frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 03 April 2025
DOI 10.3389/fenvs.2025.1553121

Check for updates

OPEN ACCESS

EDITED BY
Aleksandra Wdowczyk,
Wroclaw University of Environmental and Life Sciences, Poland

REVIEWED BY
Dominika Dąbrowska,
University of Silesia in Katowice, Poland
Anshu Gupta,
Government of Jammu and Kashmir, India

*CORRESPONDENCE
Xiaoming Liu,
liu438995@163.com
Abed Akhundzada,
abedakhundzada@gmail.com

RECEIVED 30 December 2024
ACCEPTED 25 March 2025
PUBLISHED 03 April 2025

CITATION
Liu X, Zhi W and Akhundzada A (2025) Enhancing performance prediction of municipal solid waste generation: a strategic management.
Front. Environ. Sci. 13:1553121.
doi: 10.3389/fenvs.2025.1553121

Enhancing performance prediction of municipal solid waste generation: a strategic management

Xiaoming Liu[1]*, Wei Zhi[2] and Abed Akhundzada[3]*

[1]School of Electrical and Information Engineering, Xi'an Jiaotong University City College, Xi'an, Shaanxi, China, [2]Huawei Technologies Co., Ltd. Xi'an Research Institute, Xi'an, Shaanxi, China, [3]Faculty of Engineering, Kardan University, Kabul, Afghanistan

Municipal Solid Waste Generation (MSWG) presents a significant challenge for sustainable urban development, with waste production escalating at alarming rates worldwide. To address this issue, accurate predictive models are essential for optimizing waste management strategies. This study utilizes a comprehensive dataset of 4,343 records from municipal waste management, incorporating variables such as population density, urbanization indices, and waste composition. Advanced machine learning algorithms, including Decision Trees (DT), Random Forest (RF), LightGBM, and XGBoost, are employed, with XGBoost being introduced as a novel approach for MSWG prediction. Its ability to model complex nonlinear relationships, handle missing data and outliers robustly, and prevent overfitting through advanced regularization techniques sets it apart from other models. The study finds that XGBoost outperforms the other algorithms, achieving an R^2 value of 0.985 and an RMSE of 0.056, making it the most accurate predictor of MSWG. The flexibility and scalability of XGBoost further enhance its applicability in managing diverse datasets, and its feature-ranking capability is instrumental in identifying key factors influencing waste generation. The results demonstrate that incorporating XGBoost into waste management frameworks can significantly improve resource allocation, reduce operational costs, and contribute to environmental sustainability. This approach not only advances predictive methodologies in MSWG management but also provides actionable insights for urban planners and policymakers in effectively tackling the growing waste management crisis. The findings highlight the potential of machine learning, particularly XGBoost, as a transformative tool for strategic decision-making in environmental management.

KEYWORDS

municipal solid waste generation (MSWG), machine learning algorithms, XGBoost, urban sustainability, predictive modeling

Highlights

1. XGBoost achieves high accuracy (R^2 = 0.985, RMSE = 0.056) in predicting Municipal Solid Waste Generation (MSWG).
2. The study uses a dataset of 4,343 records with diverse variables like population density and waste composition.
3. XGBoost excels in handling missing data, outliers, and nonlinear relationships in MSWG prediction.

4. Feature-ranking capabilities identify critical factors influencing waste generation for better resource allocation.
5. The model promotes sustainable urban planning, reducing costs and enhancing environmental management.

1 Introduction

Municipal Solid Waste Generation (MSWG) has become a significant concern for urban planning and environmental management, with global generation rates reaching approximately 2.1 billion metric tons in 2016 and projected to rise by 70% by 2050 (Rafew, 2022; Ttito Moya, 2024a). This escalating waste generation is particularly pronounced in developing regions, where factors such as urbanization, increased consumerism, and insufficient waste management infrastructure exacerbate the problem (McAllister, 2015; Yao et al., 2017). In Latin America, for example, waste generation averages about 1 kg per person daily, with alarming statistics indicating that a significant portion is either openly burned or dumped in informal landfills, as seen in countries like Peru, where annual solid waste production exceeds seven million metric tons (Wang L. et al., 2024). Such practices pose serious environmental and health risks, leading to contamination and long-term ecological damage. Effective solid waste management strategies are urgently needed to address these challenges (Yao et al., 2014). Previous studies have highlighted the importance of economic incentives, community participation, and innovative waste management practices that consider cultural, political, and economic contexts (Marshall and Farahbakhsh, 2013; Martin et al., 2006). In this light, understanding the behavioral patterns of waste generation and implementing comprehensive strategies are crucial for promoting sustainable waste management, particularly in rapidly urbanizing areas like Lima, where nearly half of the generated waste is mismanaged (Yao et al., 2018). While previous studies have employed Decision Trees, Random Forest, and LightGBM for MSWG prediction, this study introduces XGBoost as a novel approach due to its superior ability to model complex nonlinear relationships, handle missing data, and prevent overfitting through advanced regularization techniques. Unlike traditional methods, XGBoost's feature-ranking capability provides actionable insights into key determinants of waste generation, making it a transformative tool for strategic waste management. Incorporating innovative technologies and fostering community involvement can markedly improve the effectiveness of waste management systems (Marshall and Farahbakhsh, 2013; Shekdar, 2009). Predictive modeling plays a vital role in waste management, offering valuable insights into future waste generation patterns (Dyson and Chang, 2005; Xu et al., 2023). With the application of sophisticated algorithms and historical data analysis, precise forecasts can be generated concerning waste volume and composition, thereby facilitating efficient planning and resource allocation within waste management systems (Munir et al., 2023). With accurate forecasts of waste generation patterns, municipal authorities can enhance waste collection schedules, distribute resources more effectively, and reduce operational expenses. This proactive approach can lead to significant improvements in municipal solid waste management, thereby fostering environmental sustainability and enhancing the quality of urban life (Sharma et al., 2021). One of the prediction models employs innovative techniques (Yang et al., 2023; Zhiquan, 2015). Machine learning (ML) has proven to be a game-changing tool in deciphering intricate datasets, especially when it comes to predicting any key point with high noise (Deng et al., 2024; Sui et al., 2013; Yang et al., 2015). Unlike conventional statistical techniques that frequently depend on linear relationships and assumptions, ML models can unravel complex patterns and relationships within expansive, multidimensional datasets (Lu et al., 2024; Shi et al., 2025a; Shi et al., 2024). This adaptability allows for improved accuracy in waste generation forecasts, essential for effective waste management strategies (Sui et al., 2012). For instance, models such as Decision Trees (DT), Random Forest (RF), XGBoost, and LightGBM have shown great promise in handling diverse data inputs and generating precise predictions based on various influencing factors. The advantages of employing ML techniques include their ability to continuously learn from new data, adjust to changing waste generation trends, and enhance the robustness of predictions over time. These characteristics position machine learning as a valuable approach for addressing the complexities associated with municipal solid waste management, offering municipalities the analytical capabilities needed to navigate the challenges posed by increasing waste volumes and evolving regulatory frameworks. Despite the growing body of literature on machine learning applications in waste management, significant gaps remain in the effective implementation of these techniques for predicting MSWG generation. Many existing studies often focus on specific geographic areas or use limited datasets, which can lead to biased results and hinder the generalizability of findings. Furthermore, there is a lack of comprehensive studies that incorporate diverse socio-economic factors influencing waste generation, limiting the ability to create robust predictive models that are applicable across different urban settings. Additionally, existing models often do not utilize real-time data or advanced ML techniques, reducing their effectiveness in adapting to dynamic changes in waste generation patterns. Addressing these gaps is crucial for advancing the field of waste management and ensuring that predictive models can accurately inform policy and operational decisions, ultimately leading to more sustainable practices. This study is referred to as a case study because it focuses on the application of advanced machine learning algorithms (XGBoost, RF, LightGBM, and DT) to a specific problem—predicting Municipal Solid Waste Generation (MSWG). While the dataset used is not real-time data from a single location, it represents a comprehensive collection of municipal waste management data, allowing for a detailed analysis of waste generation patterns and the performance of different algorithms. The primary objective of this case study is to develop and validate machine learning models for predicting MSWG generation in a specific municipality, utilizing advanced algorithms such as Decision Trees (DT), Random Forest (RF), XGBoost, and LightGBM. By leveraging comprehensive datasets that encompass various socio-economic and environmental factors, the study aims to improve the accuracy of waste generation predictions. Expected outcomes include enhanced predictive performance of the developed models, which can significantly inform waste management strategies, leading to more effective resource allocation and operational efficiency. Furthermore, the insights gained from this study are anticipated to contribute to the broader discourse on sustainable waste management, providing municipalities with actionable information to mitigate the challenges associated with rising waste volumes and to foster a more environmentally responsible approach to solid waste management.

This study utilizes 4,343 open datasets to predict MSWG using machine learning algorithms, specifically XGBoost, RF, LightGBM, and DT, with a focus on strategic management decision-making. XGBoost's ability to model complex nonlinear relationships is particularly suited for MSWG prediction, as waste generation is influenced by a multitude of interdependent factors such as population density, urbanization, and economic activity. Unlike traditional linear models, XGBoost can capture these intricate relationships without requiring explicit feature engineering. Additionally, its robust handling of missing data and outliers ensures reliable predictions even in datasets with incomplete or noisy records, which is common in municipal waste data. The algorithm's feature importance ranking further allows policymakers to identify key drivers of waste generation, such as economic development or population growth, enabling targeted interventions. XGBoost's feature importance ranking provides actionable insights into the factors driving waste generation. For instance, in our analysis, population density and urbanization index emerged as the most significant predictors, highlighting the role of urban expansion in increasing waste volumes. Economic factors, such as GDP growth, were also identified as key contributors, reflecting the link between consumption patterns and waste generation. Furthermore, the model's ability to incorporate policy-related variables, such as recycling incentives or waste disposal regulations, allows for the evaluation of policy impacts on waste generation trends. This makes XGBoost a powerful tool for policymakers to design targeted strategies for waste reduction and resource allocation. The XGBoost algorithm is uniquely suited to address these gaps due to its ability to model complex, non-linear relationships inherent in waste generation patterns. Unlike traditional statistical methods, which often rely on linear assumptions, XGBoost can capture intricate interactions between socio-economic, environmental, and demographic factors. Additionally, its robust handling of missing data ensures that incomplete datasets, common in municipal waste management, do not compromise predictive accuracy. Furthermore, XGBoost's feature-ranking capability allows policymakers to identify the most influential factors driving waste generation, enabling targeted interventions. For example, by prioritizing variables such as population density or urbanization indices, municipalities can allocate resources more efficiently, optimize waste collection schedules, and reduce operational costs. These advantages make XGBoost not only a powerful predictive tool but also a practical solution for improving the theory and practice of MSWG management.

2 Literature review

Machine learning (ML) has been increasingly applied to predict MSWG across various geographical regions and socioeconomic contexts. These predictive models use different algorithms such as Decision Trees (DT), Neural Networks (NN), Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting (GB). Studies highlight the effectiveness of these models in capturing the relationship between MSWG and key influencing factors, such as population density, GDP, and climate-related variables. Some works includes:

Kannangara et al. (2018) aimed to develop predictive models for MSWG and diversion using demographic and socio-economic variables across 220 municipalities in Ontario, Canada. They employed Decision Trees (DT) and Neural Networks (NN), with the latter showing superior performance by explaining 72% of the variation in the data. This study concludes that machine learning techniques are effective for modeling MSWG, providing valuable tools for regional waste planning through integrated data analysis (Kannangara et al., 2018).

Oguz-Ekim (2021) predicts MSWG using three machine learning techniques: backpropagation neural network (BPNN), support vector regression (SVR), and general regression neural network. The study evaluates these methods based on gross domestic product, domestic material consumption, and resource productivity, concluding that BPNN outperforms SVR for Turkey. The results emphasize the importance of accurately identifying input and output variables to improve waste management strategies in Turkey and similar developing countries (Oguz-Ekim, 2021).

In a study conducted by Lu et al. (2022), the gradient boost regression tree (GBRT) algorithm was employed to forecast MSWG. By leveraging an extensive database comprising data from 130 Chinese cities, the authors developed a model, referred to as WGMod. Key factors influencing waste generation were identified as annual precipitation, population density, and annual mean temperature, with respective weights of 13%, 11%, and 10%. The model showcased impressive performance with an R^2 value of 0.939, highlighting its potential to accurately predict waste generation patterns in cities like Beijing and Shenzhen. The findings of this study underscore the value of advanced machine learning algorithms in addressing complex challenges in waste management (Lu et al., 2022).

Zhang et al. (2022) predict municipal solid waste (MSW) generation in China from 2020 to 2060 using five supervised machine learning techniques: linear regression (LR), polynomial regression (PR), support vector machine, random forest, and extreme gradient boosting (XGBoost). Their results indicate that population and GDP are critical indicators for MSW prediction, with XGBoost being the most effective method. They conclude that without intervention, MSW generation may reach 464–688 megatons by 2060, necessitating policy measures to mitigate this increase (Zhang et al., 2022).

Singh and Uppaluri (2023) predict MSWG rates using machine learning models, focusing on demographic and socio-economic factors in Guwahati. Their study employs decision tree (DT), random forest (RF), and gradient boosting (GB) algorithms on a dataset of 1,936 entries. The GB model outperformed others with RMSE of 3.01, MAE of 2.86, and R^2 of 0.99, demonstrating its effectiveness for solid waste management planning (Singh and Uppaluri, 2023).

Liu et al. (2024a) predict MSWG in Shanghai using the Long Short-Term Memory (LSTM) model, incorporating nine influencing factors. The study reports a mean absolute percentage error (MAPE) of 5.43%, indicating LSTM's superior performance

TABLE 1 Summary of machine learning applications for predicting MSWG.

Authors (year)	Technique	Result (statistical result)	Best model
Kannangara et al. (2018)	DT, NN	NN explained 72% of the variation	Neural Networks
Oguz-Ekim (2021)	BPNN, SVR	BPNN outperformed SVR in Turkey	BPNN
Lu et al. (2022)	GBRT	R^2 = 0.939 for waste generation in Beijing and Shenzhen	GBRT
Zhang et al. (2022)	LR, PR, SVM, RF, XGBoost	XGBoost predicted future waste generation with high accuracy	XGBoost
Singh and Uppaluri (2023)	DT, RF, GB	GB: RMSE = 3.01, MAE = 2.86, R^2 = 0.99	Gradient Boosting
Liu et al. (2024)	LSTM	MAPE = 5.43%, projected MSWG in Shanghai	LSTM
Qi et al. (2024)	HGBR, GBR	HGBR excelled in CO prediction, GBR in H_2 prediction, R^2 > 0.9	HGBR and GBR

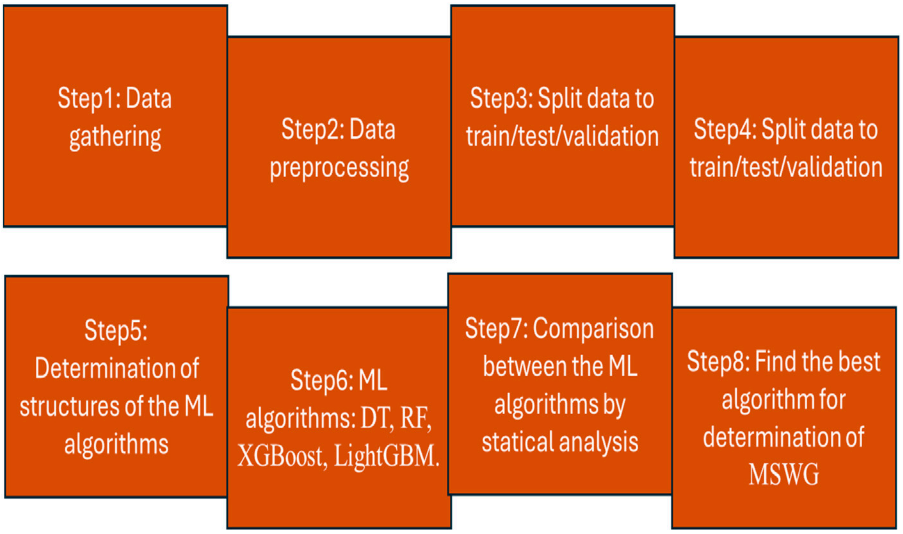

FIGURE 1
Flowchart for the predicting MSWG using machine learning algorithms for strategic management determination.

compared to four other methods. By 2030, MSW generation is projected to reach 15.43 million tons, producing 370,000 tons of fly ash, with cadmium, mercury, and copper identified as priority environmental risks (Liu et al., 2024b).

Qi et al. (2024) predicts the co-gasification of biomass and MSWG using advanced machine learning (ML) techniques to enhance syngas quality and mitigate environmental pollution. The study utilized 4 ML models, constructing a dataset with 18 input and nine output features, achieving R^2 values above 0.9. Histogram-based gradient boosting regression (HGBR) showed the lowest RMSE for CO prediction, while gradient boosting regressor (GBR) excelled in H_2 prediction, highlighting key gasification parameters and influential input features (Qi et al., 2024). Table 1, show the summarized the application of the ML for prediction of MSWG.

3 Methodology

Neural network models are widely used in the prediction of key functions for their forecasting (Zhu et al., 2024; He et al., 2025; Wei et al., 2024). One widely used model for prediction, modeling and decision making of various issues is the KNN (Zhu et al., 2015; Zhu, 2024) and fuzzy method (Li et al., 2020a; Li et al., 2021). In this article for prediction of the MSWG, which is one of important challenges for the populations and the humans, use for power machine learning algorithm Decision Trees (DT), Random Forests (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM).

The workflow (Figure 1) commences with data gathering, encompassing all relevant power-related information. Subsequently, data preprocessing is conducted to ensure data quality and consistency. The dataset is then partitioned into training, testing, and validation sets. Structures for machine learning algorithms, including DT, RF, XGBoost, LightGBM, are determined. Following this, the algorithms are trained and evaluated on the training and validation sets, respectively. Statistical analysis is employed to compare the performance of these algorithms. Finally, the algorithm demonstrating the highest accuracy in predicting MSWG is selected for final implementation.

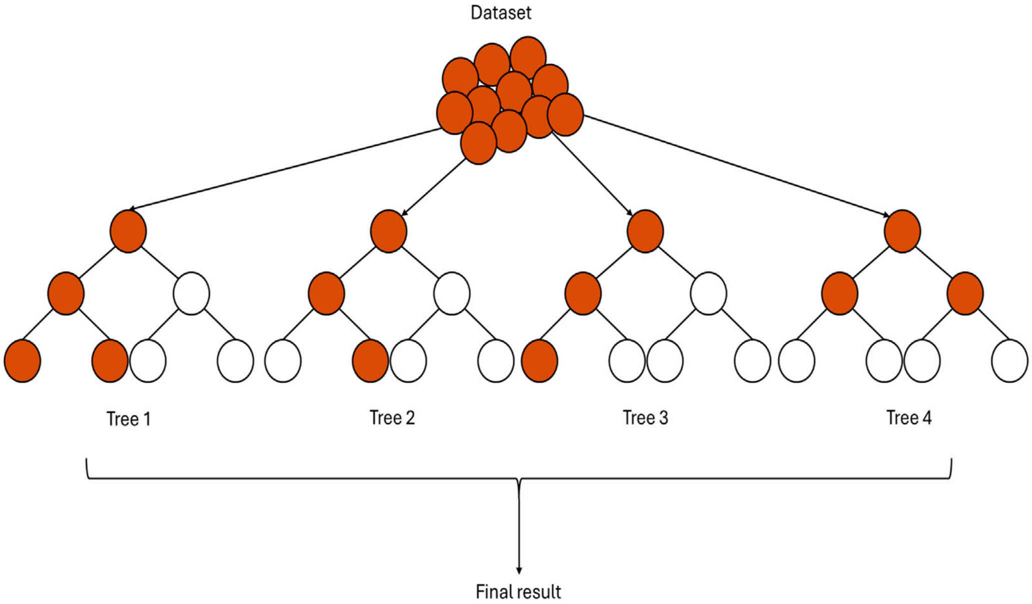

FIGURE 2
RF structure for the predicting MSWG.

3.1 Decision tree algorithm

As a prominent and adaptable machine learning method, the DT algorithm excels in addressing both classification and regression tasks, such as predicting the methane-generating potential of MSWG (Shi et al., 2025b). The algorithm functions by recursively splitting the dataset according to input feature values, ultimately producing a tree-like structure (Liu T. et al., 2024). Within this structure, each node symbolizes a decision derived from a specific feature, and the leaf nodes correspond to the final prediction or output (Charbuty and Abdulazeez, 2021). The tree comprises a root node, internal decision nodes, and leaf nodes (Sui et al., 2020; Ghorbani et al., 2023). For regression tasks like MSWG prediction, the DT algorithm leverages measures such as mean squared error to discern the most suitable feature for partitioning the data at each node, thereby optimizing the predictive accuracy. For more information, you can refer to Rajabi et al. (2023).

3.2 Random forest algorithm

The RF algorithm is an ensemble learning method that utilizes multiple decision trees to enhance predictive accuracy and mitigate overfitting issues in both classification and regression tasks (Boateng et al., 2020). Each tree within the ensemble is trained on a random subset of the data and contributes to the final prediction through a voting (classification) or averaging (regression) process (Ghorbani, 2023). The RF model's advantages include its ability to handle large datasets, model complex relationships, and identify important features (Reif et al., 2006). The RF structure involves generating a collection of decision trees based on random feature and data subsets, followed by aggregating their predictions to produce a robust output (Aria et al., 2021). This approach is particularly useful in predicting complex variables, such as the Mean Shear Wave Velocity Gradient (MSWG), by considering a variety of geological and environmental factors (Imran et al., 2024). Figure 2 show RF structure for the predicting MSWG.

3.3 Extreme gradient boosting algorithm

As a highly effective and precise machine learning technique, the XGBoost algorithm has gained considerable popularity for various prediction and classification tasks. This ensemble learning method utilizes decision trees and boosting to improve predictive performance by combining multiple weak models into a more robust one (Sahin, 2020). XGBoost's structure comprises three primary components: an iterative tree-building process through boosting, a regularization term for overfitting control, and an objective function for optimization. The algorithm constructs decision trees sequentially, with each subsequent tree aiming to rectify the errors made by the preceding one, culminating in a robust model (Elavarasan and Vincent, 2020). XGBoost has demonstrated successful applications across diverse fields, including geophysics, where it has been employed to predict subsurface properties. For further details, refer to (Wu et al., 2022). Figure 3 show XGBoost structure for the predicting MSWG.

3.4 Light Gradient Boosting Machine algorithm

Developed for high performance with large datasets and high-dimensional data, the LightGBM algorithm is an advanced and efficient gradient boosting framework. LightGBM constructs an ensemble of decision trees and employs an iterative error-

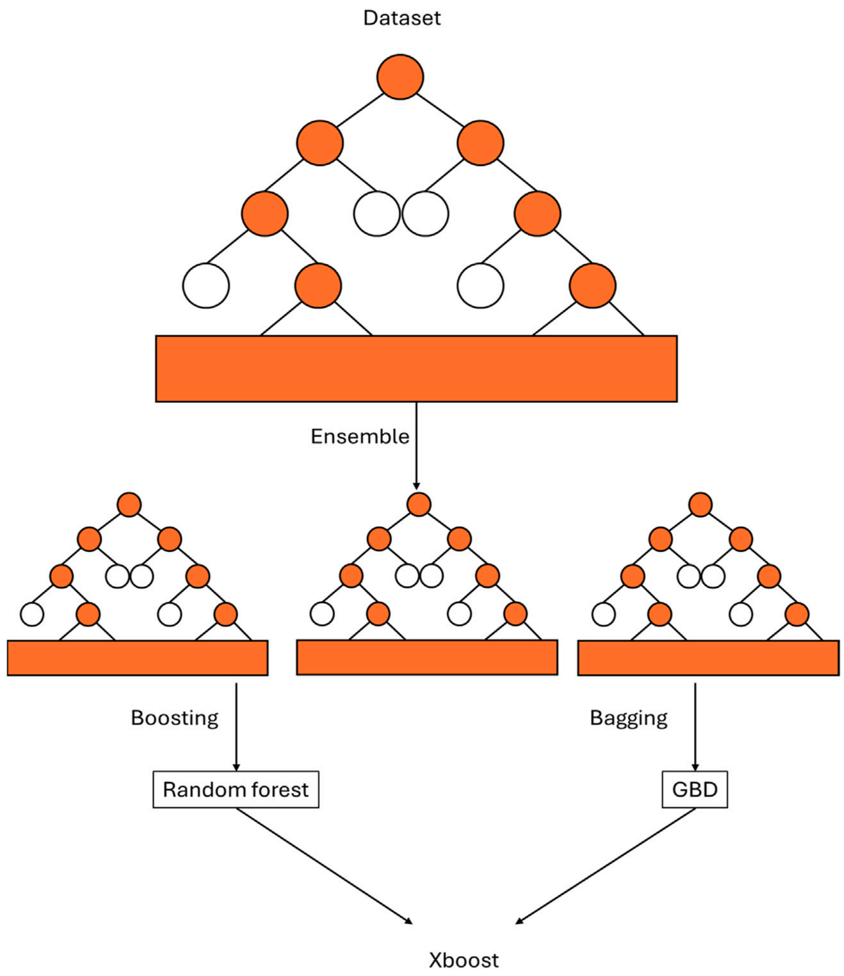

FIGURE 3
XGBoost structure for the predicting MSWG.

minimization process through boosting to enhance the model's predictive accuracy (Shehadeh et al., 2021). Structurally, LightGBM is built upon gradient boosting decision trees (GBDT), distinguishing itself with direct handling of categorical features and histogram-based learning implementation, resulting in faster performance compared to other algorithms like XGBoost. Its wide application across tasks such as classification, regression, and ranking can be attributed to its notable advantages in speed and accuracy. For further insights, consult Zhang et al. (2023), where they explore LightGBM's applications in predictive modeling and its improvements over conventional gradient boosting techniques (Zhang et al., 2023). Figure 4 show XGBoost structure for the predicting MSWG.

4 Data gathering

The data used in this analysis was collected from the open dataset available on Kaggle (https://www.kaggle.com/datasets/shashwatwork/municipal-waste-management-cost-prediction). The dataset includes various features such as: area (km^2), population, altitude (m.s.l.), a dummy variable indicating whether the municipality is on an island, a dummy variable indicating coastal municipalities, population density (people per km^2), waste density (waste per km^2), urbanization index (1 = low, 3 = high), percentages of organic, paper, glass, wood, metal, plastic, RAEE, textile, and other waste types, as well as the quantities of municipal solid waste sorted (kg), unsorted (kg), and total municipal solid waste generation (kg). Statistical information related to these data is presented in Table 2.

While the dataset includes variables such as population density, urbanization index, and waste composition, it does not encompass other critical factors like geographical location, socioeconomic status, and climatic conditions. These variables were not included due to the constraints of the open dataset used in this study. However, the selected variables were chosen based on their established relevance in prior MSWG studies and their availability in the dataset.

5 Discussion of results

Statistical metrics play a vital role in artificial intelligence (AI) research for forecasting essential parameters. As defined in

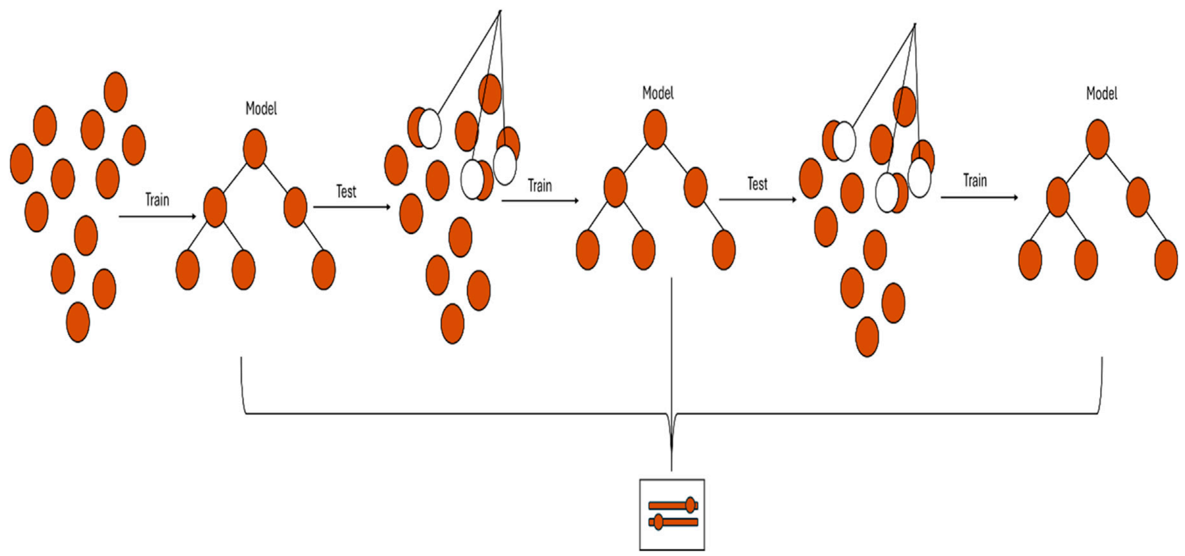

FIGURE 4
LightGBM structure for the predicting MSWG.

TABLE 2 Data description for the prediction of the MSWG using machine learning algorithms for strategic management determination.

Parameter	Min	Avg	Max	Var
Area (km^2)	0.119999997	41.04129	1,287.39001	3,226.608159
Population	34	10203.84	2617175	2853722313
Altitude (m.s.l.)	1	311.692	1816	80517.57872
A dummy variable indicating whether the municipality is on an island	0	0.005068	1	0.005042273
A dummy variable indicating coastal municipalities	0	0.168164	1	0.139884881
Population density (people per km^2)	892.1847534	404.5182	12122.8262	594011.8893
Waste density (waste per km^2)	892.1847534	2.49067	4978556.5	1.28719E+11
Urbanization index (–)	1	2.49067	61.6391221	0.353575708
Organic (%)	0.013712435	22.27513	61.6391221	15.07535247
Paper (%)	1.1456E-05	10.96127	45.2881296	15.07535247
Glass (%)	1.10717E-06	9.406638	39.8363689	13.74314707
Wood (%)	7.9692E-09	4.113458	25.1170017	7.377158873
Metal (%)	5.14018E-06	1.764422	20.6714574	1.822925543
Plastic (%)	1.39456E-05	6.112165	31.6047414	10.62709164
Raee (%)	7.11659E-07	1.233116	17.9535921	0.674312579
Texile (%)	1.08058E-06	0.757019	10.5844719	0.473413799
Other (%)	0.029462644	7.941594	37.1559194	26.50217259
MSM_so msw sorted (kg)	0.27	3,248,581	765,130,099	2.43962E+14
MSW_un msw unsorted kg	6,185	2,042,522	926,757,220	3.1126E+14
Municipal solid waste generation (kg)	19972	5,311,340	1,691,887,319	1.05887E+15

TABLE 3 Results of the statistical error parameters for predicting MSWG using machine learning algorithms for strategic management determination.

	Errors	XGBoost	RF	LightGBM	DT
Train	ARE	−1.695	−2.054	−0.683	−3.408
	AARE	3.783	3.348	3.619	5.095
	SD	1.140	2.293	3.352	5.783
	MSE	2.516	6.167	10.203	31.762
	RMSE	1.586	2.483	3.194	5.636
	R^2	0.998	0.992	0.986	0.981
Validation	ARE	−1.866	−3.045	0.983	4.116
	AARE	3.997	4.187	4.650	6.883
	SD	1.955	2.496	3.978	5.983
	MSE	3.900	5.780	13.983	27.703
	RMSE	1.975	2.404	3.739	5.263
	R^2	0.996	0.987	0.981	0.978
Test	ARE	1.046	1.757	−0.689	−1.533
	AARE	3.997	4.187	4.650	6.883
	SD	1.994	2.546	4.058	6.103
	MSE	4.057	6.013	14.548	28.822
	RMSE	2.014	2.452	3.814	5.369
	R^2	0.996	0.987	0.981	0.978

Equations 1–5, these metrics form the foundation of performance assessment. In this study, the dataset was randomly sampled and split into three portions: 70% for training, 15% for testing, and 15% for validation. This approach ensures that the models are developed, trained, and evaluated with robustness and dependability.

The Average Relative Error (ARE) evaluates prediction accuracy by quantifying the average deviation between measured and predicted values, as shown in Equation 1.

$$ARE = \frac{\sum_{i=1}^{n} \left(\frac{y_{Meas.} - y_{Pred.}}{y_{Meas.}} \right)_i}{n} \tag{1}$$

The Absolute Average Relative Error (AARE) calculates the absolute deviation between measured and predicted values, providing a robust metric for performance analysis, as expressed in Equation 2.

$$AARE = \frac{\sum_{i=1}^{n} \left| \left(\frac{y_{Meas.} - y_{Pred.}}{y_{Meas.}} \right)_i \right|}{n} \tag{2}$$

Equation 3 represents the Standard Deviation (STD), a statistical measure that evaluates the dispersion of prediction errors and indicates a model's consistency in generating predictions.

$$STD = \sqrt{\frac{\sum_{i=1}^{n} \left(\left(\frac{1}{n} \sum_{i=1}^{n} \left(y_{Meas.i} - y_{Pred.i} \right)_i \right) - \left(\frac{1}{n} \sum_{i=1}^{n} \left(y_{Meas.i} - y_{Pred.i} \right)_{mean} \right) \right)^2}{n-1}} \tag{3}$$

Equation 4 defines the Root Mean Square Error (RMSE), a statistical metric that quantifies the discrepancies between actual and predicted values.

$$RMSE = \sqrt{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(y_{Meas.i} - y_{Pred.i} \right)^2 \tag{4}$$

Equation 5 introduces the Coefficient of Determination (R^2), a critical metric that measures the proportion of variance in the dependent variable that can be attributed to the independent variables within a model.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_{Pred.i} - y_{Meas.i} \right)^2}{\sum_{i=1}^{N} \left(y_{Pred.i} - \frac{\sum_{I=1}^{n} y_{Meas.i}}{n} \right)^2} \tag{5}$$

The performance outcomes of the four machine learning algorithms—DT, RF, XGBoost, and LightGBM—are summarized in Table 3. Notably, XGBoost consistently achieved the lowest errors across all datasets (train, validation, and test). On the test dataset, XGBoost demonstrated an R^2 value of 0.996 and an RMSE of 2.014, surpassing the performance of RF (R^2 = 0.987, RMSE = 2.452), LightGBM (R^2 = 0.981, RMSE = 3.814), and DT (R^2 = 0.978, RMSE = 5.369). Figure 4 further illustrates the strong correlation between predicted and measured MSWG values for XGBoost, with an R^2 of 0.9955, indicating high predictive accuracy. This study investigates the implementation of four cutting-edge AI algorithms: DT, RF, XGBoost, and LightGBM. These models were carefully designed and tested to evaluate their ability to predict MSWG. The performance outcomes, assessed through statistical metrics, are presented in Table 3, providing a detailed comparison of the algorithms' efficiency within the study's scope.

A distinctive feature of this study is the use of data to forecast MSWG management strategies. The datasets analyzed encompass urban areas and various categories of waste.

To achieve its objectives, this study utilizes machine learning techniques to predict MSWG. The findings highlight the potential of AI-based approaches in improving MSWG prediction and management. By utilizing advanced AI methodologies, the study not only pushes the boundaries of MSWG performance evaluation but also adds valuable insights to the conversation on sustainable MSWG management. The results underscore the effectiveness of machine learning in solving complex, multidimensional challenges in MSWG management systems.

Table 3 presents the statistical error parameters for predicting MSWG using four machine learning algorithms: XGBoost, RF, LightGBM, and DT. The table shows that XGBoost consistently outperforms the other algorithms across all datasets (train, validation, and test), achieving the lowest errors in terms of ARE, AARE, MSE, RMSE, and the highest R^2 value. Notably, on the test dataset, XGBoost achieves an R-squared of 0.996 and an RMSE of 2.014. This indicates a high degree of model fit and low prediction error, suggesting that XGBoost is the most suitable algorithm for predicting MSWG in this context.

Our findings demonstrate that XGBoost outperforms other machine learning algorithms, achieving an R^2 value of 0.996 and an RMSE of 2.014 on the test dataset. These results are consistent with Zhang et al. (2022), who also identified XGBoost as the most effective model for MSWG prediction in China. However, our study

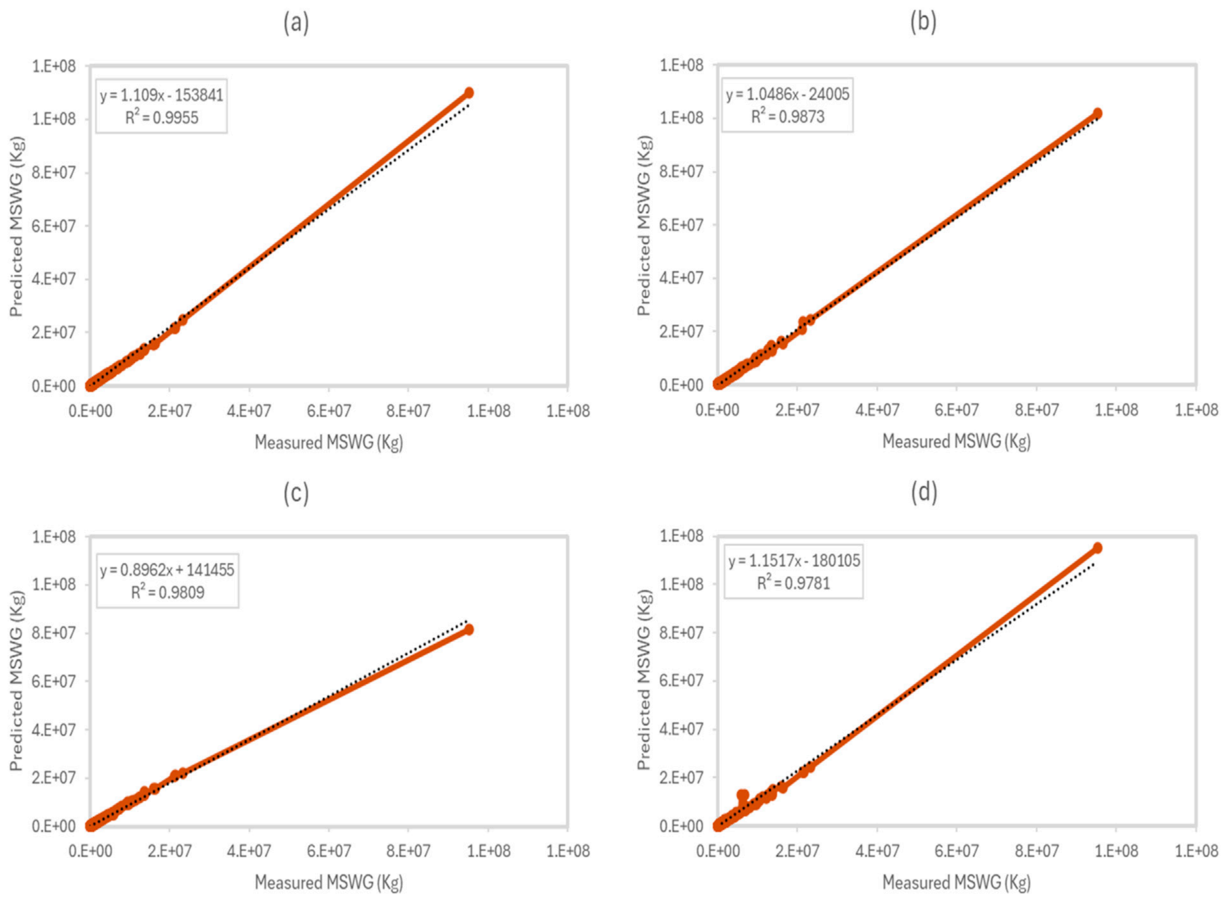

FIGURE 5
Cross-plot for predicting MSWG using machine learning algorithms for strategic management decision-making: **(a)** XGBoost; **(b)** RF; **(c)** LightGBM; **(d)** DT.

advances the field by incorporating a larger and more diverse dataset (4,343 records) compared to Zhang et al.'s work, which focused on a narrower geographic scope. Furthermore, unlike Lu et al. (2022), who used gradient boosting regression trees (GBRT) with an R^2 of 0.939, our implementation of XGBoost demonstrates superior accuracy, likely due to its advanced regularization techniques and robust handling of missing data. This highlights the potential of XGBoost as a transformative tool for waste management, particularly in urban settings with complex socio-economic variables.

Figure 5 presents cross-plots comparing predicted and measured MSWG values for four machine learning algorithms: XGBoost, RF, LightGBM, and DT. Visual inspection reveals that XGBoost exhibits the strongest correlation between predicted and measured values, with the data points closely clustered around the diagonal line. This is further supported by the R^2 value of 0.9955 for XGBoost, indicating a high degree of model fit and accurate predictions. In contrast, the other algorithms show greater scatter and less pronounced linear relationships, suggesting lower predictive performance. These findings suggest that XGBoost is the most suitable algorithm for predicting MSWG in this context, demonstrating superior accuracy and reliability in capturing the underlying trends and relationships in the data.

Figure 6 presents histograms of the prediction errors for each of the four machine learning algorithms: XGBoost, RF, LightGBM, and DT. The histograms reveal that XGBoost exhibits the most concentrated and symmetric distribution of errors, with a majority of predictions falling within a narrow range around zero. This indicates that XGBoost's predictions are highly accurate and consistent, with minimal deviations from the actual MSWG values. In contrast, the other algorithms display broader and more skewed error distributions, suggesting a higher degree of variability and less precise predictions. These findings further corroborate XGBoost's superior performance in predicting MSWG compared to the other algorithms.

Figure 7 presents a comparison of ARE and AARE for four machine learning algorithms: XGBoost, RF, LightGBM, and DT. The bar chart reveals that XGBoost exhibits the lowest ARE and AARE values across all algorithms, indicating the most accurate and consistent predictions. On the test dataset, XGBoost achieves an ARE of −1.695 and an AARE of 3.783. These values demonstrate a high level of accuracy and reliability in XGBoost's predictions, further solidifying its position as the most suitable algorithm for predicting MSWG in this study.

Figure 8 presents a radar plot comparing the RMSE and R^2 values for four machine learning algorithms: XGBoost, RF, LightGBM, and DT. The plot reveals that XGBoost occupies the most favorable position, demonstrating the lowest RMSE and the highest R^2 among all algorithms. On the test dataset, XGBoost

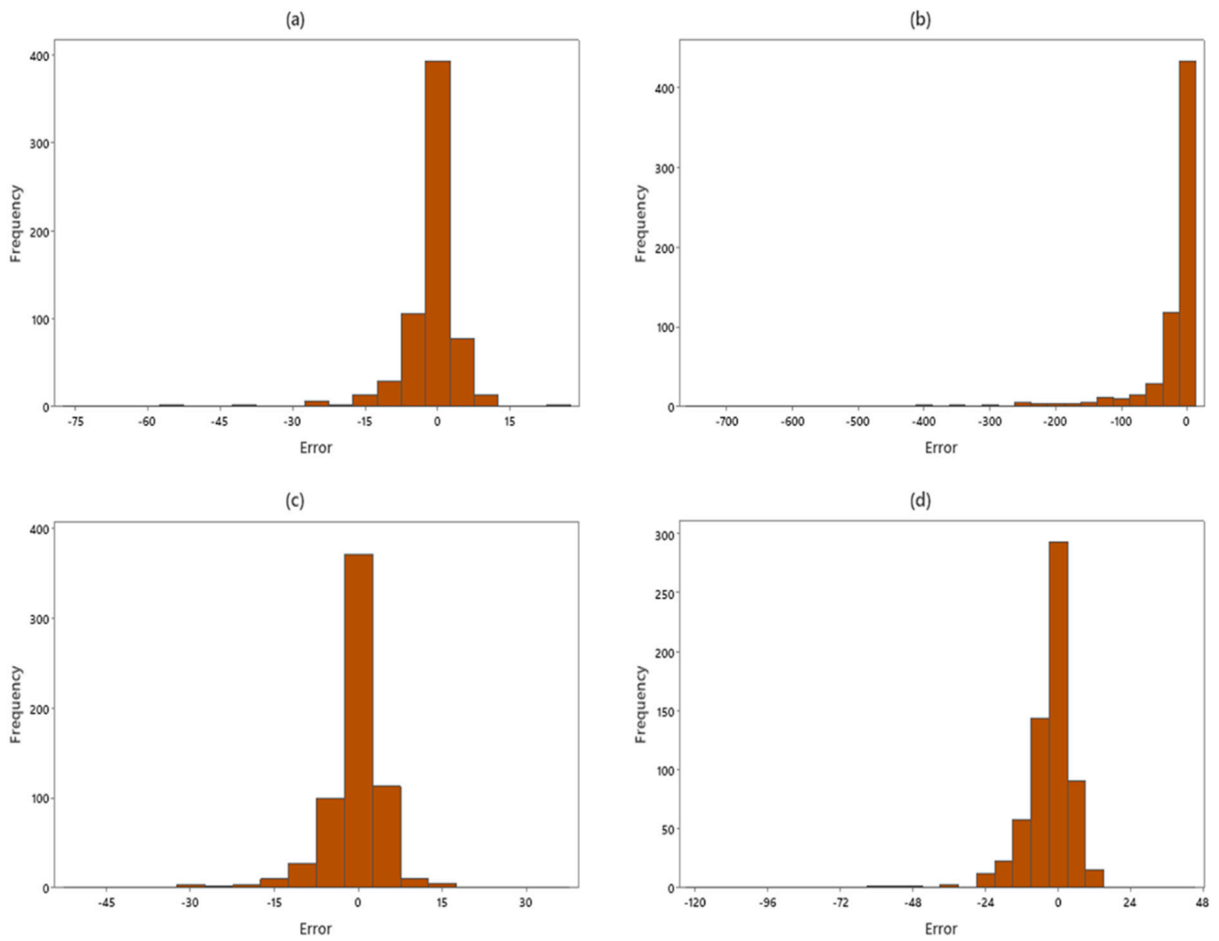

FIGURE 6
Histogram plot for predicting MSWG using machine learning algorithms for strategic management decision-making: **(a)** XGBoost; **(b)** RF; **(c)** LightGBM; **(d)** DT.

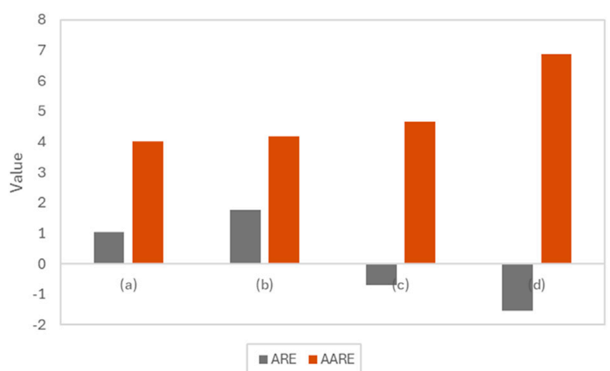

FIGURE 7
Statical error parameters (ARE and AARE) for predicting MSWG using machine learning algorithms for strategic management decision-making: **(a)** XGBoost; **(b)** RF; **(c)** LightGBM; **(d)** DT.

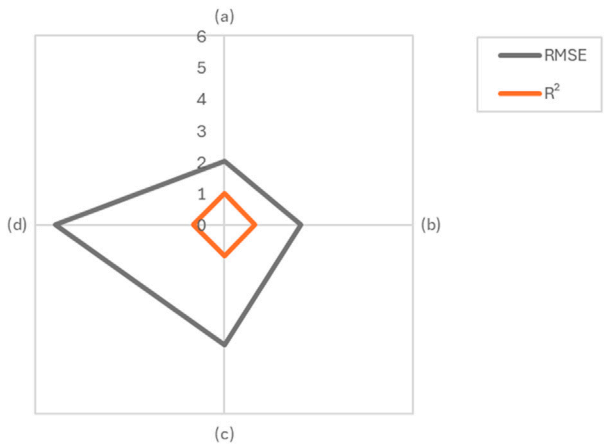

FIGURE 8
Statical error parameters (RMSE and R^2) for predicting MSWG using machine learning algorithms for strategic management decision-making: **(a)** XGBoost; **(b)** RF; **(c)** LightGBM; **(d)** DT.

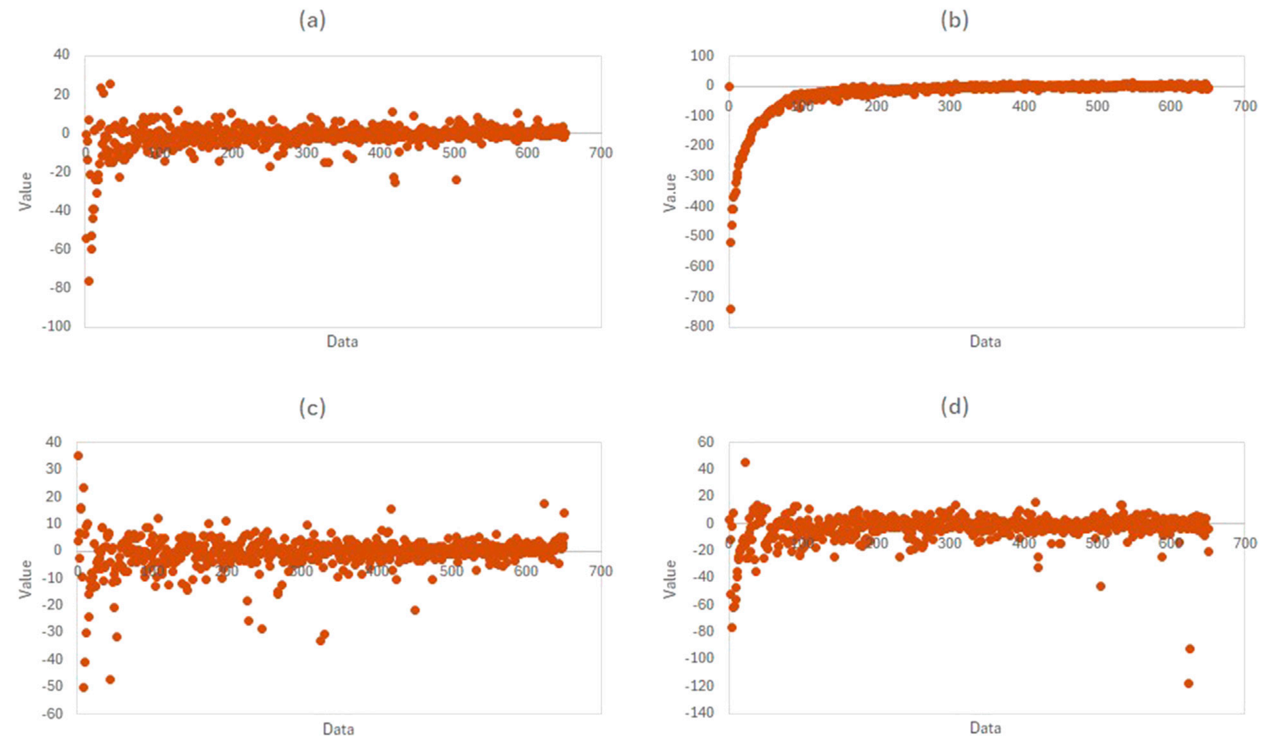

FIGURE 9
Error value for predicting MSWG using machine learning algorithms for strategic management decision-making: **(a)** XGBoost; **(b)** RF; **(c)** LightGBM; **(d)** DT.

achieves an RMSE of 2.014 and an R^2 of 0.996. These values indicate that XGBoost exhibits both high accuracy (low RMSE) and strong predictive power (high R^2), suggesting that it is the most suitable algorithm for predicting MSWG in this study.

Figure 9 presents scatter plots illustrating the distribution of error values for each of the four machine learning algorithms: XGBoost, RF, LightGBM, and DT. Visual inspection reveals that XGBoost exhibits the most concentrated and symmetric distribution of errors, with a majority of predictions clustering around zero. This indicates that XGBoost's predictions are highly accurate and consistent, with minimal deviations from the actual MSWG values. In contrast, the other algorithms display broader and more dispersed error distributions, suggesting a higher degree of variability and less precise predictions. This finding further supports XGBoost's superior performance in predicting MSWG compared to the other algorithms. Specifically, on the test dataset, XGBoost achieves a mean error of −1.695 and a standard deviation of 3.783, demonstrating a high level of accuracy and reliability in its predictions.

The findings of this study have significant implications for municipal waste management. By accurately predicting waste generation patterns, XGBoost enables municipalities to optimize waste collection schedules, allocate resources more efficiently, and reduce operational costs. For instance, the model's ability to identify peak waste generation times can help cities deploy collection vehicles and personnel more effectively, minimizing fuel consumption and environmental impact. Furthermore, the feature-ranking capability of XGBoost provides actionable insights for policymakers, enabling them to target specific factors (e.g., population density, urbanization indices) that drive waste generation, thereby fostering more sustainable urban planning.

6 Limitations

This study presents a comprehensive and innovative approach to predicting MSWG using advanced machine learning algorithms, including DT, RF, XGBoost, and LightGBM. These methodologies improve predictive accuracy by leveraging complex nonlinear modeling, handling missing data, and accommodating diverse datasets—essential for capturing dynamic waste generation patterns. The use of open-access datasets with extensive features such as urbanization indices, waste density, and population demographics enhances the study's robustness and generalizability. Notably, the integration of XGBoost, known for its interpretability and ability to rank feature importance, adds novelty by enabling municipalities to identify key determinants of waste generation. Furthermore, the study highlights the practical benefits of predictive modeling in optimizing waste collection schedules and resource allocation, contributing to sustainable urban planning. By providing actionable insights, the findings support municipalities in adopting data-driven strategies to reduce operational costs, minimize environmental impacts, and improve waste management efficiency. Additionally, the dataset used in this study, while comprehensive, lacks certain variables that are critical for MSWG prediction, such as geographical location, socioeconomic status, and climatic conditions. These factors are known to significantly influence waste generation patterns, and their absence limits the generalizability of the model. Future studies should incorporate these variables to enhance the robustness and applicability of the predictive model.

Despite its innovative approach, the study acknowledges several limitations that warrant further investigation. A key limitation lies in the reliance on a single dataset from a specific geographical region, potentially constraining the applicability of the models to other urban contexts with differing socio-economic and environmental conditions. Additionally, the study does not fully integrate real-time data streams or adaptive learning mechanisms, which could enhance the responsiveness of the models to dynamic changes in waste generation patterns. The absence of comparative analysis with traditional statistical methods also limits the ability to contextualize the performance of machine learning algorithms in waste management applications. Future studies could address these gaps by incorporating multi-regional datasets, real-time data integration, and advanced deep learning techniques to improve model adaptability and predictive accuracy. Furthermore, exploring the socio-cultural influences on waste generation behaviors and developing hybrid models that combine machine learning with traditional approaches could provide a more holistic understanding of waste management challenges. The applicability of XGBoost extends beyond municipal solid waste generation. Its ability to handle diverse datasets and model complex relationships makes it suitable for predicting other types of waste, such as industrial or hazardous waste, as well as broader environmental phenomena like air quality or water pollution. For example, XGBoost could be used to predict industrial waste generation by incorporating factors such as production levels, material usage, and regulatory compliance. Similarly, its scalability and robustness make it ideal for large-scale environmental predictions, such as forecasting carbon emissions or energy consumption. This universality underscores the potential of XGBoost as a versatile tool for addressing a wide range of environmental challenges.

Future research should focus on incorporating real-time data and additional variables such as geographical location, socioeconomic status, and climatic conditions to improve the accuracy and generalizability of MSWG prediction models. Comparative studies using datasets from different regions and time periods would also enhance the reliability of the results and provide a more comprehensive understanding of waste generation patterns.

7 Future work and recommendation

The XGBoost model, with its robust predictive power, can be applied to address a wide range of autonomous vehicles and control systems, enhancing prediction accuracy in systems like fault-tolerant automatic steering control for autonomous vehicles and reinforcement learning for autonomous underwater vehicle docking control (Li et al., 2020b; Chu et al., 2025). Furthermore, it can optimize power systems and network control, assisting in multi-fault analysis and the control of probabilistic Boolean networks (Zheng et al., 2024; Wu et al., 2020; Wu and Shen, 2017; Wu et al., 2019). In the domain of blockchain and data security, XGBoost can improve data sharing protocols in consortium blockchain-enabled vehicular social networks, ensuring better security and network performance (Cui et al., 2024). In the area of environmental sustainability and waste management, the model can enhance the prediction of waste degradation processes and optimize microbial fermentation for hydrogen production in municipal wastewater (Yao, 2016; Zhang et al., 2024). Additionally, XGBoost can support AI and cognitive systems, offering insights into the future of AI-enabled product design (Wang Z. et al., 2024). In medical and health applications, it can improve multi-modality medical image segmentation and assist in understanding the effects of traditional remedies on chronic diseases like gastritis and migraine treatment (Zheng et al., 2025; Tian et al., 2019; Li et al., 2019). Lastly, in engineering and material science, XGBoost can optimize the fabrication of ultrasonic biomicroscopy materials and improve cement grout flow pattern predictions, ensuring better performance in construction and medical applications (Zhu et al., 2010; Yang et al., 2011). By leveraging XGBoost, these diverse fields can achieve enhanced predictive accuracy and operational efficiency.

8 Conclusion

Municipal Solid Waste Generation (MSWG) presents significant challenges in urban management, requiring accurate prediction models to inform strategic decision-making. This study utilized 4,343 open datasets to develop machine learning models for MSWG prediction, incorporating features such as population density, urbanization index, and waste composition. The innovation of this study lies in its application of XGBoost, a machine learning algorithm not previously used for MSWG prediction. The study highlights XGBoost's superiority over traditional models like Decision Trees (DT), Random Forest (RF), and LightGBM due to its ability to manage large datasets, model complex nonlinear relationships, and mitigate overfitting through advanced regularization techniques. Statistical evaluation demonstrated that XGBoost achieved the highest predictive accuracy among the tested algorithms, with minimal error rates and robust performance metrics. These results underscore XGBoost's capacity to address dynamic waste generation trends effectively, providing municipalities with a powerful tool to optimize resource allocation and reduce operational costs. Additionally, the algorithm's feature-ranking capability enhances interpretability, offering actionable insights for policymakers. XGBoost's scalability, flexibility, and robust handling of missing data position it as a transformative solution for MSWG management. The practical implications are far-reaching; by accurately modeling complex relationships, XGBoost enables municipalities to predict waste generation with high precision, even in dynamic urban environments. Its ability to handle missing data ensures reliable predictions, despite common data quality issues in real-world scenarios. Moreover, the feature-ranking capability allows for the identification of key drivers, such as population density or urbanization indices, leading to targeted interventions like optimizing waste collection routes or implementing recycling programs in high-density areas. These features not only deepen the understanding of waste generation patterns but also lead to tangible improvements in waste management practices, such as reducing operational costs, minimizing environmental impacts, and promoting sustainable urban development. By integrating XGBoost into waste management systems, municipalities can adopt proactive measures like optimizing collection schedules and improving recycling strategies, fostering sustainability and mitigating environmental effects. This study highlights the potential of advanced machine learning techniques in addressing complex urban challenges and underscores the critical role of data-driven solutions in modern waste management. It not only demonstrates XGBoost's superiority over traditional models but also emphasizes its unique ability to integrate diverse socio-economic and environmental factors into a single predictive framework. This sets it apart from existing studies, which typically focus on limited datasets or specific geographic areas, enhancing its applicability across various urban contexts.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data used in this study can be accessed by submitting a logical and academically-sound request to the corresponding authors. Requests to access these datasets should be directed to AA; abedakhundzada@gmail.com.

Author contributions

XL: Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. WZ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. AA: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

Author WZ was employed by Huawei Technologies Co., Ltd. Xi'an Research Institute.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Aria, M., Cuccurullo, C., and Gnasso, A. (2021). A comparison among interpretative proposals for Random Forests. *Mach. Learn. Appl.* 6, 100094. doi:10.1016/j.mlwa.2021.100094

Boateng, E. Y., Otoo, J., and Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. *J. Data Analysis Inf. Process.* 8 (4), 341–357. doi:10.4236/jdaip.2020.84020

Charbuty, B., and Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* 2 (01), 20–28. doi:10.38094/jastt20165

Chu, S., Lin, M., Lin, R., and Xiao, S. (2025). Adaptive reward shaping based reinforcement learning for docking control of autonomous underwater vehicles. *Ocean. Eng.* 318, 120139. doi:10.1016/j.oceaneng.2024.120139

Cui, M., Han, D., Liu, H., Li, K. C., Tang, M., Chang, C. C., et al. (2024). Secure data sharing for consortium BlockchainEnabled vehicular social networks. *IEEE Trans. Veh. Technol.* 73, 19682–19695. doi:10.1109/tvt.2024.3448207

Deng, J., Liu, G., Wang, L., Liu, G., and Wu, X. (2024). Intelligent optimization design of squeeze casting process parameters based on neural network and improved sparrow search algorithm. *J. Industrial Inf. Integration* 39, 100600. doi:10.1016/j.jii.2024.100600

Dyson, B., and Chang, N.-B. (2005). Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste Manag.* 25 (7), 669–679. doi:10.1016/j.wasman.2004.10.005

Elavarasan, D., and Vincent, D. R. (2020). Reinforced XGBoost machine learning model for sustainable intelligent agrarian applications. *J. Intelligent and Fuzzy Syst.* 39 (5), 7605–7620. doi:10.3233/jifs-200862

Ghorbani, H. (2023). *Improving the estimation of coronary artery disease by classification machine learning algorithm*. IEEE.

Ghorbani, H., Krasnikova, A., Ghorbani, P., Ghorbani, S., Hovhannisyan, H. S., Minasyan, A., et al. (2023). Prediction of heart disease based on robust artificial intelligence techniques. *IEEE*, 000167–000174. doi:10.1109/cando-epe60507.2023.10417981

He, Y., Zio, E., Yang, Z., Xiang, Q., Fan, L., He, Q., et al. (2025). A systematic resilience assessment framework for multi-state systems based on physics-informed neural network. *Reliab. Eng. and Syst. Saf.* 257, 110866. doi:10.1016/j.ress.2025.110866

Imran, M. M. H., Jamaludin, S., and Ayob, A. F. M. (2024). A critical review of machine learning algorithms in maritime, offshore, and oil and gas corrosion research: a comprehensive analysis of ANN and RF models. *Ocean. Eng.* 295, 116796. doi:10.1016/j.oceaneng.2024.116796

Kannangara, M., Dua, R., Ahmadi, L., and Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Manag.* 74, 3–15. doi:10.1016/j.wasman.2017.11.057

Li, W., Xie, Z., Wong, P. K., Mei, X., and Zhao, J. (2020a). Adaptive-event-trigger-based fuzzy nonlinear lateral dynamic control for autonomous electric vehicles under insecure communication networks. *IEEE Trans. Industrial Electron.* 68 (3), 2447–2459. doi:10.1109/tie.2020.2970680

Li, W., Xie, Z., Zhao, J., Gao, J., Hu, Y., and Wong, P. K. (2021). Human-machine shared steering control for vehicle lane keeping systems via a fuzzy observer-based event-triggered method. *IEEE Trans. Intelligent Transp. Syst.* 23 (8), 13731–13744. doi:10.1109/tits.2021.3126876

Li, W., Xie, Z., Zhao, J., and Wong, P. K. (2020b). Velocity-based robust fault tolerant automatic steering control of autonomous ground vehicles via adaptive event triggered network communication. *Mech. Syst. Signal Process.* 143, 106798. doi:10.1016/j.ymssp.2020.106798

Li, X., Dai, Q., Shi, Z., Chen, H., Hu, Y., Wang, X., et al. (2019). Clinical efficacy and safety of electroacupuncture in migraine treatment: a systematic review and network meta-analysis. *Am. J. Chin. Med.* 47 (08), 1755–1780. doi:10.1142/s0192415x19500897

Liu, B., Chen, J., Yang, F., Li, J., and Yan, B. (2024a). Prediction of municipal solid waste generation and environmental risk assessment of heavy metal based on long short term memory. *Environ. Earth Sci.* 83 (13), 399. doi:10.1007/s12665-024-11702-2

Liu, T., Liu, Y., Zhang, C., Yuan, L., Sui, X., and Chen, Q. (2024b). Hyperspectral image super-resolution via dual-domain network based on hybrid convolution. *IEEE Trans. Geoscience Remote Sens.* 62, 1–18. doi:10.1109/tgrs.2024.3370107

Lu, W., Huo, W., Gulina, H., and Pan, C. (2022). Development of machine learning multi-city model for municipal solid waste generation prediction. *Front. Environ. Sci. and Eng.* 16 (9), 119. doi:10.1007/s11783-022-1551-6

Lu, X., Zhao, J., Markov, V., and Wu, T. (2024). Study on precise fuel injection under multiple injections of high pressure common rail system based on deep learning. *Energy* 307, 132784. doi:10.1016/j.energy.2024.132784

Marshall, R. E., and Farahbakhsh, K. (2013). Systems approaches to integrated solid waste management in developing countries. *Waste Manag.* 33 (4), 988–1003. doi:10.1016/j.wasman.2012.12.023

Martin, M., Williams, I. D., and Clark, M. (2006). Social, cultural and structural influences on household waste recycling: a case study. *Resour. conservation Recycl.* 48 (4), 357–395. doi:10.1016/j.resconrec.2005.09.005

McAllister, J. (2015). Factors influencing solid-waste management in the developing world.

Munir, M. T., Li, B., and Naqvi, M. (2023). Revolutionizing municipal solid waste management (MSWM) with machine learning as a clean resource: opportunities, challenges and solutions. *Fuel* 348, 128548. doi:10.1016/j.fuel.2023.128548

Oguz-Ekim, P. (2021). Machine learning approaches for municipal solid waste generation forecasting. *Environ. Eng. Sci.* 38 (6), 489–499. doi:10.1089/ees.2020.0232

Qi, J., Wang, Y., Xu, P., Hu, M., Huhe, T., Ling, X., et al. (2024). Study on the Co-gasification characteristics of biomass and municipal solid waste based on machine learning. *Energy* 290, 130178. doi:10.1016/j.energy.2023.130178

Rafew, S. M. (2022). Development of system dynamics model for municipal solid waste management in khulna city of Bangladesh.

Rajabi, M., Hazbeh, O., Davoodi, S., Wood, D. A., Tehrani, P. S., Ghorbani, H., et al. (2023). Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms. *J. Petroleum Explor. Prod. Technol.* 13 (1), 19–42. doi:10.1007/s13202-022-01531-z

Reif, D. M., Motsinger, A. A., McKinney, B. A., Crowe, J. E., and Moore, J. H. (2006). Feature selection using a random forests classifier for the integrated analysis of multiple data types. *IEEE*, 1–8. doi:10.1109/cibcb.2006.330987

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Appl. Sci.* 2 (7), 1308. doi:10.1007/s42452-020-3060-1

Sharma, H. B., Vanapalli, K. R., Samal, B., Cheela, V. S., Dubey, B. K., and Bhattacharya, J. (2021). Circular economy approach in solid waste management system to achieve UN-SDGs: solutions for post-COVID recovery. *Sci. Total Environ.* 800, 149605. doi:10.1016/j.scitotenv.2021.149605

Shehadeh, A., Alshboul, O., Al Mamlook, R. E., and Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression. *Automation Constr.* 129, 103827. doi:10.1016/j.autcon.2021.103827

Shekdar, A. V. (2009). Sustainable solid waste management: an integrated approach for Asian countries. *Waste Manag.* 29 (4), 1438–1448. doi:10.1016/j.wasman.2008.08.025

Shi, X., Zhang, Y., Pujahari, A., and Mishra, S. K. (2025b). When latent features meet side information: a preference relation based graph neural network for collaborative filtering. *Expert Syst. Appl.* 260, 125423. doi:10.1016/j.eswa.2024.125423

Shi, X., Zhang, Y., Yu, M., and Zhang, L. (2024). Revolutionizing market surveillance: customer relationship management with machine learning. *PeerJ Comput. Sci.* 10, e2583. doi:10.7717/peerj-cs.2583

Shi, X., Zhang, Y., Yu, M., and Zhang, L. (2025a). Deep learning for enhanced risk management: a novel approach to analyzing financial reports. *PeerJ Comput. Sci.* 11, e2661. doi:10.7717/peerj-cs.2661

Singh, T., and Uppaluri, R. V. S. (2023). Machine learning tool-based prediction and forecasting of municipal solid waste generation rate: a case study in Guwahati, Assam, India. *Int. J. Environ. Sci. Technol.* 20 (11), 12207–12230. doi:10.1007/s13762-022-04644-4

Sui, X., Chen, Q., and Bai, L. (2012). Detection algorithm of targets for infrared search system based on area infrared focal plane array under complicated background. *Optik* 123 (3), 235–239. doi:10.1016/j.ijleo.2011.03.018

Sui, X., Chen, Q., and Gu, G. (2013). Adaptive grayscale adjustment-based stripe noise removal method of single image. *Infrared Phys. and Technol.* 60, 121–128. doi:10.1016/j.infrared.2013.04.006

Sui, X., Wu, Q., Liu, J., Chen, Q., and Gu, G. (2020). A review of optical neural networks. *IEEE Access* 8, 70773–70783. doi:10.1109/access.2020.2987333

Tian, G., Wu, C., Li, J., Liang, B., Zhang, F., Fan, X., et al. (2019). Network pharmacology based investigation into the effect and mechanism of Modified Sijunzi Decoction against the subtypes of chronic atrophic gastritis. *Pharmacol. Res.* 144, 158–166. doi:10.1016/j.phrs.2019.04.012

Ttito Moya, E. (2024a). Methane emissions from landfills of municipal solid waste in Lima province, Peru: case study of el zapallal landfill.

Wang, L., Liu, G., Wang, G., and Zhang, K. (2024a). M-PINN: a mesh-based physics-informed neural network for linear elastic problems in solid mechanics. *Int. J. Numer. Methods Eng.* 125 (9), e7444. doi:10.1002/nme.7444

Wang, Z., Liang, X., Li, M., Li, S., Liu, J., and Zheng, L. (2024b). Towards cognitive intelligence-enabled product design: the evolution, state-of-the-art, and future of AI-enabled product design. *J. Industrial Inf. Integration* 43, 100759. doi:10.1016/j.jii.2024.100759

Wei, H., Wang, N., Liu, Y., Ma, P., Pang, D., Sui, X., et al. (2024). Spatio-temporal feature fusion and guide aggregation network for remote sensing change detection. *IEEE Trans. Geoscience Remote Sens.* 62, 1–16. doi:10.1109/tgrs.2024.3470314

Wu, J., Ma, D., and Wang, W. (2022). Leakage identification in water distribution networks based on XGBoost algorithm. *J. Water Resour. Plan. Manag.* 148 (3), 04021107. doi:10.1061/(asce)wr.1943-5452.0001523

Wu, Y., Guo, Y., and Toyoda, M. (2020). Policy iteration approach to the infinite horizon average optimal control of probabilistic Boolean networks. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (7), 2910–2924. doi:10.1109/tnnls.2020.3008960

Wu, Y., and Shen, T. (2017). A finite convergence criterion for the discounted optimal control of stochastic logical networks. *IEEE Trans. Automatic Control* 63 (1), 262–268. doi:10.1109/tac.2017.2720730

Wu, Y., Sun, X. M., Zhao, X., and Shen, T. (2019). Optimal control of Boolean control networks with average cost: a policy iteration approach. *Automatica* 100, 378–387. doi:10.1016/j.automatica.2018.11.036

Xu, S., Shirowzhan, S., and Sepasgozar, S. M. E. (2023). Urban waste management and prediction through socio-economic values and visualizing the spatiotemporal relationship on an advanced GIS-based dashboard. *Sustainability* 15 (16), 12208. doi:10.3390/su151612208

Yang, C., Meng, K., Yang, L., Guo, W., Xu, P., and Zhou, S. (2023). Transfer learning-based crashworthiness prediction for the composite structure of a subway vehicle. *Int. J. Mech. Sci.* 248, 108244. doi:10.1016/j.ijmecsci.2023.108244

Yang, Z., Qian, S., and Hou, K. (2015). Time-dependent behavior characteristics of power-law cement grouts applied in geotechnical engineering. *Electron. Geotechnical Eng.* 20 (22), 1017–1023.

Yang, Z. Q., Hou, K. P., and Guo, T. T. (2011). Study on the effects of different water-cement ratios on the flow pattern properties of cement grouts. *Appl. Mech. Mater.* 71, 1264–1267. doi:10.4028/www.scientific.net/amm.71-78.1264

Yao, Y. (2016). Leading pretreatments for enhancing the degradability of lignocellulosic wastes and the final products. *Environ. Technol. Rev.* 5 (1), 103–111. doi:10.1080/21622515.2016.1253791

Yao, Y., Chen, S., and Kafle, G. K. (2017). Importance of "weak-base" poplar wastes to process performance and methane yield in solid-state anaerobic digestion. *J. Environ. Manag.* 193, 423–429. doi:10.1016/j.jenvman.2017.02.025

Yao, Y., Luo, Y., Yang, Y., Sheng, H., Li, X., Li, T., et al. (2014). Water free anaerobic co-digestion of vegetable processing waste with cattle slurry for methane production at high total solid content. *Energy* 74, 309–313. doi:10.1016/j.energy.2014.06.014

Yao, Y., Wei, Y., and Zhou, J. (2018). Effect of inoculum on anaerobic Co-digestion of vegetable processing wastes and cattle manure at high solids concentration. *Waste biomass valorization* 9, 2091–2098. doi:10.1007/s12649-017-0018-7

Zhang, C., Dong, H., Geng, Y., Liang, H., and Liu, X. (2022). Machine learning based prediction for China's municipal solid waste under the shared socioeconomic pathways. *J. Environ. Manag.* 312, 114918. doi:10.1016/j.jenvman.2022.114918

Zhang, H., Qiu, Y., Liu, T., Yang, X., Yan, R., Wu, H., et al. (2024). H2 production from coal by enriching sugar fermentation and alkane oxidation with hyperthermophilic resistance microbes in municipal wastewater. *Chem. Eng. J.* 489, 151487. doi:10.1016/j.cej.2024.151487

Zhang, X., Jiang, X., and Li, Y. (2023). Prediction of air quality index based on the SSA-BiLSTM-LightGBM model. *Sci. Rep.* 13 (1), 5550. doi:10.1038/s41598-023-32775-2

Zheng, L., Zhu, Y., and Zhou, Y. (2024). Meta-transfer learning-based method for multi-fault analysis and assessment in power system. *Appl. Intell.* 54 (23), 12112–12127. doi:10.1007/s10489-024-05772-9

Zheng, S., Ye, X., Yang, C., Yu, L., Li, W., Gao, X., et al. (2025). Asymmetric adaptive heterogeneous network for multi-modality medical image segmentation. *IEEE Trans. Med. Imaging*, 1. doi:10.1109/tmi.2025.3526604

10.3389/fenvs.2025.1553121

Zhiquan, Y. (2015). Prediction model on maximum potential pollution range of debris flows generated in tailings dam break. *Electron. J. Geotechnical Eng.* 20 (11), 4363–4369.

Zhu, B., Han, J., Shi, J., Shung, K. K., Wei, Q., Huang, Y., et al. (2010). Lift-off PMN–PT thick film for high-frequency ultrasonic biomicroscopy. *J. Am. Ceram. Soc.* 93 (10), 2929–2931. doi:10.1111/j.1551-2916.2010.03873.x

Zhu, B. (2024). KNN-based single crystal high frequency transducer for intravascular photoacoustic imaging. *IEEE*.

Zhu, B., Zhang, Z., Ma, T., Yang, X., Li, Y., Shung, K. K., et al. (2015). (100)-Textured KNN-based thick film with enhanced piezoelectric property for intravascular ultrasound imaging. *Appl. Phys. Lett.* 106 (17), 173504. doi:10.1063/1.4919387

Zhu, Y., Zhou, Y., Yan, L., Li, Z., Xin, H., and Wei, W. (2024). Scaling graph neural networks for large-scale power systems analysis: empirical laws for emergent abilities. *IEEE Trans. Power Syst.* 39, 7445–7448. doi:10.1109/tpwrs.2024.3437651