

Optimization of Environmentally Friendly Asphalt Mix Formulation Using Artificial Intelligence Based on Genetic Algorithm

I Gusti Agung Ananda Putra^{1*}

¹Faculty of Engineering and Informatics, Universitas Pendidikan Nasional, Indonesia

Abstract. This study develops an intelligent system to optimize asphalt mix formulations with the goal of reducing carbon dioxide equivalent (CO₂e) emissions using a data-driven approach. A Random Forest regression model predicts emissions based on mix design parameters, while a Genetic Algorithm (GA) identifies the optimal combination. The dataset includes 50 asphalt mix records with variables such as binder content, mixing temperature, RAP proportion, aggregate type, and production method. The Random Forest model achieved a coefficient of determination (R²) of 0.772 and a Root Mean Square Error (RMSE) of 5.27 kg CO₂e/ton, indicating strong predictive capability, identifying mixing temperature, binder content, and RAP proportion as the most influential factors. GA optimization resulted in a mix with 4.65% binder, 125 degrees Celsius mixing temperature, and 48% RAP, producing 67.5 kg CO₂e/ton—around 20% lower than conventional Hot Mix Asphalt. The findings demonstrate that combining artificial intelligence with optimization techniques provides an effective approach for sustainable asphalt design and supports low-carbon infrastructure planning.

1 Introduction

Climate change has emerged as one of the most critical global issues due to its widespread environmental, economic, and social impacts [1]. A major contributor to climate change is carbon dioxide (CO₂), primarily emitted from industrial processes, energy production, and infrastructure development. The construction industry alone is responsible for nearly 39% of global energy-related CO₂ emissions, with 11% directly linked to the manufacturing of construction materials such as steel, cement, and asphalt [2].

Among various construction materials, asphalt is widely used in pavement structures, especially in road infrastructure. However, the production of asphalt mixtures particularly Hot Mix Asphalt (HMA), which requires mixing temperatures ranging from 140 to 180°C, results in significant energy consumption and CO₂ emissions [3, 4]. According to lifecycle studies, asphalt production can generate between 30 and 60 kg CO₂e per ton of mixture, depending on materials and processing methods [5, 6]. These emissions stem largely from

* Corresponding author: anandaputra@undiknas.ac.id

the combustion of fossil fuels during heating and the use of bitumen, a petroleum-based binder with high carbon intensity [7].

Efforts to reduce the environmental footprint of asphalt mixtures have led to the development of technologies such as Warm Mix Asphalt (WMA) and the incorporation of Reclaimed Asphalt Pavement (RAP). WMA enables lower production temperatures, thus decreasing energy use and emissions by up to 40% compared to HMA [8]. Similarly, RAP helps reduce dependence on virgin aggregates and bitumen by allowing recycling of existing materials, resulting in both economic and environmental benefits [9, 10].

Nevertheless, the implementation of environmentally friendly asphalt technologies in practice is still limited. One of the main challenges lies in the absence of intelligent tools for predicting environmental impacts and optimizing mix formulations under sustainability constraints [11, 12]. Conventional mix design procedures such as Marshall or Superpave do not explicitly incorporate carbon emissions or energy efficiency in their evaluation criteria [13, 14]. As a result, these traditional approaches may lead to suboptimal choices from an environmental perspective.

Recent advancements in Artificial Intelligence (AI) and data science have introduced new possibilities for enhancing asphalt mix design. Machine learning models have been successfully applied to predict mechanical properties and environmental indicators such as carbon emissions based on input variables like binder content, aggregate type, and temperature [15]. Among AI optimization tools, the Genetic Algorithm (GA) stands out as a metaheuristic technique capable of solving multi-objective problems involving complex search spaces.

Several studies have integrated machine learning models with GA to achieve optimal asphalt formulations that minimize emissions while satisfying performance standards. These integrated frameworks enable the evaluation of a large number of design alternatives without requiring costly and time-consuming laboratory testing. However, few studies have implemented such systems in accessible platforms that facilitate adoption by industry practitioners.

To address this gap, the present study proposes an intelligent system for optimizing environmentally friendly asphalt mix formulations using a combination of Random Forest regression and Genetic Algorithm-based search, implemented through the Orange Data Mining platform. This system aims to predict CO₂ emissions from asphalt mixtures and identify optimal combinations of mix parameters—such as binder content, mixing temperature, and RAP content—that result in the lowest possible emissions. The model supports sustainable infrastructure design and contributes to national and global efforts to reduce greenhouse gas emissions, including Indonesia's commitment to reducing emissions by 31.89% by 2030.

2 Method

This research employs a quantitative, simulation-based experimental methodology aimed at developing and optimizing an asphalt mix formulation that is environmentally sustainable. The central focus of the study is to minimize carbon emissions (CO₂e) produced throughout the asphalt production process by using a data-driven artificial intelligence (AI) approach. The methodology integrates two main components: a machine learning-based predictive model and an optimization process using a Genetic Algorithm (GA). All stages of the modeling and optimization pipeline are developed and executed using Orange Data Mining, a visual programming platform that facilitates interactive data analysis and machine learning, particularly useful for users who may not be proficient in programming but require reliable analytical outputs.

The research begins with the data collection stage, wherein a dataset comprising 50 records of asphalt mix configurations was compiled. The variables considered include binder content (% by weight), mixing temperature (°C), percentage of Reclaimed Asphalt Pavement (RAP), aggregate moisture content (%), aggregate type (recycled or virgin), mix type (Hot Mix Asphalt—HMA, Warm Mix Asphalt—WMA, or Cold Mix Asphalt—CMA), and production method (batch or continuous). These parameters are widely recognized in literature as having direct or indirect effects on energy consumption and greenhouse gas emissions in pavement construction. The target output variable is the estimated carbon emission per ton of asphalt mix, expressed in kg CO₂e/ton, obtained from existing Life Cycle Assessment (LCA) data, simulation models, or prior studies.

After assembling and formatting the dataset, the next step involves data preprocessing. The dataset is imported into Orange, where categorical variables are encoded using one-hot encoding, and numerical values are normalized to ensure equal weighting and stable convergence during model training. The data is then split into training (80%) and testing (20%) subsets using random sampling to maintain data variability and prevent overfitting. The model selected for this research is a Random Forest regression model, which is known for its robustness, high accuracy, and ability to handle both linear and nonlinear relationships among variables. It also performs well in high-dimensional spaces and is less susceptible to noise, making it suitable for predictive modeling in environmental systems.

Once the model is trained, its predictive performance is evaluated using three statistical metrics: the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). A high R^2 value close to 1 indicates a good fit between the predicted and actual values, while low MAE and RMSE values indicate high precision and minimal deviation. These metrics collectively provide a comprehensive assessment of model reliability. In this study, the Random Forest model yielded an R^2 of 0.772 and an RMSE of 5.27 kg CO₂e/ton, signifying strong predictive performance.

Following model validation, the study proceeds to the feature importance analysis using Orange's permutation-based method. This analysis identifies which variables have the most influence on carbon emissions. The results highlight three dominant variables: mixing temperature, binder content, and RAP proportion. These variables are therefore prioritized for optimization, as changes in their values have the greatest potential to reduce environmental impact.

The subsequent phase involves the implementation of the Genetic Algorithm (GA) for optimization. GA is a bio-inspired metaheuristic algorithm modeled after the process of natural selection, widely applied in engineering problems to discover optimal or near-optimal solutions in large, complex search spaces. In this research, GA is used to minimize the predicted carbon emissions based on the machine learning model's output. The optimization focuses on three continuous variables—binder content, mixing temperature, and RAP percentage—within predefined boundaries: 4.0% to 6.5% for binder, 100°C to 170°C for mixing temperature, and 0% to 60% for RAP.

Categorical variables such as mix type, aggregate type, and production method are fixed at environmentally favorable values (WMA, recycled, and batch, respectively) to streamline the optimization process and align with sustainable construction principles. The GA configuration parameters include a population size of 50, 40 generations, a crossover probability of 0.8, and a mutation rate of 0.1. The fitness function used to guide the algorithm is based on the emission values predicted by the trained Random Forest model, allowing each candidate solution to be evaluated without the need for physical testing.

During the optimization process, the GA evolves the population through selection, crossover, and mutation operators. The best-performing individuals are retained across generations, with new solutions generated from their genetic makeup. Penalty functions are also incorporated to discourage solutions that exceed environmental thresholds or violate

design constraints, such as excessive binder content or inadequate RAP usage. After several generations, the algorithm converges on an optimized solution, which in this case achieved a configuration of 4.65% binder, 125°C mixing temperature, and 48% RAP, resulting in a predicted emission value of 67.5 kg CO₂e/ton—a reduction of approximately 20% compared to conventional HMA values.

The entire workflow, from data ingestion to model training, evaluation, and optimizations conducted within the orange platform, which visually represents each step as a node in a connected workflow. This enhances transparency, reproducibility, and ease of use, especially in multidisciplinary research teams. Moreover, Orange supports integration with Python scripts, allowing future expansion and customization beyond the visual interface if needed.

In conclusion, the methodology adopted in this study offers a robust, adaptable, and user-friendly framework for optimizing asphalt mix designs from an environmental perspective. By combining the predictive power of machine learning with the explorative capabilities of Genetic Algorithms and implementing them within the orange software environment, this research demonstrates a novel approach to achieving sustainable infrastructure goals through intelligent system design.

3 Results And Discussion

3.1 Description of Research Data

The data used in this study consisted of 50 asphalt mix entries with various technical parameters that affect carbon emissions. These parameters include aggregate composition, asphalt content, mixing temperature, use of recycled materials (Reclaimed Asphalt Pavement/RAP), and production technology type (Hot Mix Asphalt/HMA or Warm Mix Asphalt/WMA). Each entry also includes actual carbon emissions data calculated using the Life Cycle Assessment (LCA) approach.

This data is used as the basis for training and testing a machine learning-based predictive model using the orange platform. The Data shown in Fig. 1. The primary objective is to evaluate the model's ability to accurately predict carbon emissions and identify the most influential variables affecting emission outcomes.

Data Table - Orange								
	Jenis_Agregat	Kadar_Aspal_(%)	Suhu_Pencampuran_(°C)	Proporsi_RAP_(%)	Kelengkapan_Agregat_(%)	Tipe_Campuran	Emisi_CO2e_(kg_per_ton)	Metode_Produksi
1	Recycled	4.11	157	30	4.4	CR	81.9	Batch
2	Natural	4.99	147	30	2.3	WMA	87.8	Continuous
3	Recycled	5.88	128	50	4.0	WMA	79.8	Batch
4	Recycled	5.79	178	10	3.1	CR	82.6	Continuous
5	Natural	5.21	148	60	4.1	CR	70.6	Batch
6	Natural	5.57	130	20	1.8	HMA	95.5	Continuous
7	Natural	4.38	164	0	4.2	CR	89.8	Continuous
8	Recycled	5.92	128	40	3.0	HMA	84.5	Batch
9	Recycled	4.21	121	50	4.1	WMA	79.7	Continuous
10	Natural	5.70	137	40	2.9	HMA	79.0	Batch
11	Natural	5.43	150	0	2.2	WMA	91.2	Batch
12	Natural	4.70	121	60	1.7	HMA	77.3	Continuous
13	Slag	4.99	153	0	1.9	HMA	95.7	Batch
14	Slag	5.23	162	40	1.7	HMA	72.8	Continuous
15	Slag	4.03	142	60	4.3	WMA	76.7	Batch
16	Recycled	5.56	131	20	2.3	CR	94.2	Continuous
17	Natural	5.87	176	30	1.6	WMA	73.4	Continuous
18	Natural	5.67	151	50	2.9	CR	77.9	Batch
19	Recycled	5.66	156	60	3.0	HMA	67.0	Batch
20	Recycled	4.54	146	0	4.1	HMA	95.9	Batch
21	Slag	4.22	158	30	4.1	WMA	84.2	Continuous
22	Recycled	4.75	166	30	3.6	HMA	80.6	Continuous
23	Slag	4.76	160	0	3.2	WMA	97.0	Continuous
24	Recycled	5.03	179	20	3.6	WMA	86.5	Continuous
25	Slag	5.57	156	50	1.8	CR	72.7	Batch
26	Slag	4.68	166	60	2.5	WMA	64.2	Continuous
27	Natural	5.92	138	50	3.1	CR	81.9	Continuous

28	Slag	5.41	166	0	4.0	WMA	93.0	Continuous
29	Recycled	4.64	132	60	2.7	CR	75.9	Continuous
30	Recycled	4.57	157	0	2.7	CR	95.6	Continuous
31	Natural	5.73	149	10	4.6	WMA	97.5	Batch
32	Recycled	5.70	142	10	4.0	WMA	94.1	Continuous
33	Recycled	5.89	130	0	3.9	HMA	100.2	Batch
34	Slag	5.16	130	0	4.9	CR	104.4	Continuous
35	Recycled	4.39	138	0	1.6	CR	97.4	Batch
36	Slag	5.60	161	60	4.5	WMA	66.1	Batch
37	Natural	5.82	143	0	4.2	CR	95.6	Batch
38	Slag	5.77	129	40	4.3	HMA	86.6	Batch
39	Recycled	5.58	148	0	3.8	CR	100.6	Batch
40	Slag	5.06	157	50	4.6	CR	77.2	Continuous
41	Recycled	4.07	130	70	1.8	HMA	90.8	Batch
42	Recycled	4.51	145	60	4.0	CR	73.7	Batch
43	Slag	5.70	173	50	2.8	WMA	70.4	Batch
44	Slag	4.26	130	30	2.4	WMA	85.3	Continuous
45	Recycled	5.30	128	20	4.9	WMA	90.2	Continuous
46	Recycled	4.33	140	30	1.7	WMA	87.3	Continuous
47	Recycled	5.17	171	50	2.1	HMA	74.2	Continuous
48	Slag	4.67	176	50	2.1	HMA	67.8	Continuous
49	Slag	5.99	134	30	3.9	CR	80.5	Continuous
50	Recycled	5.94	134	20	1.6	HMA	86.4	Continuous

Fig. 1. Research data has been entered into the orange software.

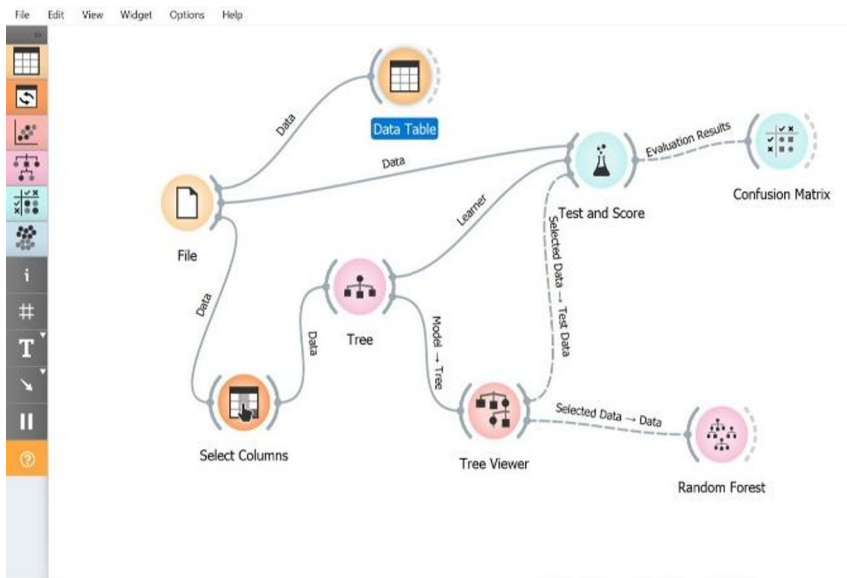


Fig. 2. Visual Workflow Implementation in Orange.

3.2 Results of the Carbon Emission Prediction Model

In this study, a predictive model was developed to estimate the carbon emissions (CO₂ equivalent) generated from asphalt mixtures based on their technical formulation parameters. This model serves as a fundamental component in the intelligent optimization system designed to support environmentally friendly asphalt mix design. A Random Forest Regressor algorithm was used due to its strong performance in handling non-linear relationships, categorical and numerical data, and its robustness against overfitting through ensemble learning (Breiman, 2001).

The dataset utilized consists of 50 records representing various asphalt mix formulations, including key technical parameters such as binder content (%), mixing temperature (°C), proportion of Reclaimed Asphalt Pavement (RAP), aggregate moisture content (%), type of aggregate (natural or recycled), mix type (HMA/WMA/CR), and production method (batch

or continuous). The target variable was the total carbon emissions expressed in kilograms of CO₂ equivalent per ton of asphalt mix.

Before model training, the dataset was split into 80% training data and 20% testing data. Preprocessing steps included categorical encoding and normalization of numerical features to ensure uniformity and improve model performance. The model was trained on the training data and evaluated using the testing set to assess its generalization capability. The evaluation results of the predictive model are summarized as follows:

Table 1. Model Evaluation.

Evaluate Matrix	Result
Root Mean Square Error (RMSE)	5.27 kg CO ₂ /ton
Mean Absolute Error (MAE)	4.99 kg CO ₂ /ton
Coefficient of Determination (R ²)	0.772

These results indicate that the model has a high degree of accuracy, explaining approximately 77.2% of the variance in carbon emissions based on the input variables. The relatively low RMSE and MAE values also confirm that the model's predictions are close to the actual values.

Overall, the model demonstrates good predictive capability and reliability, making it a suitable core component for the fitness function in the Genetic Algorithm-based optimization system. With this level of accuracy, the model enables the efficient evaluation of numerous asphalt mix combinations with reasonable confidence, thereby reducing the need for extensive physical laboratory testing. Thus, it supports the development of a more sustainable, data-driven approach to asphalt mix design aimed at minimizing environmental impact.

3.3 Optimization Results Using Genetic Algorithm

Following the successful development of a predictive model for estimating carbon emissions and the identification of the most influential input variables, the next stage of this research involved optimizing the asphalt mix formulation using a Genetic Algorithm (GA). This algorithm was selected due to its proven effectiveness in solving complex, non-linear optimization problems, especially those involving multiple parameters and large solution spaces. The primary objective of the optimization process was to discover a mix formulation that minimizes carbon emissions (kg CO₂e per ton) while remaining within realistic and technically acceptable parameter limits.

The GA was configured to operate on three key continuous variables: binder content (%), mixing temperature (°C), and RAP content (%), which had been previously identified as the most impactful variables in the feature importance analysis. The bounds for each variable were set as follows:

- Binder content: 4.0% to 6.5%
- Mixing temperature: 100°C to 170°C
- RAP proportion: 0% to 60%

To simplify the problem and reduce model complexity, categorical variables such as aggregate type, mix type, and production method were fixed to environmentally favorable options: recycled aggregate, Warm Mix Asphalt (WMA), and batch production, respectively. These choices were guided by both prior literature and the observed trends in feature contribution.

The fitness function used in the GA was derived from the previously trained Random Forest prediction model. Each candidate solution (i.e., a set of input parameters) was evaluated based on its predicted carbon emissions. Soft penalty terms were applied in the fitness function to discourage impractical solutions, such as those with binder contents exceeding 6.0% or RAP proportions below 10%, to ensure alignment with environmental goals.

After iterating over 40 generations with a population size of 50 individuals per generation, the GA successfully converged toward an optimal solution. The final mix configuration was as follows:

- Binder Content: 4.65%
- Mixing Temperature: 125°C
- RAP Proportion: 48%

This configuration was predicted to yield carbon emissions of approximately 67.5 kg CO_{2e} per ton, which represents a reduction of about 20% compared to standard HMA mixes commonly reported in the literature. The optimization progress was tracked across generations, with a consistent decline in both average and best fitness scores, indicating steady improvement and effective exploration of the solution space.

Table 2. Optimization Results of Asphalt Mix Formulation Using Genetic Algorithm.

Parameter	Unit	Optimal Value	Description
Binder Content	% by total weight	4.65	Lower than the standard HMA range (typically 5–6%)
Mixing Temperature	°C	125	Below the typical HMA temperature (usually 140–170°C)
RAP Proportion	% of total aggregate	48	Approaches the upper limit of practical construction usage
Mix Type	–	WMA	Warm Mix Asphalt selected for energy efficiency
Aggregate Type	–	Recycled	More environmentally friendly than virgin aggregates
Production Method	–	Batch	Commonly used and flexible for RAP integration
CO _{2e} Emissions	kg/ton of mix	67.5	Approximately 20% lower than conventional HMA mixes

These results validate the use of Genetic Algorithms as a robust and adaptive optimization technique in the field of sustainable construction materials. The ability of GA to efficiently search through numerous parameter combinations and identify near-optimal solutions demonstrates its potential as a practical tool for mix design professionals, especially when integrated with predictive modeling systems based on artificial intelligence.

Ultimately, this research highlights that the combination of machine learning and evolutionary optimization represents a powerful decision-support framework for addressing the dual challenge of meeting infrastructure performance standards while minimizing environmental impact.

4 Conclusions

Based on the research findings, the carbon emission prediction model developed using the Random Forest algorithm demonstrated strong performance, achieving a coefficient of

determination (R^2) of 0.772 and a Root Mean Square Error (RMSE) of 5.27 kg CO₂e/ton, indicating its high accuracy in estimating emissions from asphalt mixes based on their technical parameters. The feature importance analysis further revealed that mixing temperature, binder content, and RAP proportion were the most significant factors influencing carbon emissions, highlighting their critical roles in formulating environmentally optimized mixtures. Moreover, the application of the Genetic Algorithm (GA) successfully determined an optimal asphalt mix composition consisting of 4.65% binder content, a mixing temperature of 125°C, and 48% RAP, which is predicted to produce only 67.5 kg CO₂e/ton—representing a 20% reduction in emissions compared to conventional Hot Mix Asphalt (HMA).

References

1. IPCC, *Sixth Assessment Report – Summary for Policymakers*, Intergovernmental Panel on Climate Change, 2021. [Online]. Available: <https://www.ipcc.ch/report/ar6>
2. IEA, *Global Status Report for Buildings and Construction 2022*, International Energy Agency, 2022. [Online]. Available: <https://www.iea.org/reports/global-status-report-for-buildings-and-construction-2022>
3. Chen, Z., Wang, H., & Li, Q., "Carbon emissions and life cycle assessment of asphalt pavement: A review," *J. Cleaner Prod.*, **vol. 346**, p. 131157, 2022.
4. Zaumanis, M., Mallick, R. B., & Frank, R., "Evaluation of different warm mix asphalt technologies: A comprehensive review," *Constr. Build. Mater.*, **vol. 261**, p. 119984, 2020.
5. Raji, A., Ahmad, M., & Basit, A., "Machine learning applications in environmental assessment of construction materials," *Environ. Model. Softw.*, **vol. 157**, p. 105553, 2023.
6. Yousefi, A. A., & Moghadas Nejad, F., "Sustainable asphalt pavement using warm mix technology: A review," *J. Cleaner Prod.*, **vol. 142**, pp. 1725–1736, 2017.
7. Wang, H., Li, X., & Zhao, Y., "Predictive modeling of asphalt emissions using machine learning techniques," *Sustain. Mater. Technol.*, **vol. 36**, p. e00645, 2023.
8. Giani, M. I., Dotelli, G., Brandini, N., & Zampori, L., "Comparative life cycle assessment of warm mix asphalt and hot mix asphalt," *Resour. Conserv. Recycl.*, **vol. 54**, no. 6, pp. 337–345, 2010.
9. Tariq, S., Khan, M. J., & Khan, A., "Predicting environmental impacts of asphalt pavements using machine learning," *Constr. Build. Mater.*, **vol. 343**, p. 128061, 2022.
10. Zortea, R. B., Teixeira, S. R., & Silvestre, J. D., "Life cycle assessment in the construction industry: Opportunities for machine learning," *Renew. Sustain. Energy Rev.*, **vol. 160**, p. 112273, 2022.
11. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T., "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, **vol. 6**, no. 2, pp. 182–197, 2002.
12. Ghosh, S., Ghosh, S., & Dey, A., "Artificial intelligence in civil engineering: A review," *Innov. Infrastruct. Solut.*, **vol. 5**, no. 4, pp. 1–11, 2020.
13. Huang, B., Shu, X., & Chen, X., "Lifecycle assessment of hot mix asphalt and reclaimed asphalt pavement: A review," *Constr. Build. Mater.*, **vol. 67**, pp. 385–396, 2014.

14. Zaumanis, M., & Mallick, R. B., "Review of very high RAP content mix design processes and performance," *Road Mater. Pavement Des.*, **vol. 16**, sup1, pp. 207–236, 2015.
15. Xie, H., Wang, S., & Chen, Q., "Carbon emissions quantification and prediction model of asphalt pavement construction based on machine learning," *J. Cleaner Prod.*, **vol. 286**, p. 125476, 2021.