International Journal on Computational Engineering

Journal Homepage www.comien.org/index.php/comien

Prognostic Modelling of Biomethane Production from Waste: Application of Extreme Gradient Boosting

Thi Yen Pham [a], Lan Huong Nguyen [a,*]

[a] School of Mechanical Engineering, Vietnam Maritime University, Haiphong, Vietnam

Corresponding author: [*] huongnlvck@vimaru.edu.vn

***Abstract*— The escalating fossil fuel prices and greenhouse gases need urgent attention for a sustainable solution. The present study explores as modern machine learning approaches can be employed to prognosticate the complex biomethane generation process from organic wastes, like biowaste or food waste. The research investigates the use of organic sludges and how intelligent approaches can be employed to comprehend the complex nonlinear processes involved in biomethane production. Linear regression and Extreme gradient boosting (XGBoost) based prediction-models were developed and assessed employing a diverse set of statistical parameters, including R, R^2, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Kling-Gupta Efficiency (KGE). The results show that the XGBoost model beat the classical Linear Regression (LR) model in both the training and testing phases. During training, the XGBoost had an impressive R^2 value of 0.99994, indicating a perfect fit to the data. In contrast, LR achieved an R^2 value of 0.65464. Similarly, during the test period, XGBoost outperformed LR with R^2 values of 0.9553 to 0.9902. Furthermore, XGBoost reduced prediction errors, with significantly lower MSE and MAE values than LR. Taylor's graph better illustrates the excellent performance of the XGBoost over LR in both training and testing. These data demonstrate the ability of XGBoost to predict biomethane production, as well as its ability to improve the biomethane production process.**

***Keywords*— Biomethane production; waste; extreme gradient boosting; prognostic modelling.**

Manuscript received 15 Dec. 2023; revised 29 Jan. 2024; accepted 2 Feb. 2024. Date of publication 31 Mar. 2024.

I. INTRODUCTION

The technology known as Waste-to-Energy (WtE) provides a sustainable solution to the combined concerns of energy generation and waste management. It makes use of efficient waste management; by converting waste into energy or fuel. In this way there is less dependence on landfills, which in turn helps the environment deal with lower levels of pollution and greenhouse gases [1], [2]. Thermal processes such as incineration, flaring, and gasification are employed to convert organic waste into useful energy. This process contributes to the circular economy by recovering materials from waste streams. At the same time, WtE technologies will address the urgent need for effective waste management solutions, playing an important role in the transition to a sustainable energy environment. It does this by providing renewable energy and at the same time meeting the demand for waste management solutions [3], [4].

Additionally, waste energy provides renewable energy that can be used for a variety of purposes. The thermal energy generated from WtE plants can be used for district heating, industrial applications, or to generate electricity through steam or gas turbines. In addition, biogas produced from anaerobic digestion can be converted into biomethane, a renewable natural gas with properties similar to conventional natural gas, which is suitable for injection into natural gas pipelines, fuel transportation, or heating energy. Biomethane has emerged as a versatile sustainable energy source with many applications in various industries. As a renewable alternative to fossil fuels, biomethane can help reduce greenhouse gas emissions and combat climate change. Used in transport as compressed natural gas (CNG) or liquefied natural gas (LNG), it can significantly reduce vehicle emissions, especially for heavy transport such as in buses, trucks and ships [5], [6], [7].

Additionally, biomethane can be used to generate energy installed in districts off-grid or in remote locations, providing clean and reliable energy for communities without grid access use of ritual provides opportunities. In agriculture, biomethane derived from organic waste, such as crop residues and animal manure, offers the possibility of environmentally friendly waste management, as well as additional income for farmers through energy-renewable sales. But despite the obvious advantages, several barriers must be overcome before waste-energy biomethane technology is widely adopted technical barriers, such as efficiency and scalability, associated economic considerations investment and energy pricing, regulatory framework for waste management and renewable energy. Public acceptance and perception, are some of the factors contributing to these challenges [8], [9].

Composting constitutes a biological conversion procedure that incorporates a number of different types of microorganisms. These microorganisms break down trash that is derived from organic waste material and transform it into compost. Handling organic solid waste in a manner that is kind to the environment may be accomplished via the process of composting. A considerable market for compost may be found in agricultural regions as well as emitted zones. Maturity and stability of the compost are essential for the use that is intended for it. In addition, it takes less time than traditional composting methods. By using a wide range of chemicals and earthworms, this method has the potential to enhance both the efficiency of the process and the quality of the compost [10], [11].

Certain organic materials can only be handled via the process of co-composting, despite the fact that practically all organic wastes are capable of being composted due to the physicochemical properties they possess. The process of co-composting encompasses a multitude of advantages, such as the modification of the initial moisture content and the ratio of carbon to nitrogen, the enhancement of the effectiveness of the process, and the improvement of the quality of the compost. The co-composting of different biogenic materials together has been the subject of study in recent years [12], [13], [14]. This is because of many reasons.

There has been a significant amount of research conducted on the co-composting of a variety of organic wastes; however, there has not been a significant amount of research conducted on the co-composting of tea residues or food wastes. Leftover of the pulp and paper manufacturing process is rich in carbon and is also referred to as lignocellulosic biomass under another name. It is generally agreed that TW co-composting is an environmentally friendly method of recycling garbage that is widely used. Either as a soil amendment or a fertilizer, the finished product has the potential to be used. For this reason, it is essential to investigate a variety of composting techniques for the purpose of recycling tea waste [15], [16].

Vinasse is generally considered to be the most important waste leftover of the ethanol production industry. It is largely produced during the distillation stage of the process. Vinasse can be both caustic and alkaline because of its high pH, large quantities of sulfate and potassium, as well as considerable organic content. Vinasse also has a high pH. It is the latter that takes place as a consequence of actions that are associated with the handling of substrate for fermentation, such as the incorporation of sulfuric acid in order to regulate pH and avoid yeast flocculation [15], [17], [18]. There are three distinct varieties of sugarcane vinasse that may be produced as a consequence of the industrial process. These varieties are determined by whether the fermentable sugars that were utilized to produce ethanol originated from juice, molasses, or a mix of the two [19], [20].

Two indicators of this are the rate at which vinasse is created and the variations in the COD of the degradable biological ingredient that is present in vinasse. Both of these factors are indicators of progress. The fermentation process results in vinasse with a pH that is naturally acidic, which, if it is not neutralized, might potentially impede the process of anaerobic digestion [21], [22], [23]. Biomethane production is a complex and nonlinear process, especially when biowaste and food waste are used together This process is influenced by a variety of independent controls, making it difficult to apply strategies a they will often use modeling effectively. But modern machine learning approaches offer a promising alternative by capturing complex patterns in the input and output dynamics of food waste and agricultural waste In this context , the present study explores the use of modern machine learning techniques to model, predict and simulate complex nonlinear processes involved in organic waste co-treatment, such as artificial neural networks or support vector devices, researchers said aim to gain deeper insights into the mechanisms underlying biomethane production and improve system efficiency This new approach sustains bioenergy production to improve waste energy processes and deliver great power to improve.

II. Materials and methods

A. Raw materials and biomethane generation

The raw sugarcane vinasse was acquired from rural-based sugar manufacturing plant working on sugarcane. It was processed, filtered, and refrigerated at sun zero temperature. The subsequent analysis to determine its chemical oxygen demand (COD) and approximate content. To prepare the test substrate, the vinasse was diluted to a specific COD concentration by mixing with deionized water before introduction into the specially designed reactor. The substrate's pH was adjusted to seven by introducing 0.05 litre of sodium hydroxide per liter of substrate. Sewage treatment waste (STW), abundantly available from various sources including food factories, tea shops, and residences, was chosen as an organic waste feedstock for its potential energy production. In a circular economy framework, utilizing STW for biogas generation aligns with sustainability goals. To enhance biogas production it was supplemented with pigeon droppings. Experiments were conducted using conventional pilot-scale anaerobic digesters. The samples were stored in 5L glass reactors with airtight seals. The trials were conducted under mesophilic conditions, with mechanical stirrer facilitated regular mixing of the substrate, and biogas volume measurements were taken after manual agitation of each digester twice daily.

B. Extreme gradient boosting

XGBoost is widely recognized as an industry leader in machine learning due to its exceptional ability to solve a wide range of problems in numerous domains with remarkable accuracy and efficacy. The nomenclature of this algorithm, which stands for "eXtreme Gradient Boosting," succinctly captures its prowess: an amalgamation of gradient boosting techniques accompanied by substantial improvements in performance. XGBoost, which was initially designed by Tianqi Chen and his colleagues, has since become an indispensable instrument for data scientists and practitioners who aim to advance the limits of predictive modeling and optimization. XGBoost functions according to the ensemble learning paradigm, in which a resilient predictive model is constructed by combining multiple poor learners. On the contrary, XGBoost distinguishes itself through its clever execution of gradient boosting, a technique that enhances the performance of the model through the minimization of a predetermined loss function. By concentrating on the instances in which the model exhibits subpar performance, XGBoost iteratively enhances its predictive capabilities and consequently improves its accuracy [24], [25].

XGBoost exhibits a notable degree of versatility and adaptability, rendering it suitable for an extensive array of tasks encompassing classification, regression, ranking, and recommendation systems. Whether it be foretelling stock prices in finance, optimizing ad placements in digital marketing, or predicting customer attrition in telecommunications, XGBoost consistently provides cutting-edge outcomes, solidifying its standing as the preferred solution for numerous practical scenarios. However, what genuinely differentiates XGBoost is its capacity to efficiently process enormous datasets due to its optimized architecture and parallel processing capabilities. By capitalizing on the capabilities of distributed computing frameworks such as Apache Spark and Dask, XGBoost can scale effortlessly to accommodate datasets comprising thousands of features and millions of records, rendering it an optimal choice for industrial-scale deployments and big data analytics. Additionally, XGBoost provides an extensive array of hyperparameters that facilitate customization and fine-tuning to accommodate particular use cases and specifications. Users have complete authority over the model's behavior, including the ability to adjust learning rates, regularization penalties, tree structure, and depth. This empowers them to attain optimal performance even when confronted with the most difficult circumstances [26], [27], [28].

Notwithstanding its numerous merits, XGBoost is not devoid of obstacles and factors to be taken into account. Similar to other machine learning algorithms, optimizing its performance necessitates meticulous data preprocessing, feature engineering, and hyperparameter optimization. Furthermore, although XGBoost demonstrates exceptional predictive accuracy, its opaque structure may occasionally impede interpretability, thereby complicating the extraction of significant insights through the model's forecasts—a compromise that professionals must tactfully manage. In anticipation of the future, XGBoost exhibits encouraging signs, as continuous research and development endeavors are directed towards augmenting its functionalities. The progression of XGBoost, which involves the integration of sophisticated optimization methods and the investigation of novel approaches to enhance the interpretability and explicability of models, remains a catalyst for advancements in machine learning and predictive analytics [29], [30].

In summary, XGBoost serves as a compelling illustration of how collaboration and innovation can propel the boundaries of machine learning forward. XGBoost has solidified its position as a fundamental instrument in the repertoire of data scientists due to its exceptional performance, scalability, and adaptability. This enables professionals to address intricate challenges and discover fresh prospects within the continuously expanding domain of decision-making based on data.

III. Results and Discussion

In the present study, linear regression and XGBoost was employed for prediction modelling of biomethane generation data from anaerobic digestion of waste organic materials. The python based open access libraries were used in Jupyter.

A. *Correlational analysis of data*

The correlation heatmap is depicted in Figure 1 while the correlation matrix is listed in Table 1. An understanding of the connections that exist between the various variables in the dataset may be gained via the use of the correlation matrix. Vinasse (percent), Pigeon Dropping (percent), Incolcum to Substrate (percent), Hours of Rainfall (percent), OLR (percent), and Biomethane yield (ml/gVS) are some of the variables that correlate to each row and column in the matrix to which they belong. There is a correlation coefficient between these variables, and the numbers included inside the matrix represent that connection. One thousand is the correlation coefficient that shows a perfect positive connection, whereas one thousand is the correlation coefficient that indicates a perfect negative correlation.

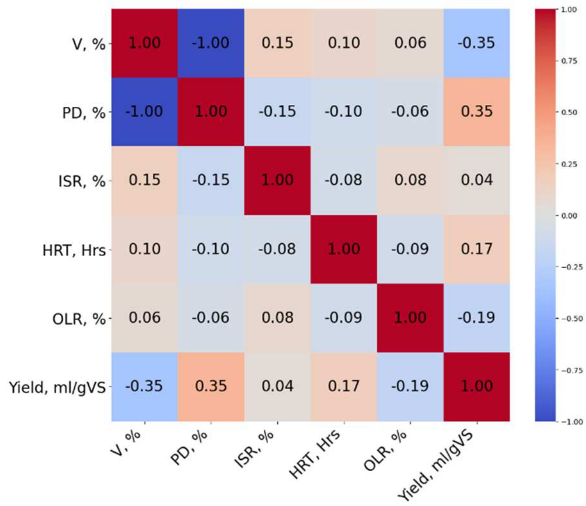


Fig. 1 Correlation heatmap of biomethane generation

The correlation coefficient of 1.000 between the percentage of Vinasse and the percentage of Pigeon Dropping implies that there is a significant negative connection between the two variables. This means that as the percentage of Vinasse, the percentage of Pigeon Dropping drops, and vice versa. Similar to the previous example, the correlation coefficient between Vinasse (%) and Yield (ml/gVS) is -0.346, which shows a somewhat negative connection. This suggests that a rise in Vinasse (%) is connected with a drop in Biomethane yield (ml/gVS). The correlation coefficient between Incolcum to Substrate (%) and OLR (%) is 0.079, which indicates that there is a rather weak positive association between the two variables. These correlations provide useful insights into the interdependencies that exist between various variables, and they have the potential to guide future analysis and decision-making processes in the context of waste-to-energy and biomethane production.

TABLE I
CORRELATION MATRIX

	V, %	PD, %	ISR, %	HRT, Hrs	OLR, %	Yield, ml/gVS
V, %	1.000	-1.000	0.149	0.096	0.056	-0.346
PD, %	-1.000	1.000	-0.149	-0.096	-0.056	0.346
ISR, %	0.149	-0.149	1.000	-0.075	0.079	0.042
HRT, Hrs	0.096	-0.096	-0.075	1.000	-0.088	0.169
OLR, %	0.056	-0.056	0.079	-0.088	1.000	-0.190
Yield, ml/gVS	-0.346	0.346	0.042	0.169	-0.190	1.000

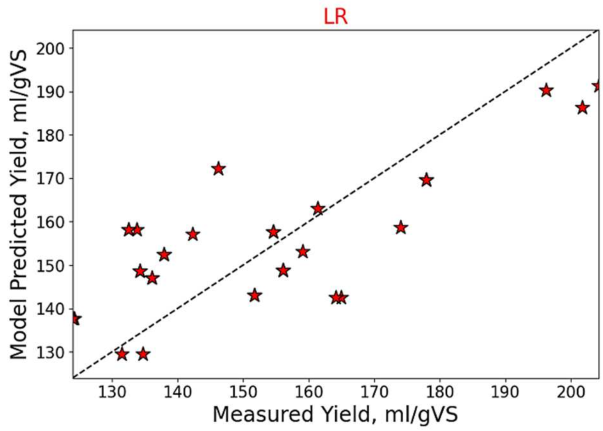
B. Model development and comparison

The models were developed and tested on different statistical metrics namely, R, R^2, Mean squared error, mean absolute error, Kling – Gupta Efficiency (KGE). This study presents the training and test results of both linear regression (LR) and XGBoost models aimed at predicting the amount of biomethane production. During the training phase, the LR model was able to obtain R^2 value of 0.65464. These results indicate that the model was able to explain about 65.46% of the variation in data related to biomethane production. The performance of the XGBoost model was significantly better than that of the LR model, with an impressive R^2 value of 0.99994, indicating an almost perfect fit to the training data. In the case of prediction errors LR had higher levels of errors as 164.97 MSE and 9.97 as MAE, compared to XGBoost with a MSE and MAE at 0.042 and 0.083, respectively.

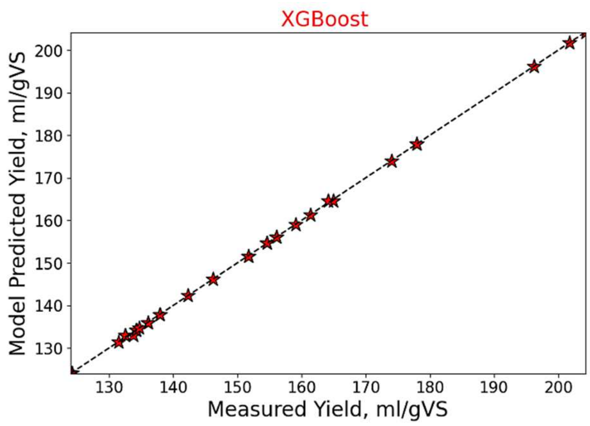
TABLE II
MODEL OUTPUTS IN STATISTICAL TERMS

Training of models					
Model	R	R^2	KGE	MSE	MAE
LR	0.8104	0.65464	0.7044	164.97	9.97
XGBoost	0.99994	0.99994	0.99994	0.042	0.083
Testing of models					
Model	R	R^2	KGE	MSE	MAE
LR	0.98103	0.9553	0.9357	30.26	3.622
XGBoost	0.9978	0.9902	0.9398	6.61	1.66

In the testing phase, the models were evaluated with a fresh potion of data. LR had fair enough predictive performance, as shown by its R^2 value of 0.9553. This figure indicates that the model was able to explain around 95.53% of the variation in the biomethane yield data. Throughout the testing process, the XGBoost model maintained its exceptional performance, reaching an R2 value of 0.9902, indicating that it provided a high fit to the data under test More specifically, LR yielded a MSE of 30.26 and mean absolute error (MAE) of 3.622, but XGBoost gave the lowest estimates of 6.61 for MSE and 1.66 for MAE. The model performance as shown in Figure 2 during model training and Figure 3 for model testing. It demonstrates that XGBoost was superior to LR based models.



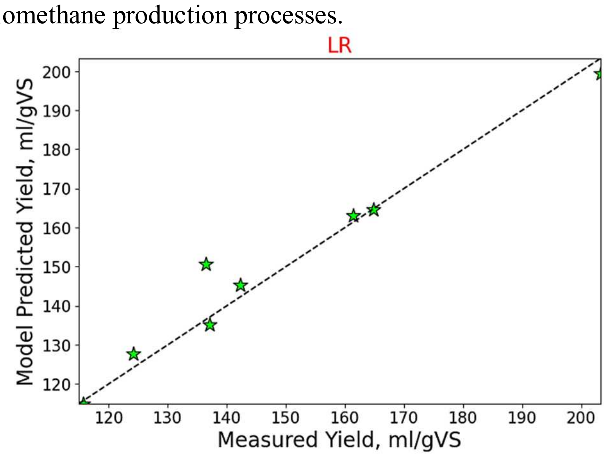
(a)



(b)

Fig. 2 Measure vs Model predicted biomethane yield during mode training for (a) linear regression (b) XGBoost

The XGBoost model outperformed the LR model in terms of accuracy, reliability, and predictive power. Generally, the XGBoost model showed impressive performance in both the training and testing phases. These findings indicate that XGBoost is a successful tool for estimating the amount of biomethane produced, and the potential to improve biomethane production processes.



(a)

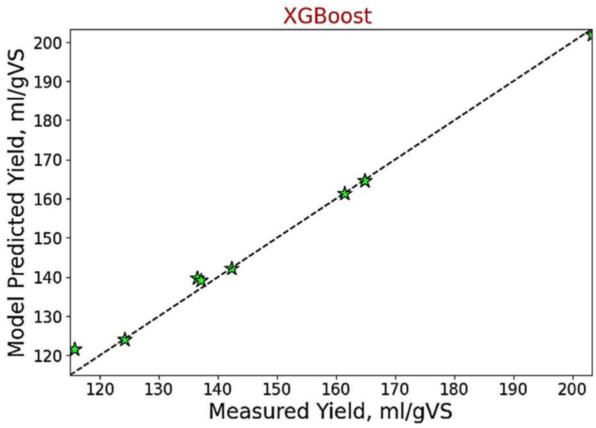
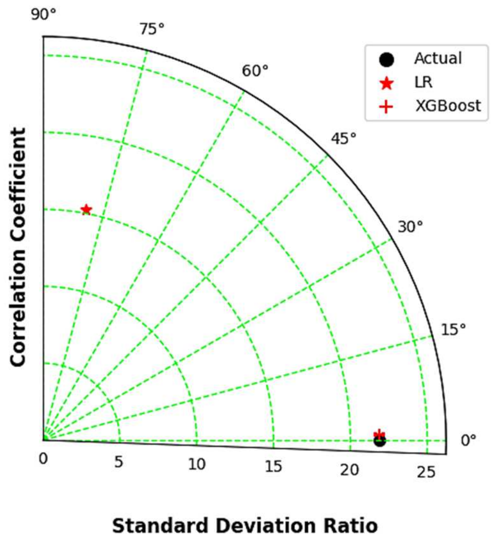
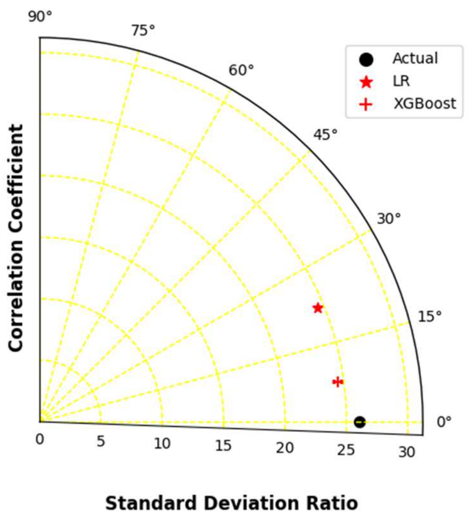


Fig. 3 Measure vs Model predicted biomethane yield during mode testing for (a) linear regression (b) XGBoost

For the purpose of comparing models, the Taylor's diagram was used, as shown in Figure 4a for the training of models and Figure 4b for the testing of models. It is not difficult to see that XGBoost has the potential to perform much better than models that are based on LR.



(a)



(b)

Fig. 4 Taylor's plots for model (a) training (b) testing

IV. Conclusion

The current work investigates how modern machine learning algorithms can be used to predict the complex biomethane generation process from organic wastes, such as biowaste or food waste. The study looks into the use of organic sludges and how intelligent methods can be used to understand the complex nonlinear processes involved in biomethane synthesis. Linear regression and Extreme gradient boosting (XGBoost)-based prediction models were created and evaluated using a wide range of statistical parameters. The following are the main results of the study: XGBoost model outperformed Linear Regression (LR) in training and testing phases; XGBoost achieved an R^2 value of 0.99994 during training, indicating perfect data fit; XGBoost outperformed LR in testing with R2 values of 0.9553 to 0.9902; XGBoost reduced prediction errors with lower MSE and MAE values; Taylor's graph demonstrates XGBoost's superior performance in both training and testing.

References

[1] J. A. Siles, I. García-García, A. Martín, and M. A. Martín, "Integrated ozonation and biomethanization treatments of vinasse derived from ethanol manufacturing," *J Hazard Mater*, vol. 188, no. 1–3, pp. 247–253, Apr. 2011, doi:10.1016/J.JHAZMAT.2011.01.096.

[2] V. R. Moreira, T. G. Carpanez, N. C. Magalhães, Y. F. X. Ladeira, L. C. Lange, and M. C. S. Amaral, "Ultrafiltration as a pre-treatment technology to improve vinasse biomethanation," *Process Safety and Environmental Protection*, vol. 169, pp. 718–724, Jan. 2023, doi:10.1016/J.PSEP.2022.11.061.

[3] V. G. Nguyen, M. H. Tran, P. Paramasivam, H. C. Le, and D. T. Nguyen, "Biomass: A Versatile Resource for Biofuel, Industrial, and Environmental Solution," *Int J Adv Sci Eng Inf Technol*, vol. 14, no. 1, pp. 268–286, Feb. 2024, doi: 10.18517/ijaseit.14.1.17489.

[4] T. T. Le *et al.*, "An Experimental Assessment of Waste Transformer Oil and Palm Oil Biodiesel Blended with Diesel Fuel on A Single Cylinder Direct in Diesel Engine," *Int J Adv Sci Eng Inf Technol*, vol. 14, no. 1, pp. 246–258, Feb. 2024, doi: 10.18517/ijaseit.14.1.15998.

[5] K. Ravi, S. Mathew, J. Pradeep Bhasker, and E. Porpatham, "Gaseous alternative fuels for CI engines - a technical review," *International Journal of Pharmacy and Technology*, vol. 8, no. 4, pp. 5257–5268, 2016.

[6] B. B. Sahoo, N. Sahoo, and U. K. Saha, "Effect of engine parameters and type of gaseous fuel on the performance of dual-fuel gas diesel engines—A critical review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6–7, pp. 1151–1184, Aug. 2009, doi:10.1016/J.RSER.2008.08.003.

[7] V. S. Yaliwal, N. R. Banapurmath, N. M. Gireesh, and P. G. Tewari, "Production and utilization of renewable and sustainable gaseous fuel for power generation applications: A review of literature," *Renewable and Sustainable Energy Reviews*, vol. 34, no. June, pp. 608–627, 2014, doi: 10.1016/j.rser.2014.03.043.

[8] M. Prussi, M. Padella, M. Conton, E. D. Postma, and L. Lonza, "Review of technologies for biomethane production and assessment of Eu transport share in 2030," *J Clean Prod*, vol. 222, 2019, doi:10.1016/j.jclepro.2019.02.271.

[9] O. A. Aworanti *et al.*, "Enhancing and upgrading biogas and biomethane production in anaerobic digestion: a comprehensive review," *Frontiers in Energy Research*, vol. 11. 2023. doi:10.3389/fenrg.2023.1170133.

[10] A. Mandpe *et al.*, "Exploring the synergic effect of fly ash and garbage enzymes on biotransformation of organic wastes in in-vessel composting system," *Bioresour Technol*, vol. 322, no. October 2020, p. 124557, 2021, doi: 10.1016/j.biortech.2020.124557.

[11] A. Le Pera, M. Sellaro, E. Bencivenni, and F. D'Amico, "Environmental sustainability of an integrate anaerobic digestion-composting treatment of food waste: Analysis of an Italian plant in the circular bioeconomy strategy," *Waste Management*, vol. 139, pp. 341–351, Feb. 2022, doi: 10.1016/j.wasman.2021.12.042.

[12] M. Alruqi and P. Sharma, "Biomethane Production from the Mixture of Sugarcane Vinasse, Solid Waste and Spent Tea Waste: A Bayesian Approach for Hyperparameter Optimization for Gaussian Process

Regression," *Fermentation*, vol. 9, no. 2, p. 120, Jan. 2023, doi:10.3390/fermentation9020120.
[13] T. B. N. Nguyen and N. V. L. Le, "Biomass resources and thermal conversion biomass to biofuel for cleaner energy: A review," *Journal of Emerging Science and Engineering*, vol. 1, no. 1, pp. 6–13, Sep. 2023, doi: 10.61435/jese.2023.2.
[14] V. H. Dong and P. Sharma, "Optimized conversion of waste vegetable oil to biofuel with Meta heuristic methods and design of experiments," *Journal of Emerging Science and Engineering*, vol. 1, no. 1, pp. 22–28, Sep. 2023, doi: 10.61435/jese.2023.4.
[15] M. J. Diaz, E. Madejón, F. López, R. López, and F. Cabrera, "Optimization of the rate vinasse/grape marc for co-composting process," *Process Biochemistry*, vol. 37, no. 10, pp. 1143–1150, May 2002, doi: 10.1016/S0032-9592(01)00327-2.
[16] S. Mahapatra, Md. H. Ali, and K. Samal, "Assessment of compost maturity-stability indices and recent development of composting bin," *Energy Nexus*, vol. 6, p. 100062, Jun. 2022, doi:10.1016/J.NEXUS.2022.100062.
[17] V. Robles-González, J. Galíndez-Mayer, N. Rinderknecht-Seijas, and H. M. Poggi-Varaldo, "Treatment of mezcal vinasses: A review," *J Biotechnol*, vol. 157, no. 4, pp. 524–546, Feb. 2012, doi:10.1016/j.jbiotec.2011.09.006.
[18] C. A. Christofoletti, J. P. Escher, J. E. Correia, J. F. U. Marinho, and C. S. Fontanetti, "Sugarcane vinasse: Environmental implications of its use," *Waste Management*, vol. 33, no. 12, pp. 2752–2761, Dec. 2013, doi: 10.1016/j.wasman.2013.09.005.
[19] L. F. Ferreira, M. Aguiar, G. Pompeu, T. G. Messias, and R. R. Monteiro, "Selection of vinasse degrading microorganisms," *World J Microbiol Biotechnol*, vol. 26, no. 9, pp. 1613–1621, Sep. 2010, doi:10.1007/s11274-010-0337-3.
[20] A. Djalma Nunes Ferraz Júnior, M. H. Koyama, M. M. de Araújo Júnior, and M. Zaiat, "Thermophilic anaerobic digestion of raw sugarcane vinasse," *Renew Energy*, vol. 89, pp. 245–252, Apr. 2016, doi: 10.1016/j.renene.2015.11.064.
[21] A. Cabello, T. Mendiara, M. Teresa Izquierdo, F. García-Labiano, and A. Abad, "Energy use of biogas through chemical looping technologies with low-cost oxygen carriers," *Fuel*, vol. 344, 2023, doi: 10.1016/j.fuel.2023.128123.
[22] S. Isarapakdeetham *et al.*, "Hydrogen production via chemical looping steam reforming of ethanol by Ni-based oxygen carriers supported on CeO2 and La2O3 promoted Al2O3," *Int J Hydrogen Energy*, vol. 45, no. 3, pp. 1477–1491, Jan. 2020, doi:10.1016/j.ijhydene.2019.11.077.
[23] D. Hu *et al.*, "Insight into the biomass pyrolysis volatiles reaction with an iron-based oxygen carrier in a two-stage fixed-bed reactor," *Chemical Engineering Journal*, vol. 465, 2023, doi:10.1016/j.cej.2023.142860.
[24] Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, and C. Li, "Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration," *Eng Comput*, vol. 38, no. S5, pp. 4145–4162, Dec. 2022, doi:10.1007/s00366-021-01393-9.
[25] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi:10.1145/2939672.2939785.
[26] O. Sagi and L. Rokach, "Approximating XGBoost with an interpretable decision tree," *Inf Sci (N Y)*, vol. 572, pp. 522–542, Sep. 2021, doi: 10.1016/J.INS.2021.05.055.
[27] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," *Int J Distrib Sens Netw*, vol. 18, no. 6, 2022, doi: 10.1177/15501329221106935.
[28] X. Cui, D. Shi, Z. Chen, and F. Xu, "Parallel Forestry Text Classification Technology Based on XGBoost in Spark Framework," *Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery*, vol. 50, no. 6, 2019, doi:10.6041/j.issn.1000-1298.2019.06.032.
[29] S. Pan, Z. Zheng, Z. Guo, and H. Luo, "An optimized XGBoost method for predicting reservoir porosity using petrophysical logs," *J Pet Sci Eng*, vol. 208, 2022, doi: 10.1016/j.petrol.2021.109520.
[30] S. Ben Jabeur, S. Mefteh-Wali, and J. L. Viviani, "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Ann Oper Res*, 2021, doi: 10.1007/s10479-021-04187-w.