Journal of Environmental Management 377 (2025) 124627

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Machine learning-based analysis of microplastic-induced changes in anaerobic digestion parameters influencing methane yield

Zhenghui Gao, Zongqiang Ren, Tianyi Cui, Yao Fu *

School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK

A R T I C L E I N F O

Keywords:
Anaerobic digestion
Explainable analysis
Machine learning
Methane yield
Microplastic

A B S T R A C T

Microplastics (MPs) present significant challenges for anaerobic digestion (AD) processes used in energy recovery from contaminated organic waste. Given that optimal AD conditions vary widely across studies when MPs are present, a robust predictive model is essential to accurately assess these complex effects. This study applied four machine learning algorithms to predict methane yield using two datasets—one with and one without MPs. Among these, gradient boosting regression demonstrated the highest prediction accuracy, with testing R^2 values of 0.996 for systems without MP pollution and 0.998 with MP pollution. This model was then further optimized by removing redundant and low-importance features, refining its predictive power. Feature importance analysis revealed that digestion time and substrate organic matter content were key parameters positively correlated with methane production. In the presence of MPs, substrate pH and inoculum total solids emerged as critical factors, with partial dependence plots offering deeper insights into their optimal conditions. This research offers new perspectives on the intricate effects of MPs on methane production, which could inform the optimization of AD processes in environments contaminated by MPs.

1. Introduction

The extensive use of fossil fuels, driven by industrial growth, contributes significantly to climate change and energy scarcity, as fossil fuel consumption accounts for over two-thirds of global greenhouse gas emissions (Risco-Bravo et al., 2024). This reliance on fossil fuels has heightened global concerns about ecological balance and environmental safety, prompting a shift towards renewable energy sources with lower greenhouse gas emissions. Anaerobic digestion (AD) is recognized as a promising waste-to-energy technology capable of producing methane from various waste sources, including wastewater, food waste, crop residues, forestry materials, and livestock manure (Li et al., 2019; Gao et al., 2022). Although biogas production reached approximately 35 million tons of oil equivalent in 2018, this remains a fraction of AD potential to meet an estimated 20% of current global gas demand (IEA, 2020). However, since AD involves a complex biochemical process in which multiple anaerobic microorganisms work together, it is inherently less stable and vulnerable to unfavorable factors leading to inefficiency (Gao et al., 2022). Key parameters such as feedstock characteristics, carbon to nitrogen ratio, temperature, pH, microbial community, and solid to liquid ratio have been shown to significantly impact AD efficiency (Kainthola et al., 2019; Kumar and Samadder, 2020). Organic wastes often contain a variety of exogenous pollutants, including heavy metal, plastic, and persistent organic pollutants, which pose ecotoxicity risks and can inhibit the AD process (Tou et al., 2017; Ajay et al., 2020; Luo et al., 2020). For instance, over 90% of plastics in domestic and industrial wastewater are retained in waste-activated sludge during treatment (Carr et al., 2016). Additionally, plastics commonly used in food packaging frequently contaminate food waste due to imperfect waste segregation systems (Li et al., 2022).

Plastics are widely used in various applications, leading to a large amount of plastic waste in the environment due to poor waste management systems. Researchers have classified plastics based on their origin and degradability into two main types: petroleum-based and bio-based, as well as non-degradable and degradable plastics (Shang et al., 2023). This study specifically focuses on petroleum-based plastics because their non-renewable and resistant nature poses serious threats to ecosystems. Plastic debris can break down through various processes, resulting in tiny particles called microplastic (MP), which are smaller than 5 mm (Akdogan and Guven, 2019). These MPs can be found in a wide range of organic wastes that are commonly used as substrates for AD. For instance, MPs generated in urban and residential areas often end

* Corresponding author.
E-mail address: Fuy49@cardiff.ac.uk (Y. Fu).

https://doi.org/10.1016/j.jenvman.2025.124627
Received 2 December 2024; Received in revised form 31 January 2025; Accepted 16 February 2025
Available online 23 February 2025

up in wastewater treatment systems alongside municipal and industrial waste, as well as urban runoff (Nizzetto et al., 2016). During wastewater treatment, over 90% of MPs are trapped in waste activated sludge (WAS) (Carr et al., 2016). Additionally, plastics used in food packaging frequently contaminate food waste due to improper waste segregation, which can lead to the generation of MPs during the recycling of food waste (Li et al., 2022). The presence of MPs in organic waste can disrupt methane production during AD, creating new challenges for system operations. Research indicated that the impact of MPs varies depending on their type, size, and concentration (Wei et al., 2019a; Chen et al., 2021; H. Chen et al., 2023; X. Wang et al., 2023). Furthermore, factors such as the type of feedstock, properties of the inoculum, and operational parameters can influence these effects. Exploring all these parameters involved in MPs effects through parallel experiments presents significant challenges. Therefore, prior to performing AD, it is crucial to determine the optimal conditions for maximum methane yield from substrates that contain MPs. However, to the best of our knowledge, the approach that can predict methane yield from MP-laden substrates while addressing the issues involved in identifying the optimal AD conditions, MPs properties, and feedstock properties to maximize methane yield has not yet been developed.

Machine Learning (ML) has gained momentum in environmental science and engineering, offering robust tools for complex regression and classification tasks by leveraging large, multidimensional datasets (Zhong et al., 2021). Within the field, ML techniques have been successfully used for MP monitoring, allowing automated counting and identification of MPs in environmental samples from images (Yurtsever and Yurtsever, 2019; Lorenzo-Navarro et al., 2021). ML models are also increasingly applied in waste-to-energy systems, including AD (Yildirim and Ozkaya, 2023; Y. Zhang et al., 2023b,c), pyrolysis (Zhu et al., 2019), hydrothermal carbonization (Li et al., 2021b), and gasification (Elmaz et al., 2020; Li et al., 2021a). For instance, recent work by Yildirim and Ozkaya (2023) demonstrated that Random Forest (RF) model achieved the highest accuracy in predicting biogas yield from operational data of a real-scale AD plant. However, an exhaustive literature review revealed that most of the ML work has focused on biogas prediction from the AD systems with different operational conditions, and few studies have been conducted to address the effects of MP contamination on biogas production. Considering MP contamination alongside other AD parameters could enhance our understanding of how these pollutants affect methane production, providing insights to improve AD performance across diverse plant operations. Additionally, the increasing volume of research on various MPs in AD systems supports ML-based analyses, which could clarify the complex influence of MP physicochemical properties and AD conditions on biogas output (Wei et al., 2019b; Chen et al., 2021).

To address the existing knowledge gap regarding methane production in AD systems impacted by MP contamination, we developed four ML models specifically tailored to predict methane yield. Through a comprehensive comparative analysis of these models, we identified the most accurate one. ML techniques offer distinct advantages over traditional statistical analyses by simultaneously accounting for multiple relevant factors while uncovering complex correlations. This capability allows for precise predictions of methane production, facilitating a better understanding of how MPs influence various AD systems. To enhance our analysis, we categorized the dataset into two groups: one without MP contamination and one with MP contamination. By applying ML approach to each condition separately, we quantitatively assessed the effects of MPs on key AD parameters, allowing for a comparative evaluation of their influence. Additionally, the insights derived from ML model can inform the design of AD systems, particularly in optimizing the use of MP-contaminated substrates through recommendations on inoculum-to-substrate ratios. In this study, we identified 18 input variables, which were categorized into four main areas: (i) MP characteristics; (ii) physicochemical properties of substrates; (iii) physicochemical properties of inoculums; and (iv) experimental conditions of AD. Prior to the development of ML models, we conducted descriptive statistical analyses on data from existing literature to establish correlations among these variables. This comprehensive approach enhances our understanding of how various factors contribute to methane production in the presence of MPs, ultimately supporting improved AD strategies in contaminated environments.

2. Materials and methods

2.1. Data collection and preprocessing

A comprehensive literature search was conducted to collect data on the methane production from AD exposed to different plastics from literature databases, including Web of Sciences, Scopus, and PubMed. The search terms included keywords related to AD and petroleum-based plastics, and then the content of the searched literature was examined to identify literature for data collection. The detailed literature search and screening process was described in Gao et al. (2024). To ensure generalizability, 18 attributes were defined as the input features during the data collection consisting of four empirical categories, (1) MP characteristics (type, particle size, and concentration), (2) substrate properties (total solid (TS), volatile solid (VS), pH, total chemical oxygen demand (TCOD), and soluble chemical oxygen demand (SCOD)), (3) inoculum properties (TS, VS, pH, TCOD, and SCOD), and (4) operating conditions for AD (feed to inoculum ratio (F/I), bioreactor volume, working volume, temperature, and digestion time), and methane yield was defined as the output target. The significance of these input features will be thoroughly examined through subsequent feature selection part. The collected publications for this study were expected to contain all the data on these four categories, and the publications were excluded because of lacking all data for the parameter under consideration. Consequently, 20 studies were selected (Table S1), providing 2878 methane yield data points for machine learning analysis. For comparably examine the impact of MPs on AD, the dataset was divided into two subsets: one without MP pollution (Table 1) and one with MP pollution (Table S2). Data were obtained directly from text and tables or were extracted manually from figures using the Web Plot Digitizer Software (https://apps.automeris.io/wpd4/).

Moreover, the units of plastic concentration and methane yield were converted to mg/L and mL/g of VS uniformly to avoid inconsistencies in the parameters for machine learning. Plastic concentrations were converted to the values in the substrate according to the method proposed by Leusch and Ziajahromi (2021), where the densities of the different plastics were polyethylene = 0.94 g/cm^3, polyethylene terephthalate = 1.38 g/cm^3, polyvinyl chloride = 1.4 g/cm^3, polystyrene = 1.06 g/cm^3, polycarbonate = 1.2 g/cm^3, polyester = 1.38 g/cm^3, and polyamide 6 = 1.14 g/cm^3. Since the data samples were compiled from a wide range of publications, inevitably inconsistencies exist leading to 6 features (TCOD, SCOD, and pH the substrate and inoculum, respectively) containing missing data points (22.2%, 22.1%, 13.8%, 13.8%, 12.2%, 10.6%) in the raw dataset. To solve this problem, the K Nearest Neighbour algorithm from the fancyimpute library 0.7.0 was implemented to the raw dataset for data imputation. In addition, the target encoder was applied to convert the categorical feature (plastic type) to the numerical feature to consist of the other features. For each category of the plastic type feature, target encoding replaces it with the mean of the target variable for that specific category. This method is particularly useful for categorical variables with a high cardinality, as it captures the relationship between the categorical feature and the target variable, while avoiding the pitfalls of one-hot encoding, such as increased dimensionality and potential sparsity (Micci-Barreca, 2001).

2.2. Model development and evaluation

To predict methane production from organic wastes with and without MP contamination, four supervised machine learning models

Table 1
Empirical categories and input features used to predict methane production from anaerobic digestion without microplastic pollution.

Empirical categories	Input features		Unit	Abbreviation	Data range	No. of datapoints
Substrate properties	1	TS	g/L	S_TS	15.3–249.2	730
	2	VS	g/L	S_VS	9.5–183.6	730
	3	pH	–	S_pH	4.7–7.4	654
	4	TCOD	g/L	S_TCOD	11.9–156.6	548
	5	SCOD	g/L	S_SCOD	0.1–31.1	630
Inoculum properties	6	TS	g/L	I_TS	15.6–163.5	730
	7	VS	g/L	I_VS	7.9–79	730
	8	pH	–	I_pH	6.4–8.1	615
	9	TCOD	g/L	I_TCOD	5.5–98	557
	10	SCOD	g/L	I_SCOD	0.2–7.5	615
Experimental conditions	11	F/I	–	F/I	0.4–14.5	730
	12	Bioreactor volume	mL	B_Vol	108–1000	730
	13	Working volume	mL	W_Vol	50–750	730
	14	Temperature	°C	Temp	35–37	730
	15	Digestion time	d	Time	1–125	730

Note: TS: total solid; VS: volatile solid; TCOD: total chemical oxygen demand; SCOD: soluble chemical oxygen demand; F/I: feed to inoculum ratio.

were selected: RF, Support Vector Regression (SVR), Gradient Boosting Regression (GBR), and eXtreme Gradient Boosting (XGB). These ML algorithms have been employed for modelling the complex relationships in AD process and have demonstrated accuracy in methane production prediction (Andrade Cruz et al., 2022). The models were trained and tested on a computer with Windows 10, using Python 3.8 and machine learning libraries including scikit-learn, XGBoost, and Pandas. Detailed descriptions of these selected ML models can be found in Text S1.

Before building each machine learning model, each of the two datasets (with and without MPs) was randomly split into a training dataset (80%) and a test dataset (20%). To ensure that each model received the same dataset, datasets were split only once during the feature selection phase and then used in turn to train the four different models (RF, SVR, GBR and XGB). Once these models have been built, their training and generalisation performance on the same dataset is evaluated using two metrics: the root mean square error (RMSE) and the coefficient of determination (R^2). The RMSE is a measure of the prediction error of the model and how well the model fits the observations (Eq. (1)), so the smaller the RMSE, the better the model performs. The R^2 (Eq. (2)) is a statistic that measures the superiority of the model over a simple average model (Y = X) and shows how well the model explains the variance of the predicted values. The R^2 ranges from 0 to 1, and the closer to 1, the better the model performs.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

where y_i is the real target value of the ith sample, $\hat{y}_i$ is the predicted target value, and $\bar{y}$ is the average of the actual value of the target in all samples (n).

2.3. ML-based feature selection

In regression problems, feature selection is a method of filtering out features that contribute little to the target result or add noise to the machine learning models. This can improve both the performance of models and the accuracy of predicted results (T. Zhang et al., 2020). ML-based feature importance and correlation analysis are essential for feature selection. Based on the performance of the selected four ML models, the best-performing model was then employed to perform feature selection. Specifically, Hierarchical Clustering (HC) and Pearson Correlation Coefficient (PCC) analysis were firstly applied to the complete dataset, including both "with plastic" and "without plastic" conditions. (Palansooriya et al., 2022). Features with high correlations, indicating they provided similar information, were grouped into one cluster. From each cluster, one representative feature was selected for ML model development. ML-based feature importance can effectively identify the representative features of a cluster. Then four machine learning models-RF, SVR, GBR, and XGB-were trained and tested on the two datasets to select the best-performing model. Feature importance rankings from best-performing model, along with the HC results, were used to remove features with low importance and low correlation to the target variable, retaining only the most relevant features for further analysis. By combining feature importance with correlation, the most crucial feature within each cluster was chosen as the representative feature. The subsequent removal of less contribution features served as a strategic optimization step, further enhancing the overall performance of the machine learning model.

2.4. ML performance analysis with updated datasets

After feature selection, two updated datasets (with and without MPs) were further applied to refine the initially optimized ML model. Retaining the same hyperparameters (n_estimators, learning rate, max_depth, min_samples_split, min_samples_leaf and subsample) allowed for direct comparison between the original and updated models, assessing the effectiveness of the feature selection method while minimizing time and computational costs (Palansooriya et al., 2022). Improved predictive performance in the updated model would validate the feature selection process, whereas any decline would indicate a need for hyperparameter retuning. The final model then provided insights into the significance and influence of each feature on the target variable. The detailed process of the ML model is provided in Text S1.

The detailed process of this study, depicted in Fig. 1, is divided into three main parts. In the first step, data was collected from the literatures according to input and output variables, and data preprocessing was carried out prior to model development. In the second phase, four ML models were constructed using the original datasets, and the best-performing model was selected for feature importance and clustering analysis. Moreover, the features contributed to the model were filtered based on feature importance and correlation, and the redundant data were removed, thus upgrading the datasets. Finally, the best predictive model for guiding practical applications was identified through AutoML based on the updated dataset.

3. Results and discussion

3.1. Statistical analysis of raw and reconstructed dataset

Following a comprehensive literature review and data collection, 18

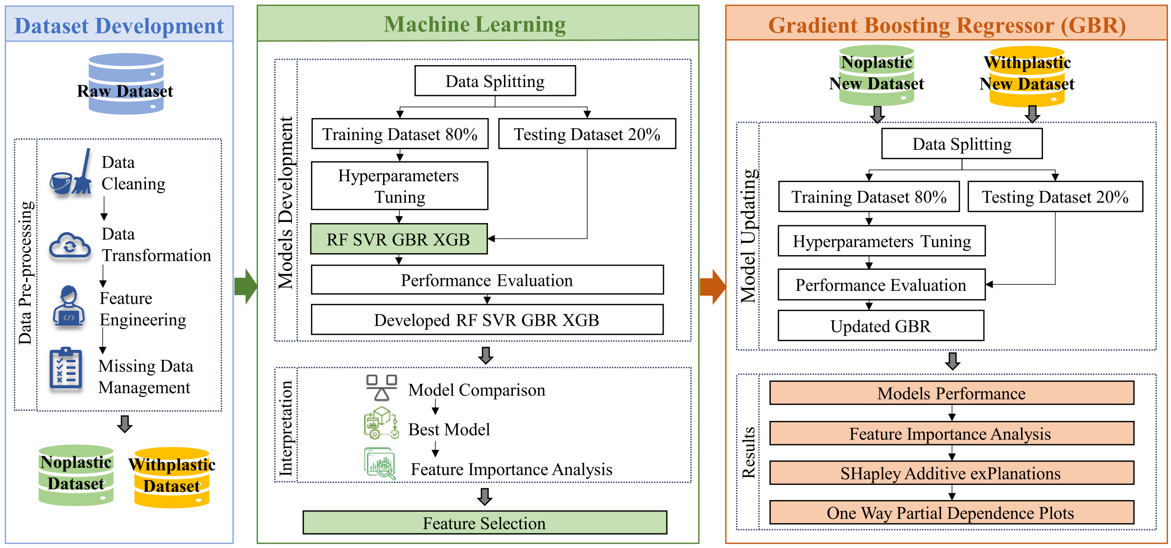

Fig. 1. Flowchart detailing the strategies of the machine learning framework to determine the effect of microplastic pollution on anaerobic digestion. RF: Random Forest; SVR: Support Vector Regression; GBR: Gradient Boosting Regressor; XGB: eXtreme Gradient Boosting.

input variables and 1 output variable were identified from 20 publications. These input variables were categorized into four empirical categories based on field knowledge, and methane yield was considered as the output variable. Detailed characteristics of the input variables for both conditions, with and without MP contamination, are presented in Table 1 and Table S2, respectively. A total of eight types of MPs were screened, which basically includes major types of petroleum-based plastics used in real life (Kwon et al., 2023). The range of digestion times in the collected dataset was extensive (1–125 d), representing the entire digestion process of substrates (Gao et al., 2022). pH plays a critical function in the operation of AD systems (Pang et al., 2023). However, since only a few of the screened publications mentioned adjusting the initial pH of the AD system to 7, and the data for this variable was missing from most of the publications, this variable was not included in the operating conditions for AD.

The statistical distribution of input features and output target for the raw data obtained from publications is shown in Fig. 2, only the final methane yield (specific methane yield) was considered. In terms of the plastic types employed in the studies, about one-third of the samples (28.4%) were PS and 15.5% focused on PVC (Fig. 2a). The PS particles were chosen as representative plastics in most of the studies due to their high global production and commercial availability; while PVC has received much attention as the most toxic plastic throughout the production and disposal process (Wei et al., 2019b). Moreover, the impact of different MP sizes on AD was variable (Fig. 2b). It was found that larger MP particles are more dispersible, resulting in increased exposure to sludge (Wang et al., 2022). Smaller MPs, on the other hand, can cause cytotoxicity by penetrating microbial cells (Zhang and Chen, 2020). The range of MP concentration was 0.012–126000 mg/L (Table S2), and the overall trend was that specific methane yield decreased with increasing plastic concentration (Fig. 2c). Fig. 2d shows the ratio of specific methane yield in the presence of different plastic types compared to the absence of plastic. The results indicated that PA6 and PC caused an increase in methane yield, but the presence of MPs generally decreased the methane yield. It has been demonstrated that different MP types can have opposite results on AD. For example, methane yield was reduced by up to 23% at PS concentration of 100 mg/L (Feng et al., 2018). In contrast, the presence of PA6 at 10 particles/g TS improved methane yield by 39.5%, from 124 to 173 mL/g of VS (Chen et al., 2021). These contrasting outcomes indicated that different MP types could affect the AD process variably due to their diverse physicochemical properties, which cannot be explained by similar mechanisms. Apart from plastic pollution, the properties of substrate and inoculum also influence the methane yield from AD. As shown in Fig. 2e, the physicochemical properties of the substrates were basically consistent, while those of the inoculum varied among the studies.

To further reveal the overall relationship of all variables, the k-nearest neighbors algorithm with $k = 3$ was used to fill in the missing values in features S_pH, I_pH, S_TCOD, I_TCOD, S_SCOD and I_SCOD, where the Euclidean distance was employed to identify the three nearest neighbors, and a weighted average of their values was used for imputation, thereby obtaining a complete dataset. The KNN imputation method was selected as it effectively preserves the local structure of the data and avoids assumptions about the underlying distribution (Salamattalab et al., 2024). The linear relationship between pairwise variables in the complete dataset was analyzed by PCC (Fig. S2). As shown in the figure, digestion time had a weak positive correlation with methane yield. Meanwhile, the TCOD and SCOD of the substrate also showed a weak positive correlation with methane yield. However, methane yield did not linearly correlate with MP type, particle size, and concentration. In addition, the statistical analysis described above did not allow for any general conclusions to be drawn or any judgements to be made in the AD processes. Therefore, the linear analyses did not well reflect the effects of the input variables on methane yield. Further internal correlations among these variables should be revealed based on big data and non-linear methods.

3.2. ML models development and feature analysis

Four ML algorithms, named RF, SVR, XGB, and GBR, were developed and assessed to predict the methane yield using 18 input features after filling in data gaps. The optimal hyper-parameters for each model were adjusted during the training step which aims to minimize the prediction error based on 5-fold cross-validation (Table S3). Fig. 3 presents the performance of each model, as measured by the R^2 and RMSE, as well as an actual values comparison of predicted values derived from the training and testing datasets. The testing R^2 values for GBR model were 0.996 without MP pollution and 0.998 with MP pollution, with corresponding RMSE values of 7.102 and 7.730. For the XGB and RF models, R^2 values remained similarly high across both datasets, approximately 0.993 and 0.996. In contrast, the SVR model exhibited the lowest predictive accuracy. These findings indicate that the GBR model outperformed the other three ML models in accurately predicting methane yield. These four ML models have been commonly used to predict

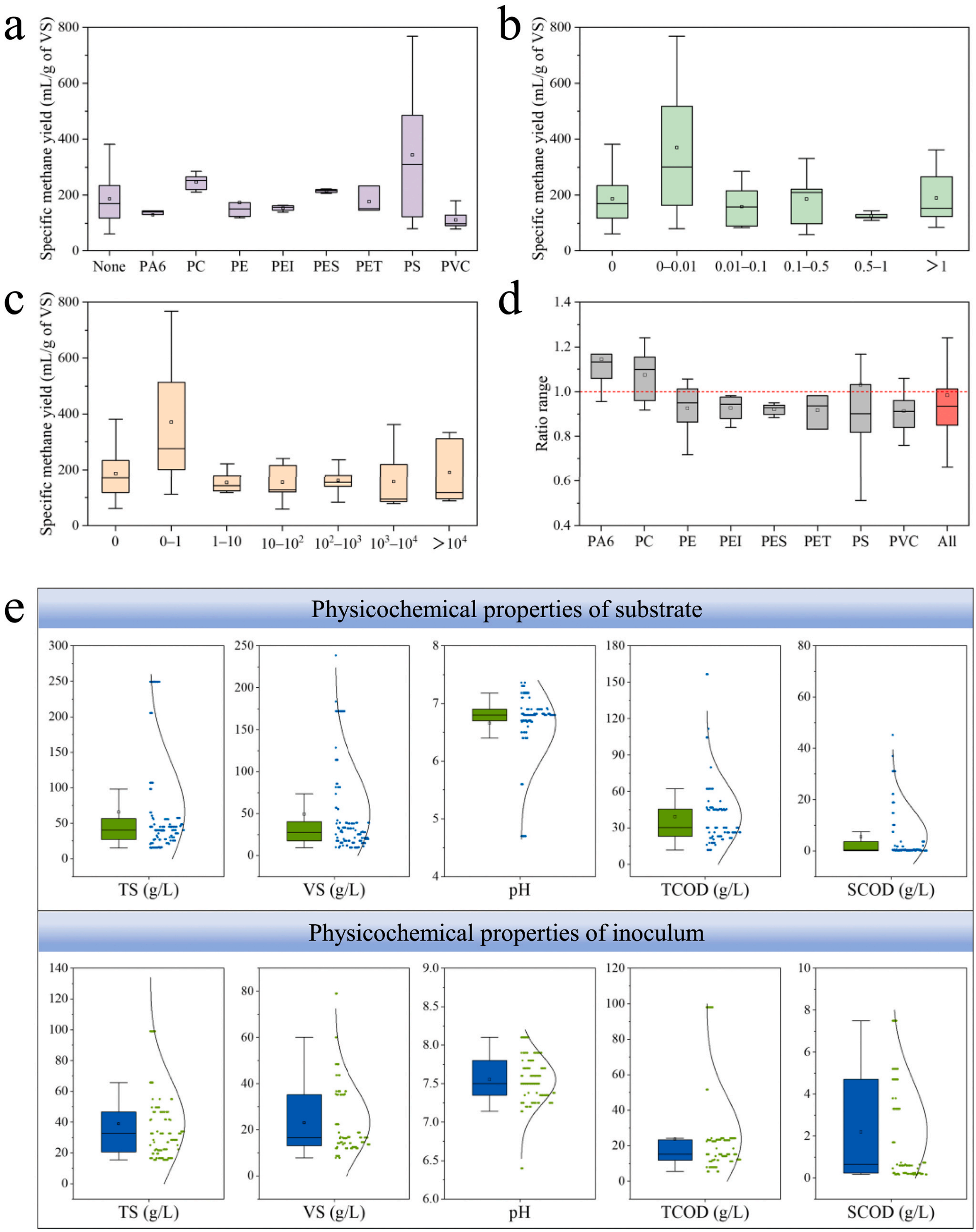

Fig. 2. Statistical data visualization of the impact of: (a) microplastic types; (b) microplastic size; and (c) microplastic concentration on specific methane yield; (d) Ratio of specific methane yield in the presence of different microplastic types compared to the absence of microplastic; (e) Box-normal plot representing the physicochemical properties of substrate and inoculum. PA6: Polyamide 6; PC: polycarbonate; PE: polyethylene; PEI: polyethyleneimine; PES: polyester; PET: polyethylene terephthalate; PS: polystyrene; PVC: polyvinyl chloride; TS: total solid; VS: volatile solid; TCOD: total chemical oxygen demand; SCOD: soluble chemical oxygen demand.

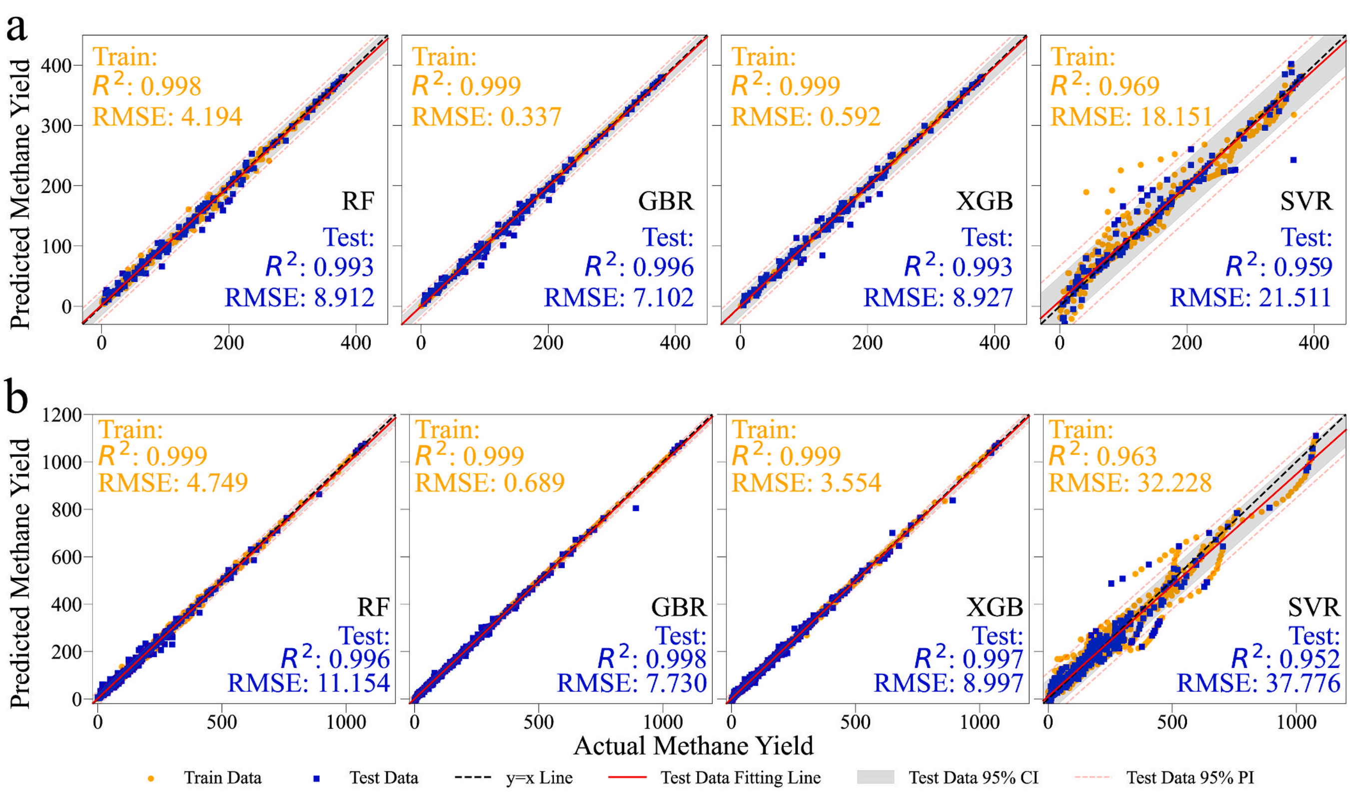

Fig. 3. Predictive performance of four machine learning models under two conditions: (a) without microplastic pollution and (b) with microplastic pollution. XGB: eXtreme Gradient Boosting; SVR: Support Vector Regression; RF: Random Forest; GBR: Gradient Boosting Regressor; RMSE: root-mean-square error.

methane production from AD. Zhang et al. employed three ML models (GBR, XGB, and RF) to reveal the role of biochar in AD, with GBR performing the best (2023a).

Although the prediction accuracy of ML models (especially GBR) was satisfactory with respect to the preliminary dataset after filling in the gaps, some input features were not important enough to build a model for methane yield prediction. The features which were highly correlated with each other may not offer unique discriminative information, and it can be easy to complicate the model by using too many of these variables, leading to generalization capability. To enhance prediction performance, ML-based feature engineering was applied to filter out less significant features and simplify the model. PCC (Fig. S2) and hierarchical clustering (Fig. 4a and c) were used to identify highly correlated input features, while the top-performing GBR model provided a ranking of feature importance (Fig. 4b and d). Based on domain knowledge, four empirical categories, including MP characteristics, substrate properties, inoculum properties, and operating conditions for AD, were identified to guide final feature selection. Notably, features within the same cluster but from different empirical categories were retained to preserve distinctive predictive information.

PCC analysis (Fig. S2) revealed a strong correlation between bioreactor volume and working volume. These two variables also clustered together and showed low feature importance values (Fig. 4). Consequently, bioreactor volume and working volume were deemed redundant for predicting methane yield. For inoculum properties, TS and VS content were strongly correlated, as were TCOD and SCOD. Additionally, TS and pH of inoculum were identified as more important features and belonged to separate clusters (Fig. 4). Finally, the VS, TCOD and SCOD content of inoculum were removed to establish a fresh dataset. Based on the feature importance values (Fig. 4b and d), the importance of the substrate properties was higher overall than that of the inoculum, except for the TS. It is reasonable that the substrate, as a raw material utilized by the microorganisms, had a higher organic matter content compared to the inoculum, providing a greater impact on the AD system (Fig. 2e). The key differences between Fig. 4b and d were the increased importance of inoculum TS and substrate pH in the presence of MPs. This is linked to the two effect pathways of MPs on AD: decreasing organic matter dissolution and increasing oxidative pressure (Gao et al., 2024). The high importance of inoculum TS can be explained by its influence on the solid-liquid ratio of the final system (Khadaroo et al., 2020), which influences the organic matter dissolution in substrate. Meanwhile, pH affects the oxidative pressure induced by MPs (H. Chen et al., 2023), further impacting methane production. The temperature can vary considerably across AD systems, which are classified into three categories based on temperature range and the activity of the microorganisms (Ryue et al., 2020). It was found that methane yield was higher in thermophilic digesters than in mesophilic digesters (Liu et al., 2022). However, the temperature in the dataset was only 35–37 °C, lacking data on continuous digestion processes in AD plants. Therefore, the feature importance of temperature was the least for model prediction. In general, the bioreactor volume, working volume, inoculum VS, inoculum TCOD, inoculum SCOD, substrate TS, substrate VS, and temperature were discarded from the dataset of input features, as they were either redundant features or had low importance to the methane yield prediction.

3.3. ML model update with filtered datasets

The top-performing GBR model was reconceptualized using streamlined datasets to enhance its generalization capability and increase computational efficiency. Fig. 5 shows the prediction accuracy of the updated GBR model. The testing R^2 values were 0.967 without MP pollution and 0.986 with MP pollution, with corresponding RMSE values of 18.645 and 20.747. Although the updated GBR model had a slightly lower R^2 value and a higher RMSE value than that of the preliminary GBR model, this slight reduction in prediction performance is reasonable. The excluded features may interact with other parameters in the original model and thus improve predictive performance (Lan et al.,

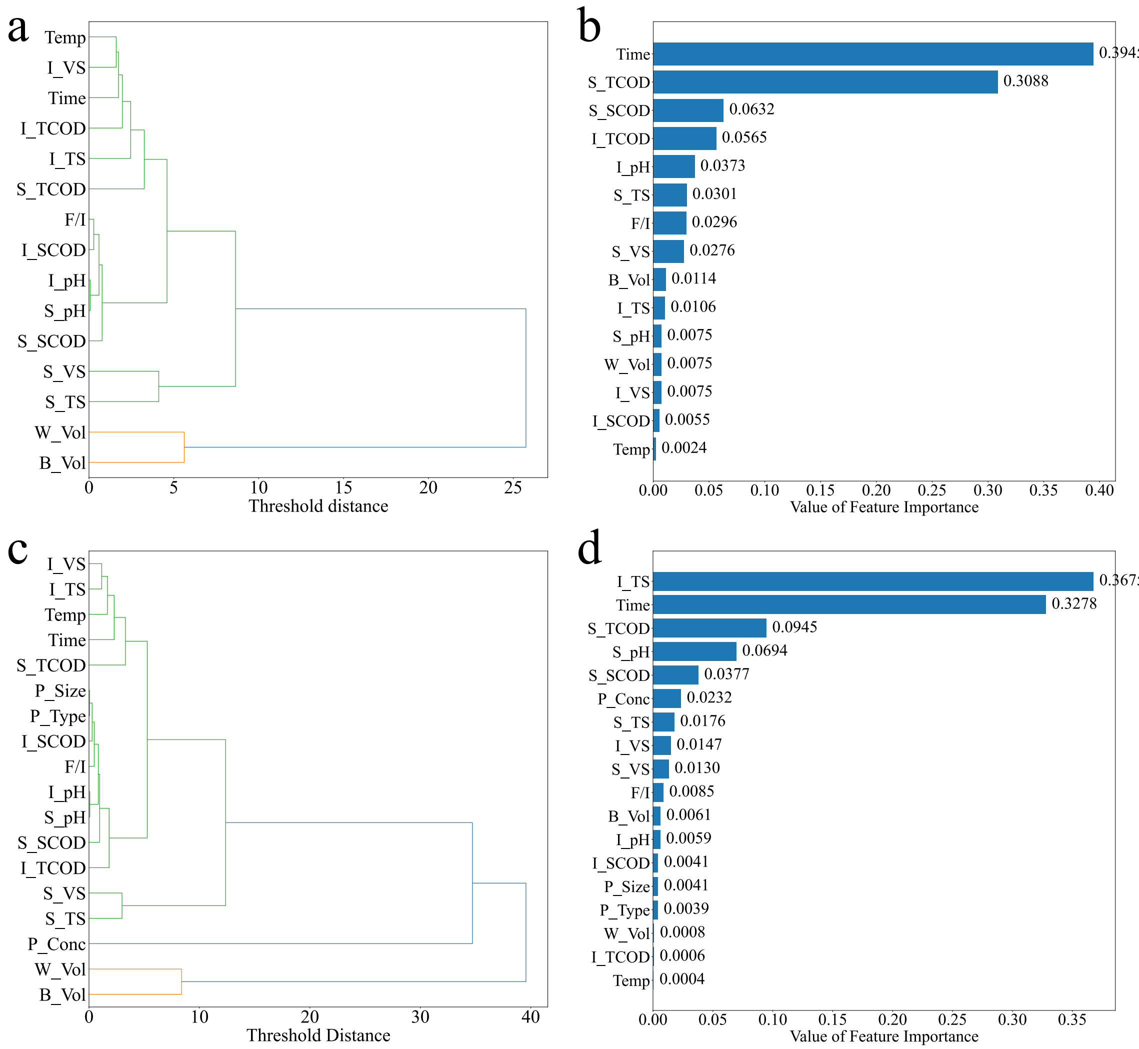

Fig. 4. Input feature analysis from the Gradient Boosting Regressor model under two conditions: (a–b) without microplastic pollution and (c–d) with microplastic pollution. Panels (a) and (c) present hierarchical clustering of features, while panels (b) and (d) display feature importance rankings. P_Type, P_Size, and P_Conc: the type, particle size, and concentration of microplastics; S_TS, S_VS, S_pH, S_TCOD, and S_SCOD: the total solid, volatile solid, pH, total chemical oxygen demand, and soluble chemical oxygen demand of substrate; I_TS, I_VS, I_pH, I_TCOD, and I_SCOD: the total solid, volatile solid, pH, total chemical oxygen demand, and soluble chemical oxygen demand of inoculum; F/I: feed to inoculum ratio; B_Vol: bioreactor volume; W_Vol: working volume; Temp: temperature; Time: digestion time.

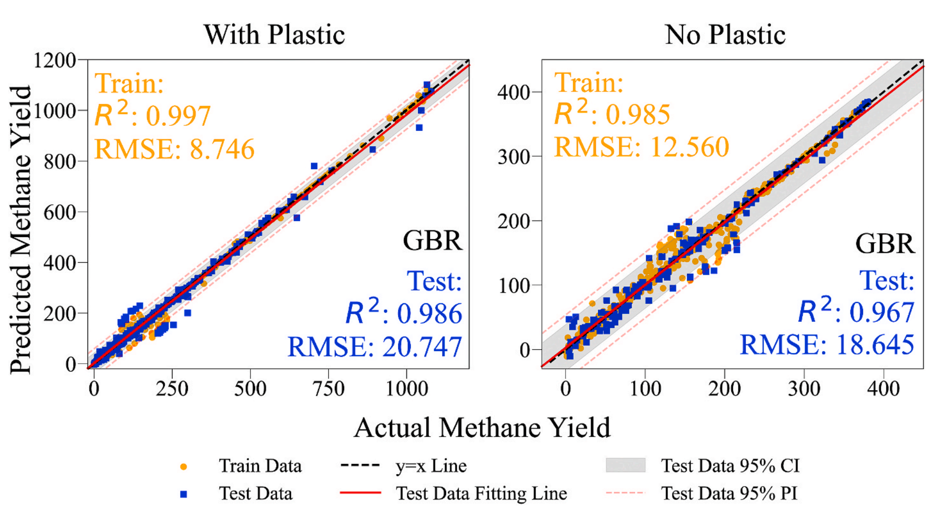

Fig. 5. Predictive performance of updated Gradient Boosting Regressor (GBR, the best performance model).

2023). Nonetheless, the employment of reduced parameters can effectively save computational cost and time in model development, and the decrease in predictive accuracy was minimal compared to the previous model. Additionally, excessive input features can lead to overfitting of the model, so the removal of redundant features from the original dataset diminishes the overfitting problem (Li et al., 2021a). Furthermore, the predictive performance of this paper was higher than the values previously reported in the literature (Baek et al., 2023; Y. Zhang et al., 2023b,c; Ghazizade Fard and Koupaie, 2024). These results implied that the feature filtering approach applied in this study, in which only 10 essential features were identified, can achieve a desirable ML model performance.

The updated GBR model was further utilized to explore the importance and impact of input features. The feature importance and their impact or contribution on methane yield were plotted according to each data point, with the importance represented by the mean absolute SHAP value (Fig. 6). The importance rankings varied considerably between datasets with and without MP pollution. In the absence of MPs, TCOD and SCOD of substrate ranked highly (Fig. 6a), reflecting the direct influence of organic matter on methane production (Soo et al., 2022). Furthermore, pH is crucial for methanogen activity in AD system, as these microorganisms thrive in neutral conditions around pH 7.2, with their growth inhibited at lower pH levels (Gao et al., 2022; Yellezuome et al., 2022). Conversely, the importance of TCOD and SCOD decreased in the presence of MPs (Fig. 6c). This decline is likely due to the interference of MPs with organic matter solubilization, which hinders methane production despite higher SCOD levels (Gao et al., 2024). The inoculum TS content directly influences the solid-liquid ratio in AD system. Increased solid loading can restrict microbial access to MPs, thereby diminishing their inhibitory effects. Additionally, pH affects plastic properties by influencing the degree of hydrolysis and degradation, potentially altering their physical and chemical characteristics. Lower pH levels may enhance the breakdown of certain MPs, leading to changes in their mechanical strength and stability (Tiwari et al., 2020). The SHAP value method not only ranks feature importance but also elucidates how input features affect methane yield (Fig. 6c and d). The input variable has a positive effect on methane yield if the SHAP value increases as the value of the input variable increases, which corresponds to the transition from blue to red in the graph. Notably, the overall SHAP value for methane yield increased proportionally with digestion time,

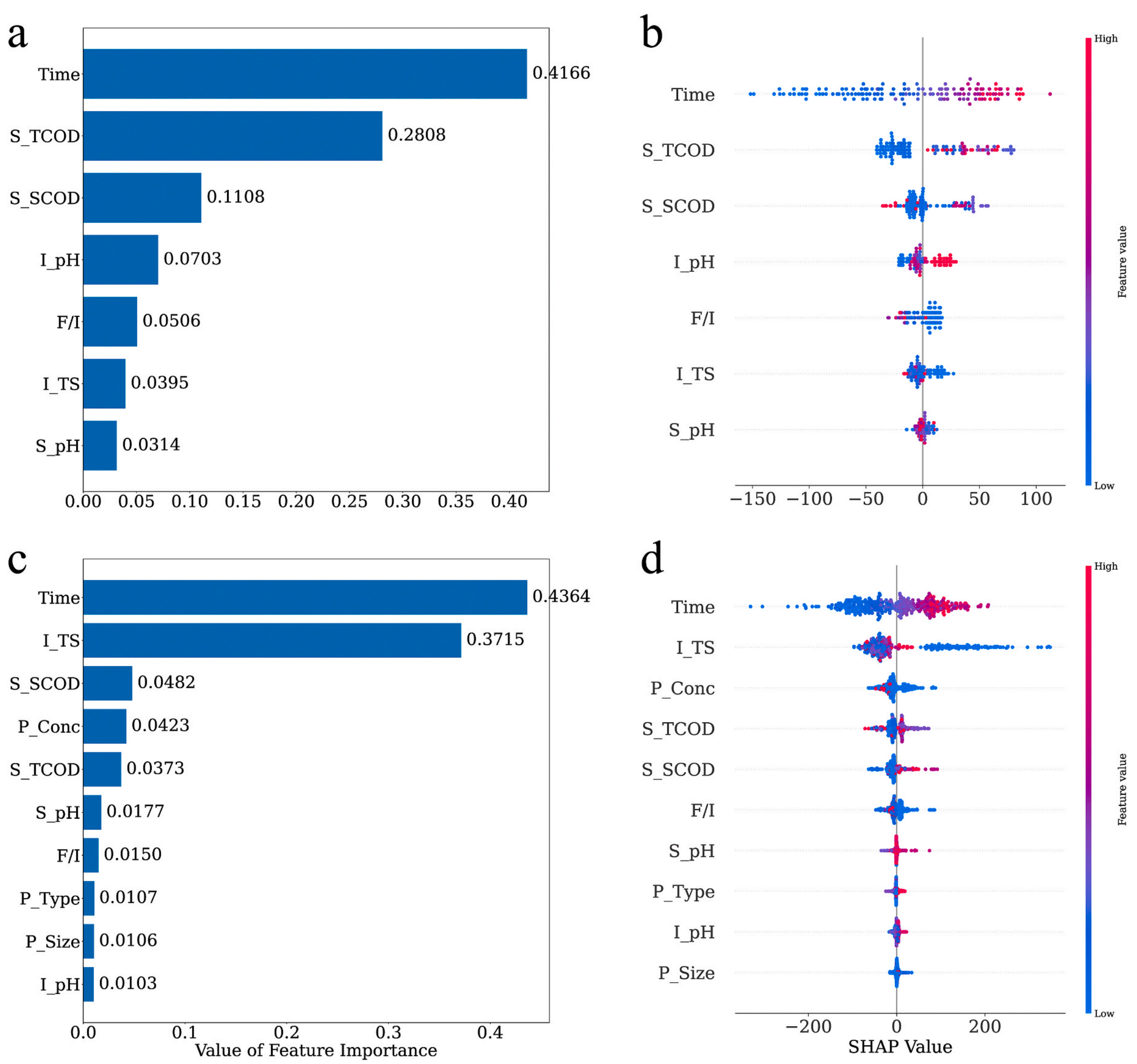

Fig. 6. Feature analysis of updated Gradient Boosting Regressor (GBR, the best performance model) using a feature-filtered dataset under two conditions: (a–b) without microplastic pollution and (c–d) with microplastic pollution. Panels (a) and (c) illustrate feature importance analysis, while panels (b) and (d) display SHAP value analysis. P_Type, P_Size, and P_Conc: the type, particle size, and concentration of microplastics; S_pH, S_TCOD, and S_SCOD: pH, total chemical oxygen demand, and soluble chemical oxygen demand of substrate; I_TS and I_pH: the total solid and pH of inoculum; F/I: feed to inoculum ratio; Time: digestion time.

suggesting that longer digestion periods enhance biodegradability and consequently methane production (Gao et al., 2022). However, the substrate consumption and the hydrolysis products (volatile fatty acid and ammonia nitrogen) accumulation during AD could prevent the production of methane or even contribute to the cessation of the system (Wu et al., 2017). As depicted in Fig. 6b and d, the impact of digestion time on methane yield diminished at higher values. Furthermore, both TCOD and SCOD of the substrate demonstrated a positive relationship with methane yield. A high COD value indicates a significant presence of organic matter in AD system, while SCOD is readily utilized by microorganisms for methane production (X. Zhang et al., 2023a).

3.4. Optimal process design based on updated model

In addition to analyzing feature importance, partial dependence plots (PDP) were performed to obtain the contribution of key features in predicting methane production. Fig. 7 illustrates how each of 10 input features affects model performance, with data distributions represented by blue and red columns along the x-axis. pH had a strong effect on methane yield with an optimal level around 6.4, while either highly acidic or alkaline conditions were detrimental (Fig. 7a). Substrate TCOD and SCOD levels were also influential, with positive correlations to methane yield (Fig. 7b and c). Ma et al. (2018) found that COD had a positive contribution to methane production. However, excessive COD can impose organic loading stress, disrupting the AD process (Braz et al., 2019). When COD levels exceed the tolerance limits of methanogens, the balance between hydrolysis/acidogenesis and methanogenesis stages is disrupted, reducing methane yield. High COD concentrations, primarily from soluble proteins and polysaccharides, are hydrolyzed into smaller molecules, which accumulate as volatile fatty acids by acidification. This volatile fatty acids buildup further inhibits methanogen activity, hindering methane production (Choong et al., 2016; Ajayi-Banji and Rahman, 2022). In fact, methane production decreased notably above 50 g/L TCOD in Fig. 7b. However, in the presence of MPs, the methane yield remained stable even as substrate SCOD levels rose above 20 g/L. This stabilization likely reflects the reduced solubility of organic matter due to MPs, which in turn limits VFA enrichment. As shown in Fig. 7e, the PDP value decreased sharply when the initial TS content increased from 15.6 g/L to 19.5 g/L, then slowed until around 40 g/L, where it began a gradual rise as TS content increased further. Initially, higher TS content of inoculum caused an increase in bioreactor TS, which prolonged the startup period and reduced the stability of the system, and this effect tended to weaken as the TS continued to increase (M. Wang et al., 2023; Yan et al., 2022). This eventual slow rise in methane production may be due to the contribution of organic matter within the inoculum itself, which provides a slight additional methane yield (Muaaz-Us-Salam et al., 2020).

Xie et al. (2021) investigated the impact of varying F/I on anaerobic digestibility, discovering that methane production decreased as F/I increased, with acetoclastic methanogens facing significant inhibition at a maximum F/I ratio of 9. Fig. 7f shows that the PDP value initially declined sharply with rising F/I but stabilized around a ratio of 5.

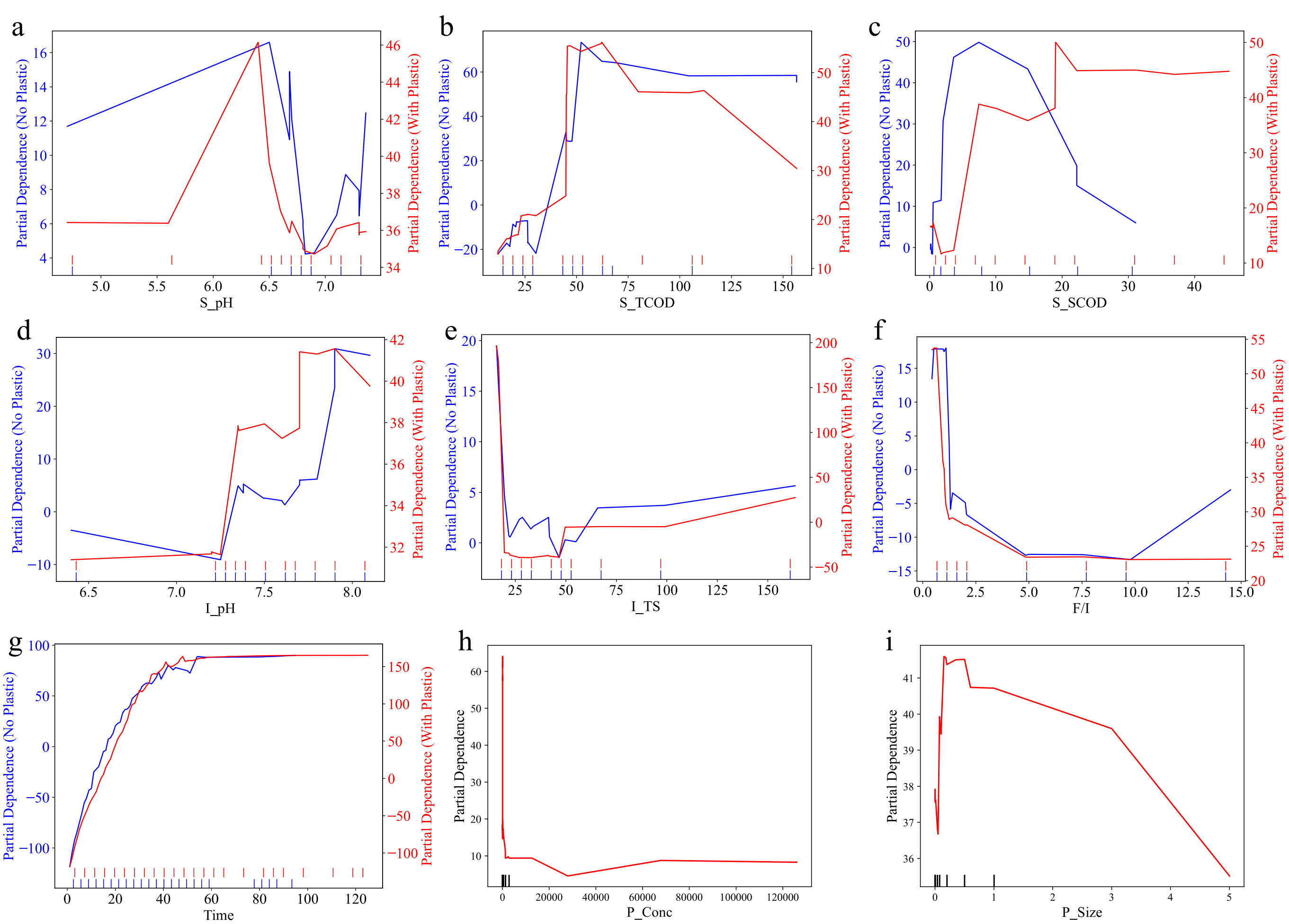

Fig. 7. Updated Gradient Boosting Regressor (GBR, the best performance model) interpretation by one-way Partial Dependence Plot (PDP). S_pH, S_TCOD, and S_SCOD: pH, total chemical oxygen demand, and soluble chemical oxygen demand of substrate; I_pH and I_TS: pH and total solid of inoculum; F/I: feed to inoculum ratio; Time: digestion time. P_Size and P_Conc: particle size and concentration of microplastics.

Digestion time emerged as the most critical factor affecting methane production. The PDP values increased rapidly at first but gradually slowed down, plateauing at around 40 days (Fig. 7g). This was consistent with the operating conditions of the practical AD system, which was generally around 35 days for a single operation in a batch AD reactor, and a 30-day hydraulic retention time for continuous AD (Muaaz-Us-Salam et al., 2020; X. Chen et al., 2023). Fig. 7h and i depict the effect of MP particle concentration and size on methane yield, respectively. The PDP value declined sharply with increasing MP concentration, indicating a significant inhibitory impact on methane yield. Notably, the effect of MPs stabilized after reaching approximately 20 g/L. Conversely, as MP particle size increased, the PDP value initially rose sharply before gradually declining, peaking at around 0.5 mm. Particle size plays a decisive role in MP toxicity; smaller sizes have been reported to cause greater cellular damage (Das et al., 2021). As particle size increased further, the inhibition of methane by the dosage effect was demonstrated. Generally, a higher AD performance can be achieved with MP particle sizes of approximately 0.5 mm and concentrations below 20 g/L.

3.5. Implications and limitations of the current study

The optimum conditions for methane production in AD containing MPs varied considerably among studies. It is challenging to examine all relevant parameters via simultaneous experiments. In this study, an ML-based empirical method was developed to overcome this challenge, which can be used to predict the methane production from the AD system with MPs pollution based on the properties of MP, substrate, and inoculum, as well as the experimental conditions. The updated GBR model has been designed as a potential technique to accurately predict methane production without performing any experiments, which could be helpful in practical AD plants. Additionally, ML-based interpretation analyses compared conditions with and without plastic contamination, highlighting the critical parameters for optimizing methane production in MP-contaminated AD systems. The dependence relationships of methane production on the operational parameters were revealed by the PDP analysis, which provided an insight into the relative importance of each variable for methane production and identified the most important variables. Meanwhile, the interaction between the parameters needs to be further analyzed carefully to more accurately guide actual operations (Haider Jaffari et al., 2023). Overall, this study can facilitate a comprehensive view of the effects of MPs in AD process and achieve maximum resource recovery through AD using MP polluted organics.

The results of this study also have several limitations due to the quality and quantity of data collected from publications. Although many publications have focused on the effects of MPs on anaerobic digestion so far, there is no available data for all the 18 variables selected for the ML dataset. For those articles that provided all the required data, the data distribution for certain input features and the output target was inconsistent because of a variety of variations in experimental goals, methodologies, and conditions. For example, the MP content in the environment was often measured by counting, i.e. the number of MP particles identified in a specific mass or volume of the environment (Murphy et al., 2016). Therefore, the collected data for MP concentration contained particles/L and the common unit (mg/L) (Feng et al., 2018; Wei et al., 2019a; J. Zhang et al., 2020). In addition, microbes play an important role in the AD processes, and MPs have been found to interfere with functional microbes during AD (Li et al., 2022). However, it is difficult to extract accurate data from stacked histograms of relative microbial content, and microbial community fluctuates greatly among different experimental conditions. Hence, future research should focus on the construction of a comprehensive database that includes studies under uniform experimental conditions and similar experimental methodologies. Furthermore, in the current dataset, there were only three types of MP features (including MP type, particle size, and concentration) that could be extracted from the publications. With the development and availability of microscopic characterization tools, more features of MPs involved in AD will be measured and discussed. In particular, the surface functional groups of MPs are directly associated with their impact on AD. The application of surface chemistry data obtained from X-ray photoelectron spectroscopy, three-dimension excitation emission matrix fluorescence spectroscopy, and Fourier transform infrared spectroscopy for developing ML models can provide a better understanding on the effects of MPs in AD process.

4. Conclusions

This study developed and compared four ML models (RF, SVR, XGB, and GBR) to predict methane yield in AD systems with and without MP pollution. To enhance the computational performance and reduce the impact of redundant features, the dataset was refined using feature importance and cluster analysis, leading to an optimized model. The following conclusions can be drawn:

a) Among the four ML models, demonstrated the highest accuracy in predicting methane yield. This superior performance underscores GBR's potential for effectively modelling complex AD systems and evaluating the impacts of MPs pollution.
b) The study identified MP concentration as a key predictor of methane yield, outweighing the influence of MP type and particle size. Additionally, feature importance results indicated that in the presence of MPs, substrate pH and SCOD, along with inoculum TS, are critical parameters and should be closely regulated in practical AD operations.
c) PDP analysis provided insights into optimal conditions for methane production, emphasizing digestion time, pH, F/I, and MP particle size and concentration. These insights provide valuable guidance for optimizing operational conditions in AD systems to maximize methane yield.

This work identifies key predictors and optimal operational conditions, offering valuable insights to improve AD efficiency and highlighting the significant impact of MPs on AD systems. These findings lay the groundwork for future research and the practical application of ML techniques to optimize AD plant operations and enhance sustainability.

CRediT authorship contribution statement

Zhenghui Gao: Writing – original draft, Investigation, Data curation, Conceptualization. **Zongqiang Ren:** Writing – review & editing, Supervision. **Tianyi Cui:** Writing – review & editing, Investigation. **Yao Fu:** Writing – review & editing, Visualization, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge financial support from the China Scholarship Council (CSC) for Zhenghui Gao (CSC No. 202006760088) and Tianyi Cui (CSC No. 202008500151).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2025.124627.

Data availability

Data will be made available on request.

References

Ajay, C.M., Mohan, S., Dinesha, P., Rosen, M.A., 2020. Review of impact of nanoparticle additives on anaerobic digestion and methane generation. Fuel 277, 118234. https://doi.org/10.1016/j.fuel.2020.118234.

Ajayi-Banji, A., Rahman, S., 2022. A review of process parameters influence in solid-state anaerobic digestion: focus on performance stability thresholds. Renew. Sustain. Energy Rev. 167, 112756. https://doi.org/10.1016/j.rser.2022.112756.

Akdogan, Z., Guven, B., 2019. Microplastics in the environment: a critical review of current understanding and identification of future research needs. Environ. Pollut. 254, 113011. https://doi.org/10.1016/j.envpol.2019.113011.

Andrade Cruz, I., Chuenchart, W., Long, F., Surendra, K.C., Renata Santos Andrade, L., Bilal, M., Liu, H., Tavares Figueiredo, R., Khanal, S.K., Fernando Romanholo Ferreira, L., 2022. Application of machine learning in anaerobic digestion: perspectives and challenges. Bioresour. Technol. 345, 126433. https://doi.org/10.1016/j.biortech.2021.126433.

Baek, G., Lee, C., Yoon, J., 2023. Machine learning approach for predicting anaerobic digestion performance and stability in direct interspecies electron transfer-stimulated environments. Biochem. Eng. J. 193, 108840. https://doi.org/10.1016/j.bej.2023.108840.

Braz, G.H.R., Fernandez-Gonzalez, N., Lema, J.M., Carballa, M., 2019. Organic overloading affects the microbial interactions during anaerobic digestion in sewage sludge reactors. Chemosphere 222, 323–332. https://doi.org/10.1016/j.chemosphere.2019.01.124.

Carr, S.A., Liu, J., Tesoro, A.G., 2016. Transport and fate of microplastic particles in wastewater treatment plants. Water Res. 91, 174–182. https://doi.org/10.1016/j.watres.2016.01.002.

Chen, H., Tang, M., Yang, X., Tsang, Y.F., Wu, Y., Wang, D., Zhou, Y., 2021. Polyamide 6 microplastics facilitate methane production during anaerobic digestion of waste activated sludge. Chem. Eng. J. 408, 127251. https://doi.org/10.1016/j.cej.2020.127251.

Chen, H., Zou, Z., Tang, M., Yang, X., Tsang, Y.F., 2023a. Polycarbonate microplastics induce oxidative stress in anaerobic digestion of waste activated sludge by leaching bisphenol A. J. Hazard Mater. 443, 130158. https://doi.org/10.1016/j.jhazmat.2022.130158.

Chen, X., Liu, W., Zhao, Y., He, H., Ma, J., Cui, Z., Yuan, X., 2023b. Optimization of semi-continuous dry anaerobic digestion process and biogas yield of dry yellow corn straw: based on "gradient anaerobic digestion reactor.". Bioresour. Technol. 389, 129804. https://doi.org/10.1016/j.biortech.2023.129804.

Choong, Y.Y., Norli, I., Abdullah, A.Z., Yhaya, M.F., 2016. Impacts of trace element supplementation on the performance of anaerobic digestion process: a critical review. Bioresour. Technol. 209, 369–379. https://doi.org/10.1016/j.biortech.2016.03.028.

Das, R.K., Sanyal, D., Kumar, P., Pulicharla, R., Brar, S.K., 2021. Science-society-policy interface for microplastic and nanoplastic: environmental and biomedical aspects. Environ. Pollut. 290, 117985. https://doi.org/10.1016/j.envpol.2021.117985.

Elmaz, F., Yücel, Ö., Mutlu, A.Y., 2020. Predictive modeling of biomass gasification with machine learning-based regression methods. Energy 191, 116541. https://doi.org/10.1016/j.energy.2019.116541.

Feng, Y., Feng, L.-J., Liu, S.-C., Duan, J.-L., Zhang, Y.-B., Li, S.-C., Sun, X.-D., Wang, S.-G., Yuan, X.-Z., 2018. Emerging investigator series: inhibition and recovery of anaerobic granular sludge performance in response to short-term polystyrene nanoparticle exposure. Environ. Sci.: Water Res. Technol. 4, 1902–1911. https://doi.org/10.1039/C8EW00535D.

Gao, Z., Alshehri, K., Li, Y., Qian, H., Sapsford, D., Cleall, P., Harbottle, M., 2022. Advances in biological techniques for sustainable lignocellulosic waste utilization in biogas production. Renew. Sustain. Energy Rev. 170, 112995. https://doi.org/10.1016/j.rser.2022.112995.

Gao, Z., Qian, H., Cui, T., Ren, Z., Wang, X., 2024. Comprehensive *meta*-analysis reveals the impact of non-biodegradable plastic pollution on methane production in anaerobic digestion. Chem. Eng. J., 149703 https://doi.org/10.1016/j.cej.2024.149703.

Ghazizade Fard, M., Koupaie, E.H., 2024. Machine learning assisted modelling of anaerobic digestion of waste activated sludge coupled with hydrothermal pre-treatment. Bioresour. Technol. 394, 130255. https://doi.org/10.1016/j.biortech.2023.130255.

Haider Jaffari, Z., Jeong, H., Shin, J., Kwak, J., Son, C., Lee, Y.-G., Kim, S., Chon, K., Hwa Cho, K., 2023. Machine-learning-based prediction and optimization of emerging contaminants' adsorption capacity on biochar materials. Chem. Eng. J. 466, 143073. https://doi.org/10.1016/j.cej.2023.143073.

IEA, 2020. Outlook for biogas and biomethane: prospects for organic growth – analysis [WWW Document]. IEA. URL. https://www.iea.org/reports/outlook-for-biogas-and-biomethane-prospects-for-organic-growth, 1.11.24.

Kainthola, J., Kalamdhad, A.S., Goud, V.V., 2019. A review on enhanced biogas production from anaerobic digestion of lignocellulosic biomass by different enhancement techniques. Process Biochem. 84, 81–90. https://doi.org/10.1016/j.procbio.2019.05.023.

Khadaroo, S.N.B.A., Grassia, P., Gouwanda, D., Poh, P.E., 2020. The influence of different solid-liquid ratios on the thermophilic anaerobic digestion performance of palm oil mill effluent (POME). J. Environ. Manag. 257, 109996. https://doi.org/10.1016/j.jenvman.2019.109996.

Kumar, A., Samadder, S.R., 2020. Performance evaluation of anaerobic digestion technology for energy recovery from organic fraction of municipal solid waste: a review. Energy 197, 117253. https://doi.org/10.1016/j.energy.2020.117253.

Kwon, G., Cho, D.-W., Park, J., Bhatnagar, A., Song, H., 2023. A review of plastic pollution and their treatment technology: a circular economy platform by thermochemical pathway. Chem. Eng. J. 464, 142771. https://doi.org/10.1016/j.cej.2023.142771.

Lan, H., Hou, H., Cynthia, Gou, Z., 2023. A machine learning led investigation to understand individual difference and the human-environment interactive effect on classroom thermal comfort. Build. Environ. 236, 110259. https://doi.org/10.1016/j.buildenv.2023.110259.

Leusch, F.D.L., Ziajahromi, S., 2021. Converting mg/L to particles/L: reconciling the occurrence and toxicity literature on microplastics. Environ. Sci. Technol. 55, 11470–11472. https://doi.org/10.1021/acs.est.1c04093.

Li, J., Pan, L., Suvarna, M., Wang, X., 2021a. Machine learning aided supercritical water gasification for H2-rich syngas production with process optimization and catalyst screening. Chem. Eng. J. 426, 131285. https://doi.org/10.1016/j.cej.2021.131285.

Li, J., Zhu, X., Li, Y., Tong, Y.W., Ok, Y.S., Wang, X., 2021b. Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: application of machine learning on waste-to-resource. J. Clean. Prod. 278, 123928. https://doi.org/10.1016/j.jclepro.2020.123928.

Li, Y., Chen, Y., Wu, J., 2019. Enhancement of methane production in anaerobic digestion process: a review. Appl. Energy 240, 120–137. https://doi.org/10.1016/j.apenergy.2019.01.243.

Li, Y., Li, X., Wang, P., Su, Y., Xie, B., 2022. Size-dependent effects of polystyrene microplastics on anaerobic digestion performance of food waste: focusing on oxidative stress, microbial community, key metabolic functions. J. Hazard Mater. 438, 129493. https://doi.org/10.1016/j.jhazmat.2022.129493.

Liu, Y., Wang, T., Xing, Z., Ma, Y., Nan, F., Pan, L., Chen, J., 2022. Anaerobic co-digestion of Chinese cabbage waste and cow manure at mesophilic and thermophilic temperatures: digestion performance, microbial community, and biogas slurry fertility. Bioresour. Technol. 363, 127976. https://doi.org/10.1016/j.biortech.2022.127976.

Lorenzo-Navarro, J., Castrillón-Santana, M., Sánchez-Nielsen, E., Zarco, B., Herrera, A., Martínez, I., Gómez, M., 2021. Deep learning approach for automatic microplastics counting and classification. Sci. Total Environ. 765, 142728. https://doi.org/10.1016/j.scitotenv.2020.142728.

Luo, J., Zhang, Q., Zhao, J., Wu, Y., Wu, L., Li, H., Tang, M., Sun, Y., Guo, W., Feng, Q., Cao, J., Wang, D., 2020. Potential influences of exogenous pollutants occurred in waste activated sludge on anaerobic digestion: a review. J. Hazard Mater. 383, 121176. https://doi.org/10.1016/j.jhazmat.2019.121176.

Ma, Y., Gu, J., Liu, Y., 2018. Evaluation of anaerobic digestion of food waste and waste activated sludge: soluble COD versus its chemical composition. Sci. Total Environ. 643, 21–27. https://doi.org/10.1016/j.scitotenv.2018.06.187.

Micci-Barreca, D., 2001. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. SIGKDD Explor. Newsl. 3, 27–32. https://doi.org/10.1145/507533.507538.

Muaaz-Us-Salam, S., Cleall, P.J., Harbottle, M.J., 2020. Application of enzymatic and bacterial biodelignification systems for enhanced breakdown of model lignocellulosic wastes. Sci. Total Environ. 728, 138741. https://doi.org/10.1016/j.scitotenv.2020.138741.

Murphy, F., Ewins, C., Carbonnier, F., Quinn, B., 2016. Wastewater treatment works (WwTW) as a source of microplastics in the aquatic environment. Environ. Sci. Technol. 50, 5800–5808. https://doi.org/10.1021/acs.est.5b05416.

Nizzetto, L., Futter, M., Langaas, S., 2016. Are agricultural soils dumps for microplastics of urban origin? Environ. Sci. Technol. 50, 10777–10779. https://doi.org/10.1021/acs.est.6b04140.

Palansooriya, K.N., Li, J., Dissanayake, P.D., Suvarna, M., Li, L., Yuan, X., Sarkar, B., Tsang, D.C.W., Rinklebe, J., Wang, X., Ok, Y.S., 2022. Prediction of soil heavy metal immobilization by biochar using machine learning. Environ. Sci. Technol. 56, 4187–4198. https://doi.org/10.1021/acs.est.1c08302.

Pang, H., Xu, Y., Ren, R., He, J., Pan, X., Wang, L., 2023. Enhanced anaerobic digestion of waste activated sludge by alkaline protease-catalyzing hydrolysis: role and significance of initial pH adjustment. Chem. Eng. J. 467, 143323. https://doi.org/10.1016/j.cej.2023.143323.

Risco-Bravo, A., Varela, C., Bartels, J., Zondervan, E., 2024. From green hydrogen to electricity: a review on recent advances, challenges, and opportunities on power-to-hydrogen-to-power systems. Renew. Sustain. Energy Rev. 189, 113930. https://doi.org/10.1016/j.rser.2023.113930.

Ryue, J., Lin, L., Kakar, F.L., Elbeshbishy, E., Al-Mamun, A., Dhar, B.R., 2020. A critical review of conventional and emerging methods for improving process stability in thermophilic anaerobic digestion. Energy for Sustainable Development 54, 72–84. https://doi.org/10.1016/j.esd.2019.11.001.

Salamattalab, M.M., Zonoozi, M.H., Molavi-Arabshahi, M., 2024. Innovative approach for predicting biogas production from large-scale anaerobic digester using long-short term memory (LSTM) coupled with genetic algorithm (GA). Waste Manag. 175, 30–41. https://doi.org/10.1016/j.wasman.2023.12.046.

Shang, Z., Wang, R., Zhang, X., Tu, Y., Sheng, C., Yuan, H., Wen, L., Li, Y., Zhang, J., Wang, X., Yang, G., Feng, Y., Ren, G., 2023. Differential effects of petroleum-based and bio-based microplastics on anaerobic digestion: a review. Sci. Total Environ. 875, 162674. https://doi.org/10.1016/j.scitotenv.2023.162674.

Soo, P.L., Bashir, M.J.K., Wong, L.-P., 2022. Recent advancements in the treatment of palm oil mill effluent (POME) using anaerobic biofilm reactors: challenges and future perspectives. J. Environ. Manag. 320, 115750. https://doi.org/10.1016/j.jenvman.2022.115750.

Tiwari, N., Santhiya, D., Sharma, J.G., 2020. Microbial remediation of micro-nano plastics: current knowledge and future trends. Environ. Pollut. 265, 115044. https://doi.org/10.1016/j.envpol.2020.115044.

Tou, F., Yang, Y., Feng, J., Niu, Z., Pan, H., Qin, Y., Guo, X., Meng, X., Liu, M., Hochella, M.F., 2017. Environmental risk implications of metals in sludges from waste water treatment plants: the discovery of vast stores of metal-containing nanoparticles. Environ. Sci. Technol. 51, 4831–4840. https://doi.org/10.1021/acs.est.6b05931.

Wang, C., Wei, W., Zhang, Y.-T., Dai, X., Ni, B.-J., 2022. Different sizes of polystyrene microplastics induced distinct microbial responses of anaerobic granular sludge. Water Res. 220, 118607. https://doi.org/10.1016/j.watres.2022.118607.

Wang, M., Wang, Y., Peng, J., Wang, L., Yang, J., Kou, X., Chai, B., Gao, L., Han, X., 2023a. A comparative study on Mesophilic and thermophilic anaerobic digestion of different total solid content sludges produced in a long sludge-retention-time system. Results in Engineering 19, 101228. https://doi.org/10.1016/j.rineng.2023.101228.

Wang, X., Zhang, Y., Zhao, Y., Zhang, L., Zhang, X., 2023b. Inhibition of aged microplastics and leachates on methane production from anaerobic digestion of sludge and identification of key components. J. Hazard Mater. 446, 130717. https://doi.org/10.1016/j.jhazmat.2022.130717.

Wei, W., Huang, Q.-S., Sun, J., Dai, X., Ni, B.-J., 2019a. Revealing the mechanisms of polyethylene microplastics affecting anaerobic digestion of waste activated sludge. Environ. Sci. Technol. 53, 9604–9613. https://doi.org/10.1021/acs.est.9b02971.

Wei, W., Huang, Q.-S., Sun, J., Wang, J.-Y., Wu, S.-L., Ni, B.-J., 2019b. Polyvinyl chloride microplastics affect methane production from the anaerobic digestion of waste activated sludge through leaching toxic bisphenol-A. Environ. Sci. Technol. 53, 2509–2517. https://doi.org/10.1021/acs.est.8b07069.

Wu, D., Lü, F., Shao, L., He, P., 2017. Effect of cycle digestion time and solid-liquid separation on digestate recirculated one-stage dry anaerobic digestion: use of intact polar lipid analysis for microbes monitoring to enhance process evaluation. Renew. Energy 103, 38–48. https://doi.org/10.1016/j.renene.2016.11.016.

Xie, A., Deaver, J.A., Miller, E., Popat, S.C., 2021. Effect of feed-to-inoculum ratio on anaerobic digestibility of high-fat content animal rendering wastewater. Biochem. Eng. J. 176, 108215. https://doi.org/10.1016/j.bej.2021.108215.

Yan, J., Zhao, Yehua, He, H., Cai, Y., Zhao, Yubin, Wang, H., Zhu, W., Yuan, X., Cui, Z., 2022. Anaerobic co-digestion of dairy manure and maize stover with different total solids content: from the characteristics of digestion to economic evaluation. J. Environ. Chem. Eng. 10, 107602. https://doi.org/10.1016/j.jece.2022.107602.

Yellezuome, D., Zhu, X., Wang, Z., Liu, R., 2022. Mitigation of ammonia inhibition in anaerobic digestion of nitrogen-rich substrates for biogas production by ammonia stripping: a review. Renew. Sustain. Energy Rev. 157, 112043. https://doi.org/10.1016/j.rser.2021.112043.

Yildirim, O., Ozkaya, B., 2023. Prediction of biogas production of industrial scale anaerobic digestion plant by machine learning algorithms. Chemosphere 335, 138976. https://doi.org/10.1016/j.chemosphere.2023.138976.

Yurtsever, M., Yurtsever, U., 2019. Use of a convolutional neural network for the classification of microbeads in urban wastewater. Chemosphere 216, 271–280. https://doi.org/10.1016/j.chemosphere.2018.10.084.

Zhang, J., Zhao, M., Li, C., Miao, H., Huang, Z., Dai, X., Ruan, W., 2020a. Evaluation the impact of polystyrene micro and nanoplastics on the methane generation by anaerobic digestion. Ecotoxicol. Environ. Saf. 205, 111095. https://doi.org/10.1016/j.ecoenv.2020.111095.

Zhang, T., Zhu, T., Xiong, P., Huo, H., Tari, Z., Zhou, W., 2020b. Correlated differential privacy: feature selection in machine learning. IEEE Trans. Ind. Inf. 16, 2115–2124. https://doi.org/10.1109/TII.2019.2936825.

Zhang, X., Yang, H., Wei, D., Chen, Z., Wang, Q., Song, Y., Ma, Y., Zhang, H., 2023a. Tolerance of anaerobic digestion sludge to heavy metals: COD removal, biogas production, and microbial variation. J. Water Proc. Eng. 55, 104157. https://doi.org/10.1016/j.jwpe.2023.104157.

Zhang, Y., Feng, Y., Ren, Z., Zuo, R., Zhang, T., Li, Y., Wang, Y., Liu, Z., Sun, Z., Han, Y., Feng, L., Aghbashlo, M., Tabatabaei, M., Pan, J., 2023b. Tree-based machine learning model for visualizing complex relationships between biochar properties and anaerobic digestion. Bioresour. Technol. 374, 128746. https://doi.org/10.1016/j.biortech.2023.128746.

Zhang, Y., Jing, Z., Feng, Y., Chen, S., Li, Y., Han, Y., Feng, L., Pan, J., Mazarji, M., Zhou, H., Wang, X., Xu, C., 2023c. Using automated machine learning techniques to explore key factors in anaerobic digestion: at the environmental factor, microorganisms and system levels. Chem. Eng. J. 475, 146069. https://doi.org/10.1016/j.cej.2023.146069.

Zhang, Z., Chen, Y., 2020. Effects of microplastics on wastewater and sewage sludge treatment and their removal: a review. Chem. Eng. J. 382, 122955. https://doi.org/10.1016/j.cej.2019.122955.

Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021. Machine learning: new ideas and tools in environmental science and engineering. Environ. Sci. Technol. 55, 12741–12754. https://doi.org/10.1021/acs.est.1c01339.

Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527. https://doi.org/10.1016/j.biortech.2019.121527.