focus on enhancing the external validity of the framework through the addition of more materials and environmental conditions into the dataset. Besides, further study of sophisticated optimization methods such as deep reinforcement learning would enhance the trade-off between exploration and exploitation in large and complicated design spaces. The additional implementation of the sustainability metrics, including life cycle analysis (LCA), will also give a more detailed analysis of the material performance, including the impact on the entire environment through the lifecycle of the material.

References

1. Dowden, J. M. (2024). *The Mathematics of Thermal Modeling: An Introduction to the Theory of Laser Material Processing, 2e*. CRC Press.
2. Mokhtarinejad, A., Mokhtarinejad, O., Kafaki, H. B., & Ebrahimi, S. M. H. S. (2017). Investigating German Language Education through Game (Computer and non-Computer) and its Correspondence with Educational Conditions in Iran. *International Academic Journal of Innovative Research*, *4*(2), 1-9.
3. Yonis, E. G. (2024). The Effect of Thickness on the Thermal Conductivity Coefficient and Some Mechanical Properties of Acrylic Material Used in the Manufacture of Dentures. *International Academic Journal of Science and Engineering*, *11*(2), 8-14. https://doi.org/10.9756/IAJSE/V11I2/IAJSE1143

4. Gundersen, Ø., Kluken, A. O., Myhr, O. R., Jones, J. E., Rhoades, V., Day, J., ... & Krygowski, B. (2024). The use of an integrated multiple neural network structure for simultaneous prediction of weld shape, mechanical properties, and distortion in 6063-T6 and 6082-T6 aluminium assemblies. In *Mathematical Modelling of Weld Phenomena: No. 5* (pp. 671-700). CRC Press.
5. Al-Jizani, H. N. Z., & Kayabaş, A. (2023). Students Real Data Features Analyzing with Supervised Learning Algorithms to Predict Efficiency. *International Journal of Advances in Engineering and Emerging Technology*, *14*(1), 34-45.
6. Wang, H. S. H., & Yao, Y. (2023). Machine learning for sustainable development and applications of biomass and biomass-derived carbonaceous materials in water and agricultural systems: A review. *Resources, Conservation and Recycling*, *190*, 106847.
7. Ulkilan, A., Mara, G. R., & Aleem, F. M. (2022). Harnessing high-temperature superconducting resonance coils and future perspectives. *National Journal of Antennas and Propagation*, *4*(1), 8-13.
8. Orozco, L., & Ttofis, H. (2025). Energy harvesting techniques for sustainable embedded systems: Design and applications. *SCCTS Journal of Embedded Systems Design and Applications*, *2*(1), 67-78.
9. Yao, Z., Lum, Y., Johnston, A., Mejia-Mendoza, L. M., Zhou, X., Wen, Y., ... & Seh, Z. W. (2023). Machine learning for a sustainable energy future. *Nature Reviews Materials*, *8*(3), 202-215.
10. Chia-Hui, C., Ching-Yu, S., Fen, S., & Ju, Y. (2025). Designing scalable IoT architectures for smart cities: Challenges and solutions. *Journal of Wireless Sensor Networks and IoT*, *2*(1), 42-49.
11. Zhang, X., & Jia, H. (2023). Low impact development planning through a comprehensive optimization framework: Current gaps and future perspectives. *Resources, Conservation and Recycling*, *190*, 106861.
12. Liu, Y., Yang, Z., Yu, Z., Liu, Z., Liu, D., Lin, H., ... & Shi, S. (2023). Generative artificial intelligence and its applications in materials science: Current situation and future perspectives. *Journal of Materiomics*, *9*(4), 798-816.
13. Ahmad, T., Madonski, R., Zhang, D., Huang, C., & Mujeeb, A. (2022). Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renewable and Sustainable Energy Reviews*, *160*, 112128.
14. de Souza Pereira Moreira, G., Rabhi, S., Lee, J. M., Ak, R., & Oldridge, E. (2021, September). Transformers4rec: Bridging the gap between nlp and sequential/session-based recommendation. In *Proceedings of the 15th ACM conference on recommender systems* (pp. 143-153).
15. Sajid, S., Haleem, A., Bahl, S., Javaid, M., Goyal, T., & Mittal, M. (2021). Data science applications for predictive maintenance and materials science in context to Industry 4.0. *Materials today: proceedings*, *45*, 4898-4905.