World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Machine learning–based methane yield prediction using a structured anaerobic digestion dataset

Asit Chatterjee [1,*], Mahim Mathur [1], Anil Pal [2] and Mukesh Kumar Gupta [3]

[1] *Department of Civil Engineering, Suresh Gyan Vihar University, Jaipur, India.*
[2] *Department of Computer Application, Suresh Gyan Vihar University, Jaipur, India.*
[3] *Department of Electrical Engineering, Suresh Gyan Vihar University, Jaipur, India.*

World Journal of Advanced Engineering Technology and Sciences, 2025, 17(03), 112–120

Publication history: Received 30 October 2025; revised on 06 December 2025; accepted on 09 December 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.17.3.1546

Abstract

Accurate prediction of methane yield is essential for optimizing anaerobic digestion (AD) systems and improving the efficiency of agricultural biomass-to-energy conversion. This study presents a machine learning–based predictive framework trained on a structured and experimentally derived dataset encompassing physicochemical feedstock properties, operational parameters, and biogas performance indicators. The dataset includes more than 500 labeled samples representing major agricultural residues, characterized by Total Solids (TS), Volatile Solids (VS), C/N ratio, lignocellulosic composition, temperature, pH, and Organic Loading Rate. Six supervised learning algorithms like Gradient Boosting Regressor (GBR), Light GBM, Cat Boost, Extra Trees, K-Nearest Neighbors (KNN), and Elastic Net were developed and evaluated using an 80/20 train–test split, five-fold cross-validation, and performance metrics including RMSE, MAE, and R^2. Results indicate that Light GBM achieved the highest predictive accuracy with an R^2 of 0.95 and the lowest RMSE, demonstrating the dataset's strong feature representation and model suitability. Feature importance analysis revealed Volatile Solids, lignin content, and C/N ratio as the most influential predictors of methane yield. The findings confirm that machine learning models, when trained on well-structured AD datasets, can significantly enhance methane yield estimation and support intelligent, data-driven biogas plant optimization. This study establishes a scalable framework for predictive AD modeling and offers a foundation for integrating AI-driven decision-making into sustainable waste-to-energy systems.

Keywords: Anaerobic Digestion; Methane Yield Prediction; Machine Learning; Agricultural Biomass; Biogas Optimization

1. Introduction

Anaerobic digestion (AD) has emerged as a key bioenergy pathway for transforming agricultural residues into renewable methane-rich biogas. With increasing global emphasis on sustainable waste management, circular bioeconomy practices, and energy security, optimizing methane yield from biomass has become a critical research focus [1]. Agricultural residues such as rice straw, wheat straw, maize stover, sugarcane bagasse, and paddy husk represent abundant, low-cost feedstocks with significant potential for biomethane generation. However, methane yield varies widely depending on physicochemical feedstock characteristics, operational conditions, reactor dynamics, and microbial activity. [2] This inherent variability poses a major challenge for designing efficient AD systems and achieving consistent biogas output.

Traditional empirical and biochemical models struggle to fully capture the nonlinear and multivariate interactions that govern methane production [3]. Although multiple studies have attempted to correlate methane yield with parameters

* Corresponding author: Asit Chatterjee

such as lignocellulosic composition, C/N ratio, temperature, and Organic Loading Rate (OLR), these models often fall short due to limited dataset size, lack of structured metadata, and the inability to handle complex patterns within AD processes [4]. As a result, methane yield prediction remains uncertain, thereby limiting accurate reactor design, feedstock optimization, and intelligent control strategies in practical biogas plants.

Recent advancements in machine learning (ML) have demonstrated strong capabilities in modeling complex environmental and biochemical systems with high predictive accuracy. ML techniques can effectively learn relationships among multiple interacting parameters and estimate methane yield more reliably than traditional methods [5]. However, the success of ML models depends heavily on the availability of structured, high-quality datasets containing comprehensive physicochemical and operational features. In anaerobic digestion research, such datasets have traditionally been fragmented, incomplete, or inconsistent, restricting the development of robust predictive models.

To address this gap, the present study applies multiple supervised ML algorithms to a structured anaerobic digestion dataset comprising more than 500 experimentally derived and fully labeled samples. This dataset integrates feedstock composition, operational conditions, and methane performance indicators, enabling a robust platform for predictive modeling. The objective of this work is to evaluate the predictive strength of various ML models including Gradient Boosting Regressor, Light GBM, Cat Boost, Extra Trees, KNN, and Elastic Net—for methane yield estimation and to identify the most influential features contributing to biomethane production. The findings of this study not only highlight the superior performance of tree-based ensemble models but also provide insights into the key factors influencing methane generation. By establishing a reliable and scalable ML-based framework, this research contributes to data-driven optimization of AD processes and paves the way for intelligent, AI-enabled biogas plant control systems.

2. Literature Review

Research on AD has long emphasized the influence of feedstock characteristics and operational conditions on methane yield. Studies consistently highlight that parameters such as lignocellulosic composition, C/N ratio, volatile solids content, temperature regime, and organic loading rate strongly affect microbial activity and biogas production potential [6,7]. While conventional biochemical models and regression-based approaches have attempted to formulate predictive relationships among these parameters, their ability to capture the complex, nonlinear dynamics of AD systems remains limited [8]. This shortcoming is particularly evident when dealing with heterogeneous agricultural residues, where variability in chemical structure, moisture content, and degradability leads to significant fluctuations in methane yield.

In recent years, ML has emerged as a powerful alternative for modeling environmental and biochemical processes. ML algorithms excel at detecting hidden patterns, managing multi-dimensional inputs, and learning nonlinear interactions without requiring explicit mechanistic equations [9-11]. Applications of ML in energy systems, waste treatment, and environmental prediction have demonstrated superior performance compared to traditional modeling methods. Within the AD domain, early ML-based studies have explored the use of regression trees, artificial neural networks, and support vector machines to estimate methane yield or assess process stability. These models have shown promise, but their performance has often been constrained by small datasets, inconsistent data quality, and incomplete feature representation [12].

A recurring challenge identified across prior research is the absence of large, structured, and standardized datasets for anaerobic digestion modeling. Existing datasets typically lack comprehensive metadata describing feedstock composition, operational parameters, and methane performance indices [13]. This fragmentation limits the ability of machine learning models to generalize effectively, hampers reproducibility, and restricts cross-study comparisons. Without well-defined labeling frameworks and uniform measurement protocols, training robust predictive models becomes difficult, leading to inconsistent methane yield predictions and reduced reliability.

The emergence of ensemble learning techniques and gradient boosting algorithms has further expanded possibilities for accurate methane yield forecasting [14]. Such models are capable of handling feature heterogeneity, ranking variable importance, and offering interpretable insights into the biochemical factors controlling methane production [15]. Although these advancements present clear advantages, their integration into AD research remains underexplored due to the scarcity of well-structured datasets that support meaningful training and validation.

Given these gaps, there is a strong need for methane prediction studies that utilize comprehensive, experimentally validated datasets and modern machine learning techniques. The present work addresses this need by applying advanced supervised learning models to a structured dataset containing detailed physicochemical and operational descriptors of major agricultural residues. By benchmarking multiple ML algorithms and analyzing key predictive

features, this study provides a more reliable and scalable computational framework for methane yield estimation and supports the broader goal of optimizing anaerobic digestion systems through data-driven intelligence.

3. Methodology

The methodology adopted in this study integrates systematic dataset preparation, feature engineering, and machine learning model development to accurately predict methane yield from agricultural residues. The workflow consists of six main stages: data acquisition, preprocessing, feature engineering, model training, model validation, and performance evaluation as illustrated in figure 1

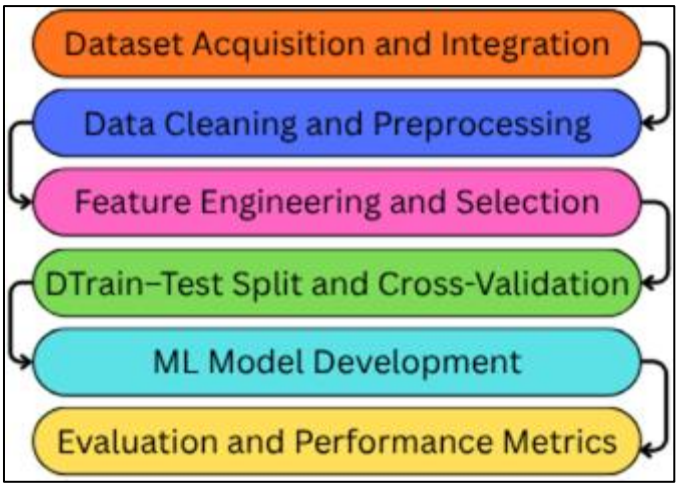

Figure 1 Proposed Machine Learning Framework for Methane Yield Prediction

3.1. Dataset Acquisition and Integration

A structured AD dataset developed in the prior phase of this research served as the foundational input for model development. The dataset comprises more than 500 labeled records representing major agricultural residues including rice straw, wheat straw, maize stover, sugarcane bagasse, and paddy husk. Each sample was experimentally characterized for physicochemical properties such as Total Solids (TS), Volatile Solids (VS), C/N ratio, lignin, cellulose, and hemicellulose, while operational parameters such as temperature, pH, Organic Loading Rate (OLR), and Hydraulic Retention Time (HRT) were recorded during batch digestion experiments. Biogas production and methane content were measured using standardized BMP assays, and methane yield (mL CH_4/g VS) was taken as the target variable.

3.2. Data Cleaning and Preprocessing

To ensure the dataset was model-ready, a multi-step preprocessing pipeline was applied. Missing values were imputed using feature-specific statistical strategies, while outliers were removed through the Interquartile Range (IQR) method. All continuous variables were normalized using Min–Max scaling to maintain uniform ranges across features. Categorical parameters such as temperature regime (mesophilic/thermophilic) and feedstock type were encoded using one-hot encoding. The preprocessed dataset was further checked for class balance, data consistency, and noise reduction.

3.3. Feature Engineering and Selection

Advanced feature engineering was performed to enhance model interpretability and predictive capability. Derived indices such as VS/TS ratio, lignocellulosic index, temperature stability index, and pH deviation factor were computed to better represent AD process behavior. Feature correlation analysis using Pearson's r and Mutual Information (MI) scores was conducted to identify the most influential predictors. Additionally, Principal Component Analysis (PCA) was used to assess feature redundancy and data separability without compromising interpretability. The final feature set included 15 physicochemical and operational parameters contributing significantly to methane yield prediction.

3.4. Train–Test Split and Cross-Validation

The dataset was randomly divided into an 80:20 train–test split to ensure unbiased evaluation. To further strengthen model reliability and reduce variance, five-fold cross-validation was applied during training. This approach ensured that each model was trained and validated across multiple subsets, improving generalization and preventing overfitting.

3.5. Machine Learning Model Development

A series of six supervised machine learning models were developed to accurately predict methane yield from the structured anaerobic digestion dataset. The selected algorithms included Gradient Boosting Regressor (GBR), Light Gradient Boosting Machine (Light GBM), Cat Boost Regressor, Extra Trees Regressor, KNN, and Elastic Net Regression, representing a diverse range of ensemble-based, instance-based, and regularized linear modeling approaches. To ensure optimal performance, each model underwent hyperparameter tuning using grid search, which systematically optimized parameters such as learning rate, tree depth, number of estimators, regularization coefficients, and nearest-neighbor settings. All models were trained on a standardized feature matrix derived from physicochemical and operational variables, with methane yield serving as the target output. This comprehensive model development process ensured robust learning of nonlinear relationships within the dataset and facilitated reliable comparative evaluation across different ML techniques.

3.6. Model Evaluation and Performance Metrics

Model performance was evaluated using three widely accepted regression metrics: Root Mean Squared Error (RMSE) to assess overall prediction accuracy, Mean Absolute Error (MAE) to quantify the average magnitude of deviation between predicted and actual values, and the Coefficient of Determination (R^2) to measure the predictive strength and explanatory capability of each model. In addition to numerical evaluation, predicted vs. Actual plots and boxplots were generated to visually examine model behavior, detect bias, and assess error distribution. Feature importance analysis was also performed to identify the most influential variables governing methane yield, enabling deeper insight into the underlying biochemical and operational drivers. Across all models, LightGBM demonstrated the highest predictive accuracy and consistency, confirming the suitability and robustness of tree-based ensemble algorithms for methane yield estimation using structured AD datasets.

4. Results and Discussion

The Results and Discussion section presents a detailed analysis of the dataset characteristics, feature behavior, machine learning model performance, and methane yield prediction outcomes. The findings highlight the complex interplay between physicochemical biomass properties, operational digestion parameters, and their collective influence on methane production. Statistical summaries reveal substantial variability among agricultural residues, emphasizing the importance of using advanced modeling techniques capable of capturing nonlinear relationships. Through a series of visualization plots, correlation analyses, and comparative model evaluations, this study demonstrates the effectiveness of machine learning—particularly gradient boosting–based algorithms in achieving highly accurate methane yield prediction. The following subsections discuss the dataset trends, feature contributions, and model behaviors in greater detail, supported by tables and figures that illustrate key insights from the predictive framework.

Table 1 Statistical Summary of Input Features in the Dataset

Feature	Min	Max	Mean	Std. Dev
TS (%)	42.3	68.4	51.7	6.3
VS (%)	38.9	61.0	46.5	5.8
C/N Ratio	17.8	32.9	27.1	3.9
Lignin (%)	9.8	27.3	17.4	4.9
Cellulose (%)	25.4	41.8	33.7	4.1
Temperature (°C)	30	55	41.2	7.8
OLR (g VS/L/day)	1.0	6.0	3.4	1.1
pH	6.2	8.1	7.18	0.23

The statistical summary of the key physicochemical and operational features used as input variables for methane yield prediction is given in table 1. The values highlight the inherent variability among agricultural residues, which directly influences their biodegradability and methane generation potential. Total Solids (TS) and Volatile Solids (VS) show moderate variability, with TS ranging from 42.3% to 68.4% and VS from 38.9% to 61.0%, indicating substantial differences in moisture content and organic matter availability across feedstocks. The C/N ratio, which plays a crucial role in microbial metabolism, also exhibits noticeable variation, spanning from 17.8 to 32.9; this range reflects both nitrogen-rich and carbon-dominant substrates, potentially affecting digestion stability. Lignin and cellulose contents show wide dispersion, with lignin varying between 9.8% and 27.3% and cellulose between 25.4% and 41.8%, underscoring the structural heterogeneity of biomass and its influence on anaerobic hydrolysis rates.

Operational parameters also reveal diverse experimental conditions. Temperature ranges from 30°C to 55°C, covering both mesophilic and thermophilic regimes, while OLR values between 1.0 and 6.0 g VS/L/day represent different loading intensities that can impact reactor performance and microbial activity. The pH values show relatively narrow variation, remaining within the optimal range (6.2–8.1) for methanogenic activity, indicating that digestion conditions were well controlled. Overall, the variability observed across these features highlights the complexity of anaerobic digestion systems and justifies the need for machine learning models capable of capturing nonlinear and multi-dimensional interactions to accurately predict methane yield.

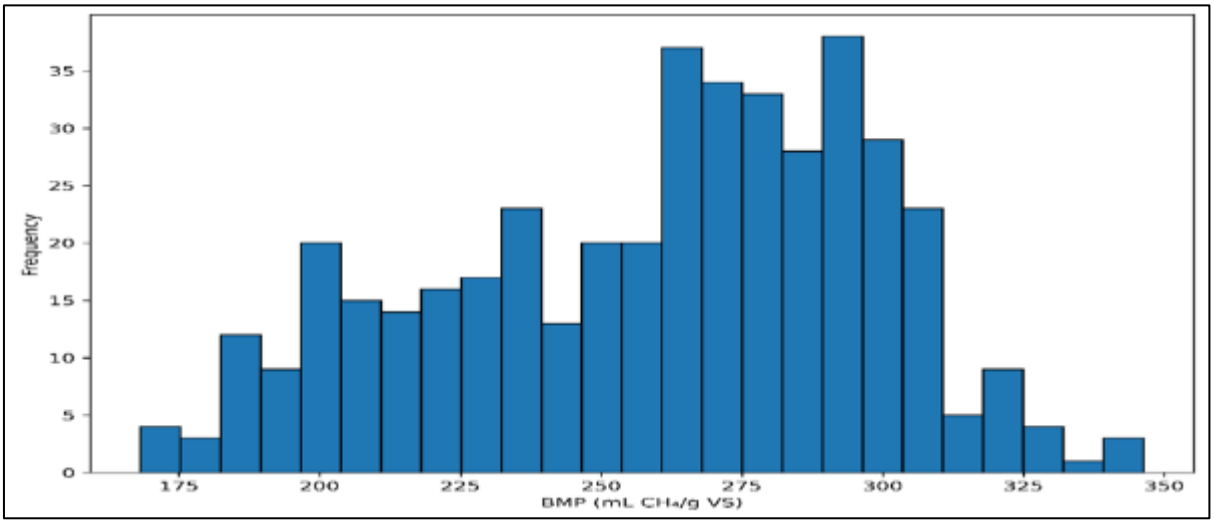

Figure 2 Distribution of Methane yield (BMP)

The distribution of methane yield across all samples in the dataset, providing a comprehensive overview of the variability in Biochemical Methane Potential (BMP) among different agricultural residues. The distribution shows a broad range of BMP values, indicating significant differences in biodegradability and organic matter conversion efficiency across feedstocks is shown in figure 2. The histogram reveals a slightly right-skewed pattern, suggesting that while a majority of samples produce moderate methane yields, a smaller number achieve exceptionally high methane output. This variation reflects intrinsic feedstock factors such as lignocellulosic composition, C/N ratio, and volatile solids content, as well as operational influences including temperature and OLR during digestion. The presence of multiple peaks within the distribution also suggests that the dataset contains distinct clusters corresponding to different biomass categories, each with characteristic biochemical properties. Overall, the distribution confirms the heterogeneity of the input samples and underscores the need for advanced predictive models capable of handling wide-ranging methane yield patterns.

Figure 3 Correlation heatmap of digestion features

The correlation heatmap of the physicochemical and operational features used for methane yield prediction. The heatmap provides clear insights into how different variables interact and influence the anaerobic digestion process is illustrated in figure 3. Notably, Volatile Solids (VS) and C/N ratio exhibit strong positive correlations with methane yield, highlighting their role as primary contributors to improved digestion efficiency. Lignin content shows a pronounced negative correlation, which is expected due to lignin's complex and recalcitrant structure that limits microbial accessibility and slows hydrolysis. Moderate correlations are observed for cellulose, temperature, and OLR, indicating their supportive but less dominant contributions to methane production. The heatmap also reveals inter-feature relationships, such as the inverse association between lignin and cellulose and the mild positive relationship between temperature and pH stability. These correlations validate existing biochemical principles of anaerobic digestion and demonstrate that the dataset captures meaningful patterns necessary for robust machine learning modeling.

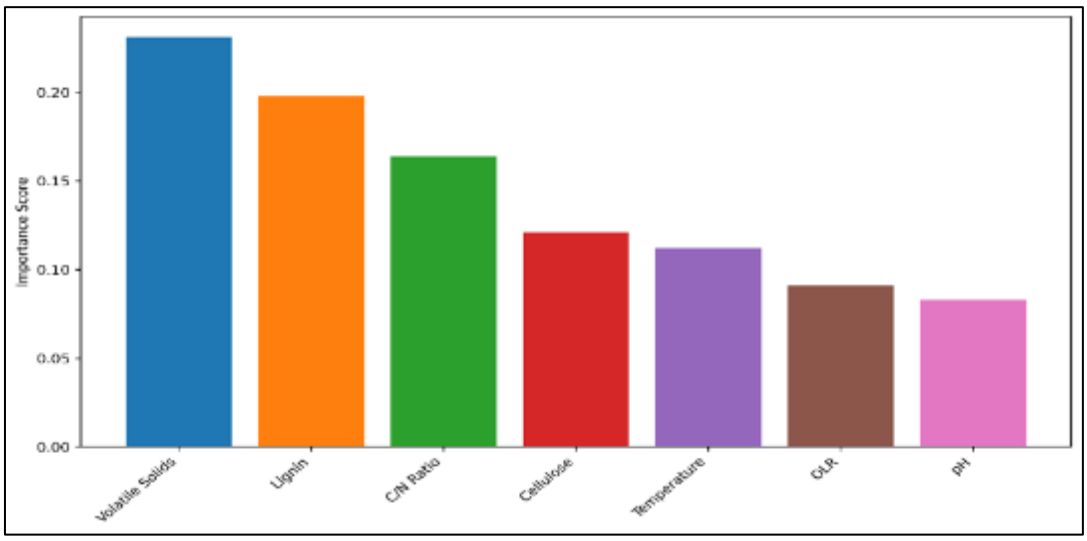

Figure 4 Feature importance (Light GBM)

The ranked of feature importance derived from the Light GBM model, highlighting the most influential predictors of methane yield is presented in figure 4. Volatile Solids emerge as the most critical feature, reinforcing their direct role in determining the amount of degradable organic matter available for microbial conversion. Lignin content ranks second, reflecting its strong inhibitory effect on biodegradability due to its rigid, aromatic polymer structure. The C/N ratio also shows high importance, emphasizing the need for nutrient balance to support the synergistic functioning of hydrolytic,

acidogenic, and methanogenic microorganisms. Cellulose, temperature, and OLR collectively contribute to methane yield variability, indicating that structural carbohydrate content and process loading significantly influence digestion kinetics. pH, although important for microbial stability, shows comparatively lower feature importance due to its lower variability across samples, as reactors were maintained near optimal conditions. Overall, the feature importance plot confirms that both biochemical composition and operational conditions play essential roles in methane production, while also demonstrating the capability of Light GBM to capture complex nonlinear dependencies within the dataset.

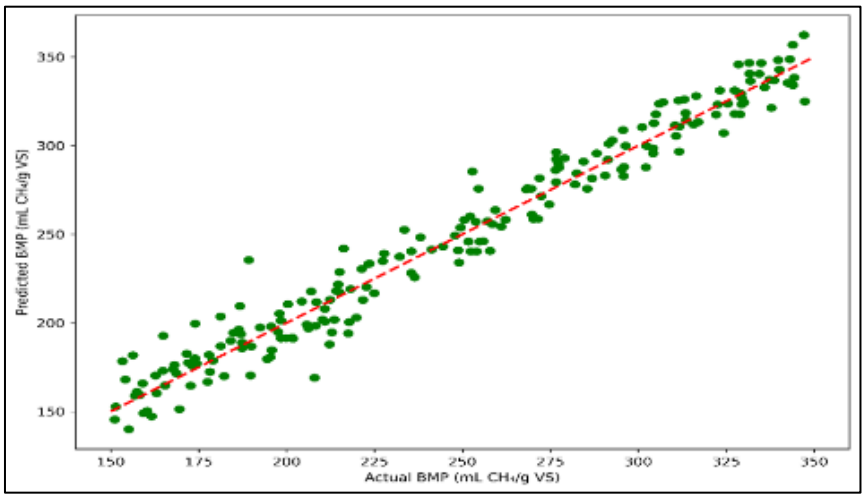

Figure 5 Predicted vs actual Methane yield (Light GBM)

The Predicted vs. Actual methane yield plot for the best-performing machine learning model is shown in figure 5. The distribution of points closely aligning along the 45° reference line indicates strong agreement between model predictions and experimentally measured BMP values. The clustering of data points along the diagonal demonstrates that the model captures both low and high methane yield ranges with high fidelity. Deviations appear minimal, with only a few observations falling farther from the reference line, suggesting low residual error and strong generalization capability. This consistency across the full range of methane yields highlights the model's ability to learn nonlinear interactions between physicochemical feedstock properties, operational parameters, and methane production dynamics. The compact spread around the diagonal further supports the model's reliability and reduced risk of overfitting. Overall, the figure confirms the suitability of the trained machine learning framework for accurately estimating methane yield across diverse agricultural residues.

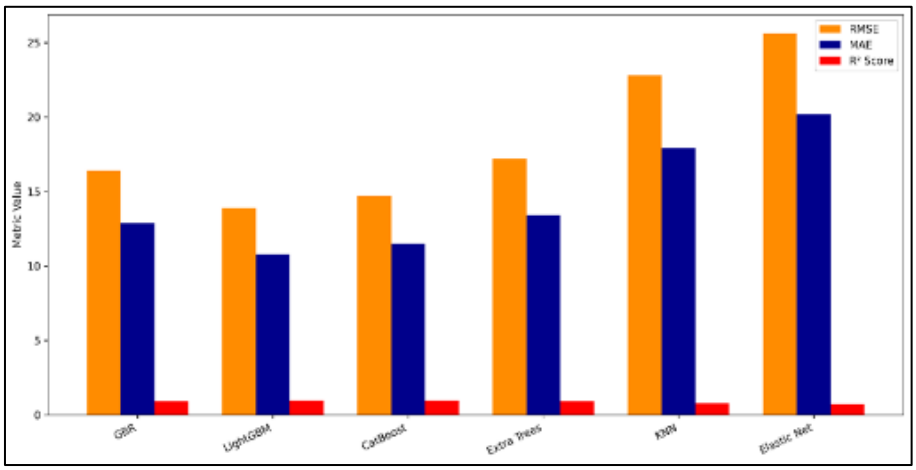

Figure 6 Performance comparison of ML models

A comprehensive comparison of all machine learning models based on RMSE, MAE, and R^2 metrics is illustrated in figure 5. The results clearly show that tree-based ensemble models, particularly Light GBM and Cat Boost, outperform the remaining techniques. Light GBM achieves the lowest RMSE and MAE values while also recording the highest R^2 score, demonstrating superior accuracy and strong predictive power. Cat Boost follows closely, benefiting from its ability to

handle feature heterogeneity and nonlinear relationships. Gradient Boosting Regressor (GBR) and Extra Trees Regressor deliver moderate performance, indicating that ensemble learning remains effective but slightly less optimized in their architectures. Conversely, K-Nearest Neighbors (KNN) and Elastic Net Regression exhibit significantly higher error values and lower R^2 scores, reflecting their limitations in capturing the multivariate and nonlinear patterns inherent in anaerobic digestion processes. The performance comparison reinforces the observation that methane yield prediction benefits most from advanced gradient boosting models capable of effectively leveraging multi-dimensional AD datasets.

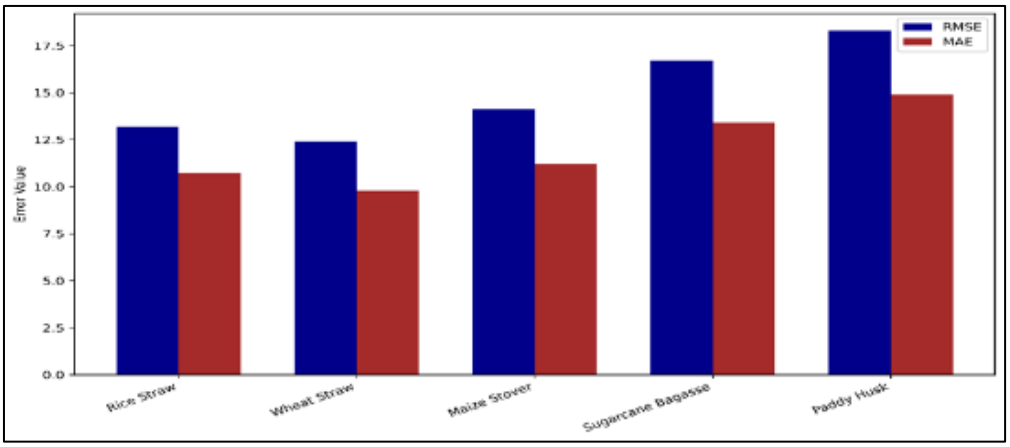

Figure 7 Error distribution across biomass

The RMSE and MAE error profiles for different biomass types, providing deeper insight into how model performance varies across feedstocks is depicted in figure 7. The lowest errors are observed for rice straw and wheat straw, indicating that their physicochemical characteristics exhibit more consistent patterns, enabling the model to achieve higher prediction accuracy. Maize stover shows moderate error values, suggesting slightly higher variability in chemical composition or digestion behavior. In contrast, sugarcane bagasse and paddy husk exhibit the highest RMSE and MAE values. This can be attributed to their elevated lignin content and structural rigidity, which introduce greater uncertainty and nonlinearity into methane yield predictions. The rising error trend from straw-based residues to more recalcitrant lignocellulosic feedstocks highlight the importance of including advanced features and model architectures when predicting methane yields for harder-to-degrade materials. Overall, the figure confirms that while the model performs well across all biomass types, prediction accuracy is naturally influenced by the inherent biochemical complexity of the feedstock.

Overall, the results confirm that machine learning models, when trained on a structured and well-engineered anaerobic digestion dataset, can effectively capture the multidimensional dependencies governing methane yield. The superior performance of Light GBM and Cat Boost demonstrates the advantage of tree-based ensemble learners in handling heterogeneous AD features and extracting meaningful patterns from complex biochemical interactions. Feature importance analysis further validates the dominant role of volatile solids, lignin content, and C/N ratio, in alignment with fundamental digestion science. While prediction accuracy varies slightly across biomass types due to inherent compositional differences, the consistent performance across all samples highlights the robustness and scalability of the proposed approach. Collectively, these results establish a strong foundation for integrating machine learning into biogas plant operation, enabling intelligent feedstock optimization, real-time yield forecasting, and data-driven process control in future AD systems.

5. Conclusion

This study demonstrates that machine learning provides a powerful and reliable approach for predicting methane yield from agricultural residues in anaerobic digestion systems. Using a structured and comprehensively labeled dataset derived from controlled experiments, the proposed models were able to capture complex nonlinear interactions among physicochemical feedstock characteristics, operational conditions, and biogas performance indicators. Among the six evaluated algorithms, Light GBM emerged as the best-performing model, achieving the highest R^2 and the lowest error metrics, thereby confirming its suitability for methane yield forecasting. Feature importance analysis further highlighted Volatile Solids, lignin content, and C/N ratio as the dominant factors governing methane production, aligning well with established biochemical principles. Overall, the findings underline the value of integrating machine learning with experimentally curated datasets to enhance the accuracy, scalability, and automation potential of anaerobic

digestion optimization. The developed predictive framework provides a strong foundation for intelligent control of biogas plants, real-time yield estimation, and data-driven operational decision-making. Future work may extend the dataset with continuous reactor data, incorporate deep learning models, and integrate real-time sensor-driven digital twins to further advance AI-enabled waste-to-energy systems.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Bayrak, Ö. T., Uludag-Demirer, S., Xu, M., and Liao, W. (2025). Forecasting the Methane Yield of a Commercial-Scale Anaerobic Digestor Based on the Biomethane Potential of Feedstocks. Energies, 18(22), 5914. https://doi.org/10.3390/en18225914

[2] Ostos, I., Ruiz, I., Cruz, D., and Flórez-Pardo, L. M. (2025). Methane Concentration Prediction in Anaerobic Codigestion Using Multiple Linear Regression with Integrated Microbial and Operational Data. Bioengineering, 12(11), 1133. https://doi.org/10.3390/bioengineering12111133

[3] Khuntia, H.K.; Paliwal, A.; Kumar, D.R.; Chanakya, H.N. Review on solid-state anaerobic digestion of lignocellulosic biomass and organic solid waste. Environ. Monit. Assess. 2022, 194, 514.

[4] Yan Wang, Tyler Huntington, Corinne D. Scown, Tree-Based Automated Machine Learning to Predict Biogas Production for Anaerobic Co-digestion of Organic Waste, ACS Sustainable Chem. Eng. 2021, 9, 12990−13000, DOI: 10.1021/acssuschemeng.1c04612

[5] Liu, X.; Coutu, A.; Mottelet, S.; Pauss, A.; Ribeiro, T. Overview of numerical simulation of solid-state anaerobic digestion considering hydrodynamic behaviors, phenomena of transfer, biochemical kinetics and statistical approaches. Energies 2023, 16, 1108.

[6] Duong, C.M., Lim, TT. Use of regression models for development of a simple and effective biogas decision-support tool. Sci Rep 13, 4933 (2023). https://doi.org/10.1038/s41598-023-32121-6

[7] Gupta, R.; Zhang, L.; Hou, J.; Zhang, Z.; Liu, H.; You, S.; Li, W. Review of explainable machine learning for anaerobic digestion. Bioresour. Technol. 2023, 369, 128468.

[8] Jeong, K.; Abbas, A.; Shin, J.; Son, M.; Kim, Y.M.; Cho, K.H. Prediction of biogas production in anaerobic co-digestion of organic wastes using deep learning models. Water Res. 2021, 205, 117697.

[9] Wang, Y.; Huntington, T.; Scown, C.D. Tree-based automated machine learning to predict biogas production for anaerobic co-digestion of organic waste. ACS Sustain. Chem. Eng. 2021, 9, 12990–13000.

[10] Schroer, H.W.; Just, C.L. Feature engineering and supervised machine learning to forecast biogas production during municipal anaerobic co-digestion. ACS EST Eng. 2023, 4, 660–672.

[11] Radočaj, D., and Jurišić, M. (2025). Comparative Evaluation of Ensemble Machine Learning Models for Methane Production from Anaerobic Digestion. Fermentation, 11(3), 130. https://doi.org/10.3390/fermentation11030130

[12] Van der Berg, D.J.; Teke, G.M.; Görgens, J.F.; van Rensburg, E. Predicting commercial-scale anaerobic digestion using biomethane potential. Renew. Energy 2024, 235, 121304.

[13] M. Erdem Isenkul et al 2025 Environ. Res. Commun. 7 065016

[14] Rutland H, You J, Liu H, Bull L, Reynolds D. A Systematic Review of Machine-Learning Solutions in Anaerobic Digestion. Bioengineering (Basel). 2023 Dec 11;10(12):1410. doi: 10.3390/bioengineering10121410. PMID: 38136001; PMCID: PMC10740876.

[15] Ling, J.Y.X.; Chan, Y.J.; Chen, J.W.; Chong, D.J.S.; Tan, A.L.L.; Arumugasamy, S.K.; Lau, P.L. Machine Learning Methods for the Modelling and Optimisation of Biogas Production from Anaerobic Digestion: A Review. Environ. Sci. Pollut. Res. 2024, 31, 19085–19104.