Innovations

Comparative Analysis of Biogas Production from Anaerobic Co-digestion of Lignocellulosic Biomass and Cow dung: A Kinetic and Machine Learning Approach

Dimoha Chinwendu[a*], Godwin Mbamalu[a], Fadoju Sunkanmi[b], Okafor Joshua[c], Ezeasor Ikenna[a], Obiora Blessing[a]

[a]Centre of Excellence in Environmental Management and Green Energy, University of Nigeria Nsukka, Enugu State, Nigeria

[b] University of Abuja , Nigeria

[c] Malaria Consortium, Abuja, Nigeria

[a]orcid.org/0009-0007-4722-5601

Corresponding Author: **Dimoha Chinwendu**

Abstract: *This study investigates the prediction of biogas production from the anaerobic co-digestion of cabbage waste, leaf litter of sandbox and cow dung using both traditional kinetic models and a machine learning algorithm. Specifically, the Modified Gompertz, Modified Richards, and Logistic Function models were compared against the XGBoost algorithm to evaluate their goodness-of-fit to experimental data. The goal of this study was to determine which model best predicts biogas production with respect to retention time, thereby providing insights into the dynamic behavior of the anaerobic digestion process. Nine (9) 32-liter capacity biodigesters labeled A-I, control (A-C), pretreated (D-F) and blended (G-I) waste samples were utilized for the experiment. Biodigesters A-C contained 6 kg of waste samples and 18 litres of water. Similarly, biodigesters D-F contained 6 kg of wastes samples pretreated with 0.6% ash, while biodigesters G-I contained a mixture of the three wastes blended in varying proportions of 20:40:40, 50:25:25 and 60:20:20. The daily biogas output was evaluated using the water displacement method. The volume of gas in the feedstocks within the biodigester was measured daily for 35 days. Experimental data from the study was analyzed using basic statistics, non-linear regression models and Machine Learning Algorithm (XGBoost).The study found that pretreatment and blending using varying waste blending proportions significantly increased biogas production when compared to the control waste samples. Also amongst the models compared, the logistic function model produced the highest coefficient of determination (R^2 = 0.82), followed by the modified Richards model (R^2 = 0.81) and the modified Gompertz model (R^2 = 0.79). The XGBoost model yielded a lower R^2 of 0.71. These results indicate that traditional kinetic models are more accurate in predicting biogas production than the XGBoost model. This study demonstrated that traditional kinetic models, particularly the Logistic Function model, was more effective for predicting anaerobic digestion and biogas production processes compared to the XGBoost machine learning model. The* *findings of the study confirmed that kinetic models are reliable tools for accurately modeling and forecasting biogas production in anaerobic digestion systems.*

Keywords: *Biogas Production, Kinetic Models, Machine Learning Algorithm, Pretreatment, Anaerobic Co-digestion*

Correspondence: Dimoha Chinwendu. Centre of Excellence in Environmental Management and Green Energy, University of Nigeria Nsukka, Enugu State, Nigeria.

1. Background

Crude oil has been the foundation of the global economy and modern lives, providing 1/3 of the world's energy for several decades. Oil is utilized in many activities, such as transportation, powering gadgets, appliances, and most industrial processes. This excessive dependence on oil as a principal energy source has resulted in environmental pollution, greenhouse gas emissions, climate change and other associated illnesses (1) Constant power generation, power supply and failure have grossly affected economic growth and development, especially in the agricultural, power, industrial, and mining sectors (2). The overwhelming energy demands of both urban and rural residents suggest that more renewable energy sources should be investigated (3). In developing nations, large volumes of wastes are being generated and waste disposal has become a significant issue which has led to health risks, contamination of the environment and the need for sustainable waste management solutions to lessen its detrimental effects on ecosystems and communities. Therefore, converting these wastes into energy (biogas) may provide a means to address some of these energy-related issues

Biogas is a viable renewable energy source produced from the anaerobic co-digestion of wastes which helps lower greenhouse gas emissions and promote a circular economy. Nevertheless, for the anaerobic codigestion process to be effective, the biogas generation capacity of various waste substrates must be assessed (4). Due to the strong dependence of biogas quality and cumulative yield on feedstock composition, agricultural and green waste may not be suitable for use as feedstock alone. According to (5-6), lignocellulosic biomass is a suitable feedstock for biogas production which comprises of plant-based materials rich in cellulose, hemicellulose, and lignin which is a key resource for sustainable energy production. In the context of biogas generation, its complex structure poses challenges for microbial digestion, influencing the efficiency of anaerobic co-digestion processes. Amongst other lignocellulosic biomass/wastes, cabbage waste and leaf litter of sand box were selected as feedstock for biogas production due to their high lignocellulosic, carbon to nitrogen and high-water content which enhances microbial decomposition and biogas yield during anaerobic co-digestion. Although, biogas production from cabbage was found to be relatively low, but the viability of increasing the biogas yield from cabbage waste by codigesting it with animal dung and the kinetics of the process are largely unexplored. As a result, the quality of the feedstock, co-digestion and the operating environment of the digester are required to increase the efficiency of biogas generation. Pre-treatments like ash application can help to break down these tough fibers, increasing the surface area available for microbial action and improving the overall digestibility of the biomass. Ash, being alkaline, also helps to adjust the pH levels of the substrate, creating a more favorable environment for anaerobic bacteria, which are essential for efficient biogas production. Additionally, ash contains minerals that can serve as micronutrients, further stimulating microbial activity and enhancing biogas yield. The choice of ash for pre-treatment is particularly justified as it is a readily available, low-cost byproduct of agricultural or forestry processes, making it an environmentally friendly and economically viable option for improving biogas generation from co-digested green waste and cow dung. According to a study by (7), increased biogas production and faster degradation rates are two advantages of anaerobic co-substrate digestion. The formation of volatile fatty acids (VFAs), a balanced C/N ratio, nutrients from organic waste, an increase in pH buffering capacity, a decrease in ammonia toxicity, and improved biochemical conditions for microbial development are the main benefits of mixed feedstock (8).

Due to the complexity of the anaerobic digestion process, several computational models have been developed to represent the organic waste's biochemical transformation during the AD process. These models aim to minimize costs and process variability, minimize energy loss, predict and optimize mixing ratios and Organic Loading Rate (OLR) shorten process times, and maximize biogas production (9-10). The first-order kinetic, modified Gompertz models (11-13), the Monod model, and the transference function (14-15) are some of the mathematical models that were utilized to characterize the kinetic parameters.

Kinetic and machine learning models can be difficult to employ because of their high data requirements, numerous state variables, and constants that require calibration. As a result, simpler models that can be tailored to certain circumstances may be needed (16). When models such as the first-order kinetic model attempt to adapt to different substrates and changing situations, they encounter slight difficulties that are resolved by dynamic parameter optimization with machine learning (17). Machine learning models can effectively capture the complex, nonlinear relationships that govern the dynamic processes involved in acidogenesis, hydrolysis, and AD by analyzing large datasets generated from biogas production (18). Therefore, to enable dynamically adjusted model settings and produce more accurate process forecasts across a variety of circumstances, machine learning techniques, modified Richards models, logistic function model and other relevant tools can be helpful (19). The understanding of the intricacies of anaerobic digestion and the prediction of its outcomes have advanced with the recent application of machine learning methods such as support vector machines (SVMs), artificial neural networks (ANNs), and ensemble approaches (20). Although ANN is the most widely used machine learning model and is well known for its high prediction accuracy, models can occasionally become complex and data-driven, which frequently leads to low interpretability of the model outputs (21). Thus, it is necessary to assess several machine learning models to ascertain whether it is possible to attain a balance between interpretability and accuracy, which is crucial for comprehending intricate procedures such as anaerobic digestion (22). Additionally, new machine learning models such as the generalized additive models (GAMs), XGBoost, have recently acquired popularity due to their remarkable interpretability and accuracy (23). Furthermore, hybrid models that include machine learning and classical kinetics may be able to accurately depict the intricacies of anaerobic digestion (24). For example, these models might apply machine learning to adjust to changing conditions and retain accuracy in dynamic scenarios while using traditional kinetic models for steady-state forecasts (25). By accounting for dynamic substrate properties and variations in process conditions, these models can enhance prediction capacities (26) and ultimately result in significant advancements in the field by offering rigorous insights into the dynamics of anaerobic digestion systems. Therefore, including machine learning and kinetic models such as modified Gompertz, Richards, Logistic function and XGBoost in AD process modeling should lead to a better understanding of the complexities of anaerobic digestion. Therefore, the goal of this study is to determine which model best predicts biogas production with respect to retention time during the anaerobic digestion process.

2.0. Methods

2.1. Materials and Sample Collection.

The wastes used for this research were cabbage waste obtained from a vegetable market located in Nsukka, Enugu State. The leaf litter of the sandbox was collected from leaf droppings of sandbox trees found within the University of Nigeria, Nsukka and its environs. Cow dung was obtained from a slaughterhouse in the Nsukka market, Enugu State, southeastern Nigeria. The empty fruit bunches (EFBs) were collected from a local palm oil mill in Nsukka, Enugu State, Southeast Nigeria. The empty fruit bunch was burned, and the ash was collected for use in alkaline treatment of the mixed vegetable waste. The biogas production experiment was conducted at National Centre for Energy Research and Development (NCERD), University of Nigeria, Nsukka.

2.2. Biogas Experimental Procedures

Nine (9) 32-L biodigesters (figure 1) labeled A-I were used for this experiment. Waste substrates (cabbage +leaf litter from the sandbox) were collected and allowed to degrade for 4 weeks, then chopped into small pieces, and soaked in a water bath for 7 days to allow partial decomposition of the waste by aerobic microbes to breakdown the cellulose before being charged into the biodigesters. Empty fruit bunches (EFB) were burned, and the residue collected as ash was used for the pretreatment of the waste. Cow dung slurry was used as a source of inoculum for this study since it contains anaerobic microbial population

The moisture level of the feed stocks determined the water to waste ratios used for charging the biodigesters; the pre-decayed control, pretreated and blended waste samples were charged into individual biodigesters of 32L capacity labelled A-I . The slurries were prepared by taking and mixing 6 kg of each waste sample and 18 L of water, resulting in a waste-to-water ratio of approximately 1:3. A total of 6 kg each of cabbage waste, sandbox, cow dung and 18 L of water was added to the control biodigesters: A=cow dung (CD) alone, B=cabbage (CB) alone, and C=sandbox (SB) alone. The pretreated biodigesters D=Cow dung +Ash, E=Cabbage +Ash, and F= Cow dung +Ash contained 6 kg of cow dung, cabbage waste and sandbox pretreated with 0.6% ash, while biodigesters G-I contained a mixture of the three wastes blended in varying proportions by weight, G=20:40:40, H=50:25:25, and I=60:20:20. Waste and water occupied 75% of the digester, which means that 24 L of the digester was occupied, while the remaining space and volume were occupied by the gas that was generated. The substates were charged into the biodigesters, and readings were taken thrice daily at 7:00 am, 1:00 pm and 7:00 pm daily. The experiment was carried out for a 35-day retention period. The quantity of biogas produced in liters was obtained by downward displacement of water. A liquid in glass thermometer was used to measure and record the ambient and slurry temperatures daily, while a digital pH meter (Jenway, 3510) was used to monitor the pH of the waste. This experiment was carried out under mesophilic temperature conditions. The daily average volume of gas, PH and temperature was recorded and used for the data analysis. The gas produced was analyzed using gas chromatography that accurately measures the concentration of biogas composition.

2.3. Analytical Methods.

The AOAC (1990) method [27] was utilized to ascertain the contents of ash, moisture, and fiber. Using the techniques outlined in Pearson (1976) [28], soxhlet extraction and micro-Kjedhal were used to determine the amounts of fat, crude nitrogen, and protein. The Meynell (1976) method [29] was used to determine total and volatile solids, whereas the Walkey and Black (1934) method [30] was used to assess carbon content.

Figure 1: A 32 L biodigester set

2.4. Model description for biogas production.

The traditional kinetic models employed for this study were the modified Gompertz model, modified Richards model, and logistic function model, as shown in the equations below. They were utilized to demonstrate the lag time, experimental maximum cumulative biogas production, and forecast of maximum biogas production. The machine learning algorithm (XG boost) used for this study was developed based on conventional procedures.

2.4.1. Kinetic Model Formulas:

Equation 1 represents the modified Gompertz equation (31).

$$y = A \times \exp\{-\exp[Rmax \times e \times (\lambda - t)/A + 1] \quad \text{(eqn 1)}$$

where y represents the cumulative biogas yield (L) with respect to time t (d),

A is the maximum cumulative biogas yield (L),

R_{max} is the maximum biogas production rate (L/d),
λ is the lag phase time (d), and e is a constant equal to 2.71.

The logistic function model (32) is represented by equation 2.

$$y = A/\{1 + \exp [4 \times R_{max} \times (\lambda - t)/A + 2]\} \quad \text{(eqn 2)}$$

where A, Rmax, and λ have the same meanings as those of the modified Gompertz equation above.

c). The modified Richards model is represented by equation 3 (33)

$$y = A \times \{1 + v \times \exp (1 + v) \times \exp [R_{max} \times (1 + v) \times (1 + 1/v) \times (\lambda - t)/A]\} (-1/v) \quad \text{(eqn 3)}$$

where A, R_{max}, and λ have the same meanings as above and v is the shape coefficient.

Nonlinear regression analysis (modified Richards, Gompertz and logistic function equation) was performed using Microsoft Excel and Origin 9.0 Pro software. The software searched for values of A, R_{max}, and λ that had minimum residual sums of squares along with their respective 95% confidence intervals. During the fitting process, the biogas yield curve and degree of correlation between variables was indicated by R^2 were determined.

2.4.2. XGBoost Model for machine learning

This model, in contrast to analytical function, follows no predetermined mathematical form (34). The biogas dataset included features such as waste sample, No of Days, Ambient Temperature (AT ℃), Slurry Temperature (ST ℃), pH, and Volume of Gas. The categorical column (Sample) was encoded using Label Encoding. The key features selected for model training were Sample, No of Days, AT ℃, ST ℃, and pH, while Volume of Gas served as the target variable. The dataset was divided into training and testing sets with an 80:20 ratio and XGBoost regressor was configured with specific parameters: objective set to reg, col sample_by tree to 0.3, learning rate to 0.1, max depth to 5, alpha to 10, and n_estimators to 10. The model was trained on the training data, and its performance was evaluated on the test set using metrics such as Mean Squared Error (MSE) and R-squared (R^2). A scatter plot of actual vs. predicted values visually showed the model's performance (35).

3.0. Results.

3.1. Biogas yields from feedstocks.

The daily biogas production from the anaerobic digestion of the control, pretreated, and blended waste samples (A-I) was monitored over a 35-day retention period as shown in figure 2a. On the first day of digestion, the daily biogas production of the pretreated and blended waste groups was greater than that of the control group. The pretreated waste samples D and E exhibited peak biogas production on the 2nd and 11th day, while the control groups (A, B, and C) peaked on the 21st, 31st and 35th, day respectively. For the blended waste groups (G, H, and I) reached their peak on the 14th, 18th, and 31st days of the experiment respectively. The pretreated waste sample (CB + ash) produced the highest quantity of daily biogas (8.4 L/day), followed by the pretreated and blended waste samples

CD + ash, SB + ash and CD+CB+SB (50:25:25) which produced 6.2L/day of biogas each, all of which surpassed those of the control group.

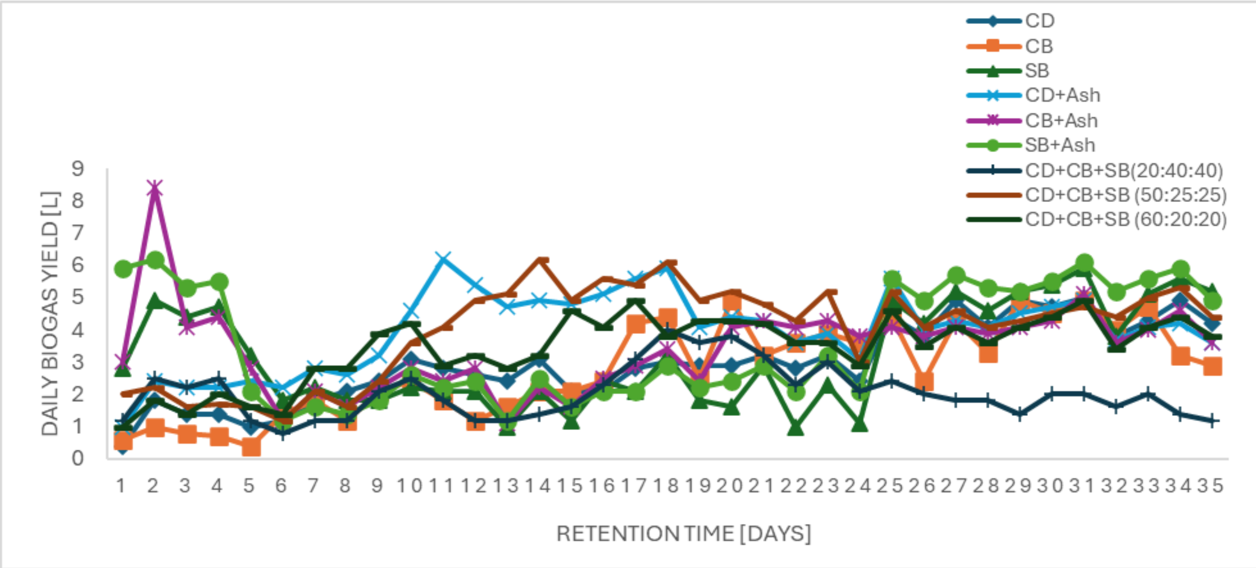


Figure 2 a. The daily biogas yields of the control, pretreated and blended waste substrates of cabbage, leaf litter from sandboxes and cow dung.

The graph (Figure 2b) shows the cumulative biogas yield for various waste substrates and combinations during a 35-day period. Compared to other single substrates, cow dung (CD+ ash) produces the most biogas, yielding 140.3L. With a yield of approximately 60 L, cabbage biomass (CB) has the lowest yield. Although adding ash only slightly affects the CD and SB yields, it greatly increases the CB yield. In comparison, combinations of CD, CB, and SB, especially at a 50:25:25 ratio, produce better yields of 135.9 L. This implies that although CD, CB and SB alone produces efficient yields of biogas, blending in varying proportions may improve the output of biogas from these waste samples.

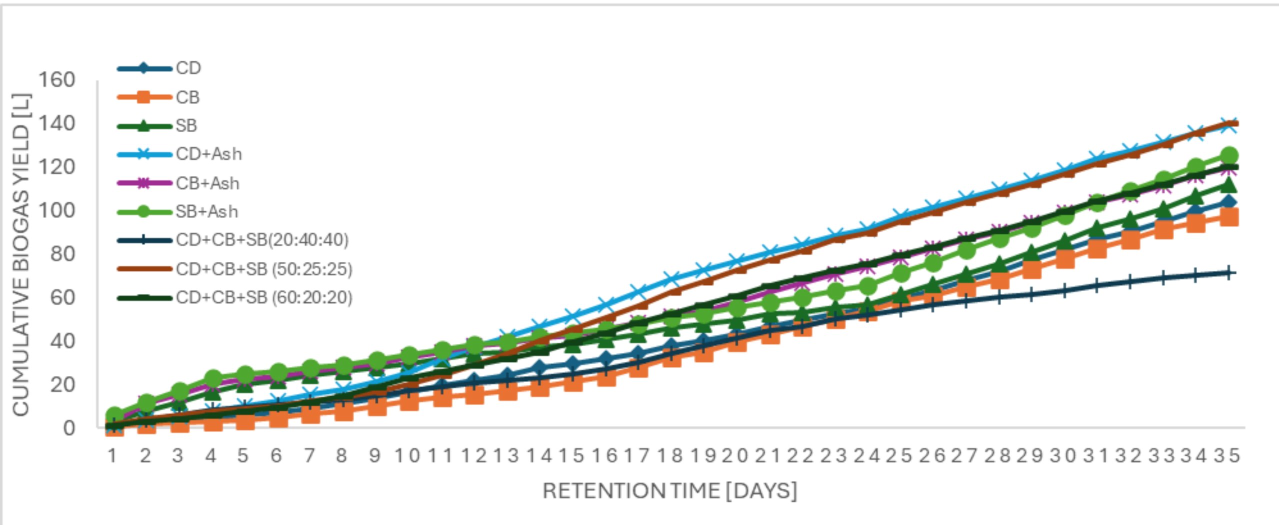


Figure 2b. The cumulative biogas yields of control, pretreated and blended waste substrates of cabbage, leaf litter from sandboxes and cow dung.

3.2: Physicochemical characteristics of waste samples.

Table 1 shows that sample A presented a higher total solid content (3.09), a volatile solid content (2.63) and the lowest C:N ratio (21:48). On the other hand, sample C had the lowest values for T.S. (0.38) and V.S. (0.31) and the second highest for C:N (33.78), while sample B had the highest carbon content (12.52) and highest C:N ratio (47.07). These results are attributed to the nature and composition of the waste substrates used. The results also revealed that sample D had the highest value for ash (0.50) and the lowest value for moisture (95.11). On the other hand, sample E has the highest value for moisture (97.31) and fiber (0.50), while sample H has the lowest value for ash (0.05). This variation is due to the variation in waste composition and the overall effect of pretreatment on the waste.

Table 1: Proximate composition and physicochemical properties of the feedstock.

Samples	Parameters (%)							
	Moisture	**Ash**	**Fiber**	**N**	**TS**	**VS**	**Carbon content**	**C: N**
A	95.11	0.47	0.40	0.446	3.09	2.63	9.58	21.48
B	95.50	0.16	0.60	0.266	0.49	0.33	12.52	47:07
C	95.61	0.13	0.50	0.307	0.38	0.31	10.37	33.78
D	95.11	0.50	0.40	0.404	2.89	2.35	8.78	21.73
E	97.31	0.17	0.50	0.238	0.68	0.34	10.95	46.00
F	96.70	0.47	0.40	0.272	1.26	0.36	9.58	35.22
G	95.90	0.07	0.50	0.322	0.25	0.21	11.17	34.69
H	95.51	0.05	0.40	0.370	0.50	0.39	10.74	29:03
I	95.80	0.19	0.40	0.384	1.19	1.00	10.17	26.48

3.3. Biogas Percentage Compositions (Gas Analysis) for Sample Biodigesters Using Gas Chromatography.

The result shown in Table 2 shows the GC results and composition of biogas produced from the different biodigesters (A-I) respectively. The biodigester (I) containing cow dung, cabbage and sandbox in ratio (60:20:20) had a higher percentage of methane (41%) followed by digester D and E (40%) with biodigester A, B, C, F, G, H producing the least quantity of methane (22%). However, biodigester D produced no CO2 at all followed by biodigester E, F ,1 producing the least quantity of carbon dioxide (6%) followed by biodigester A, B, C with the highest quantity of CO2 generation (14%).

Table 2: Percentage Composition of Biogas Produced in the Digesters Using Gas Chromatography

Parameters	Digester A	Digester B	Digester C	Digester D	Digester E	Digester F	Digester G	Digester H	Digester I
Methane	22%	22%	22%	40%	40%	22%	22%	22%	41%
Phenol	4%	4%	4%	2%	-	8%	8%	14%	7%
CO2	14%	14%	14%	11%	11%	6%	6%	-	6%
Methanol	8%	8%	8%	12%	15%	9%	9%	9%	8%
Ethanol	-	1%	1%	10%	3%	-	-	7%	-
Acetone	10%	17%	17%	-	-	17%	17%	19%	-

3.4. Kinetic Modeling

The results of the analysis of the models for the waste samples are presented in Figure 3. The modified Richards, Gompertz and logistic function models were used to fit the cumulative biogas yields from the waste samples (Figure 3). The figure compares the experimental data with the model fitting results from the modified Gompertz, Richards and logistic models.

3.4.1. Modeling Cumulative Biogas Yield Kinetics Using Modified Gompertz Model.

The results obtained from the modified Gompertz model are shown in Table 3. Blended waste sample G had the highest R2 value of 0.79, showing a good curve fit for the experimental and predicted data, while control waste sample C had the lowest R^2 value of 0.25, indicating a poor curve fit for the data. Pretreated waste sample D had the highest A value of 111.65 L, while blended waste sample G had the lowest A value of 96.85 L. The control and blended waste samples showed the best and lowest fitting R_{max} values. The highest R_{max} value of 21.08 was obtained for control waste sample B, while the lowest value of 11.09 was obtained for blended waste sample G, showing that the maximum and minimum fitting R_{max} values were obtained from different waste sample groups. The lag phase (λ) values show that pretreated waste sample G has the highest value of 1.58, and the blended waste sample has the lowest value of 0.94.

3.4.2 Modeling Cumulative Biogas Yield Kinetics Using Modified Richard's Model.

Table 3 shows that the highest Rmax value of 15.66 was obtained for pretreated waste sample E, while the lowest value of 8.16 was obtained for control sample B. The highest A value of 275.3 L was obtained from waste sample C, while blended waste sample H had the lowest value of 201.1 L. The lag phase (λ) values show that pretreated waste sample E has the highest value of 0.64, and control waste sample A has the lowest value of 0.51. Although the R2 value of waste sample F is 0.29, which is a poor curve fit for the data, for the blended waste substates, the R^2 value of G is 0.81, which indicates that the model curve fits the experimental data with precision.

3.4.3. Modeling Cumulative Biogas Yield Kinetics Using a Logistic Function Model

The biogas yields fitted by the logistic equation are shown in Table 3. As can be observed from the data, the highest Rmax value of 8.53 was obtained for control sample B, while the lowest value of 2.19 was obtained for control sample A, showing that the maximum and lowest fitting R_{max} values were obtained in the same waste sample group (control). The highest A value of 31.87 was obtained from pretreated waste sample D, while blended waste sample G had the lowest value of 4.37. The lag phase (λ) values show that blended waste sample G has the highest value of 1.36, and blended waste sample H has the lowest value of -0.06. The highest coefficient of determination (R^2) is 0.82 for blended waste sample G, which shows that the model curve fits the experimental data accurately, and in the case of blended waste sample F, the R^2 value is 0.38, showing a poor curve fit for the data.

Overall, the coefficient of determination (R2) measures how well the experimental data are fitted by the predictive model curve (Table 2). The modified Richards and logistic function model curves fit the data slightly better than the modified Gompertz model, despite the three models having desirably high R2 values. The logistic model showed the highest R2 value of 0.82 and a greater average value of 0.59, indicating that the logistic model curve accurately fits the experimental data. This is followed by the modified Richards model (0.81), with an average value of 0.57, and the Gompertz model (0.79), with an average value of 0.56. Among the kinetic models employed, the logistic function model adequately fit the model-predicted and experimental results with a high coefficient of determination (R^2). The goodness-of-fit across the kinetic models in predicting daily biogas production with respect to retention times for the waste samples did not vary significantly. Although no significant difference was detected between the modified Gompertz and modified Richards models, with p values of 0.9 and 0.98, respectively, the logistic function model exhibited a significant difference, with a p value of 0.00.

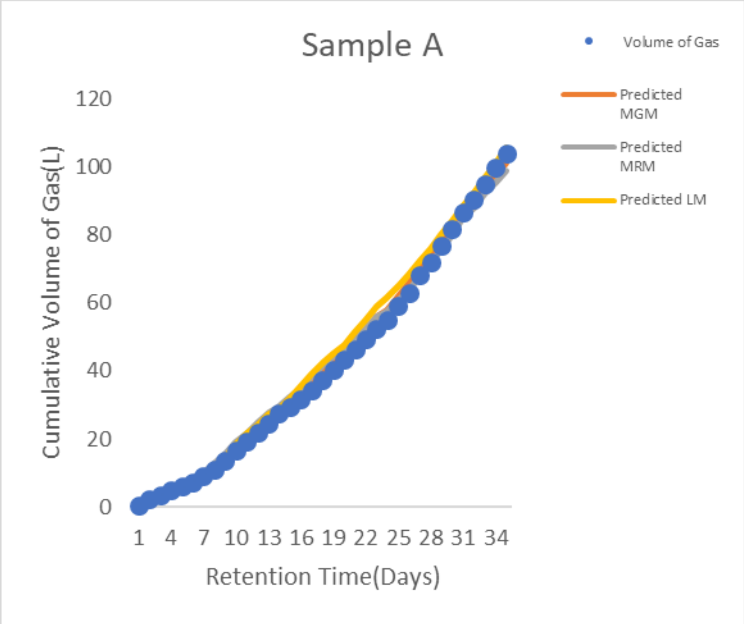
Table 3: Model fitting parameters for the samples.

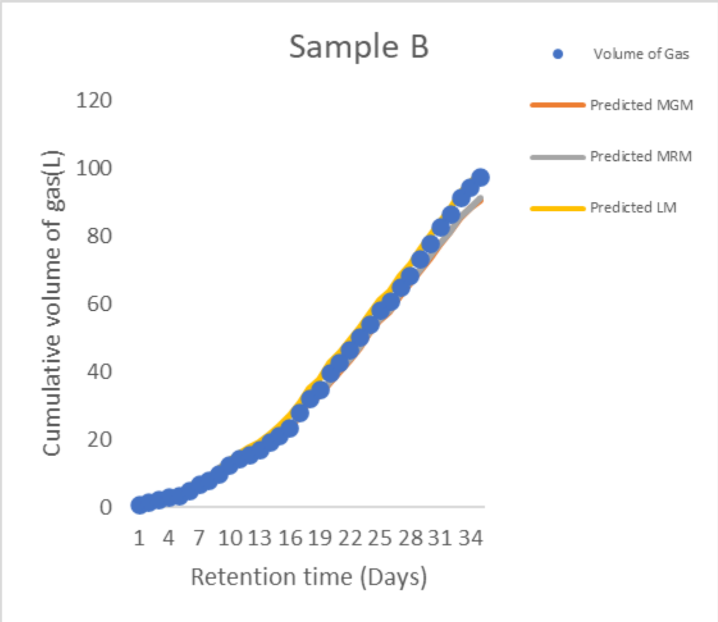
Samples	Modified Gompertz Model (MGM)				Modified Richards Model (MRM)				Logistic Model			
	A(L)	Rmax(l/day)	λ (days)	R-Squared	A(L)	Rmax(L/day)	λ (days)	R-Squared	A(L)	Rmax(L/day)	λ (days)	R-Squared
A	105.58	20.14	1.01	0.49	203.79	10.18	0.51	0.46	9.01	2.19	0.39	0.50
B	98.61	21.08	1.01	0.71	218.27	8.61	0.53	0.67	25.73	8.53	0.05	0.69
C	104.44	20.78	1.01	0.25	275.31	11.50	0.52	0.42	16.82	4.73	0.34	0.39
D	111.65	18.46	0.95	0.71	230.54	12.56	0.53	0.69	31.87	7.12	-0.18	0.67
E	97.44	14.96	1.14	0.51	206.31	15.66	0.64	0.55	15.30	4.14	0.59	0.62
F	101.08	15.39	0.99	0.42	259.32	12.06	0.52	0.29	12.06	2.48	0.30	0.38
G	96.85	11.09	1.58	0.79	233.51	11.92	0.57	0.81	4.37	2.88	1.36	0.82
H	108.35	19.56	0.94	0.62	201.01	11.14	0.52	0.74	16.50	3.12	-0.06	0.58
I	104.82	17.27	1.10	0.61	237.11	11.89	0.53	0.66	9.86	2.80	0.89	0.73

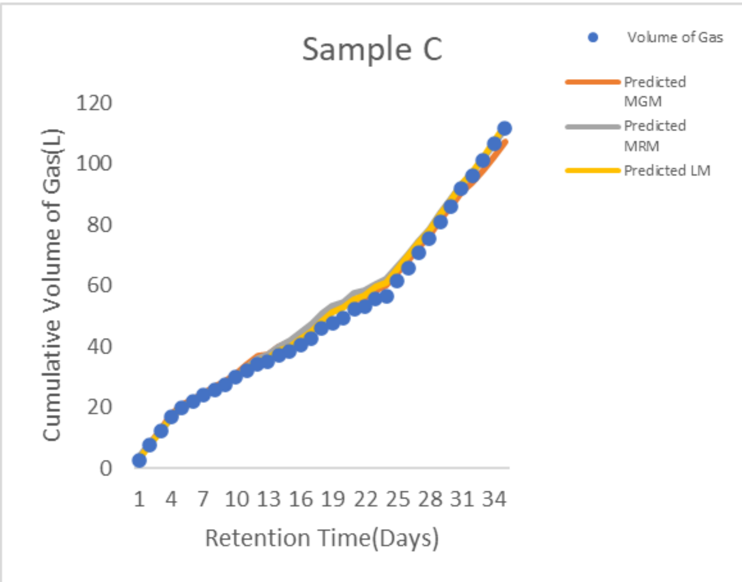
Table 4: Model fitting parameters for the samples cont'd.

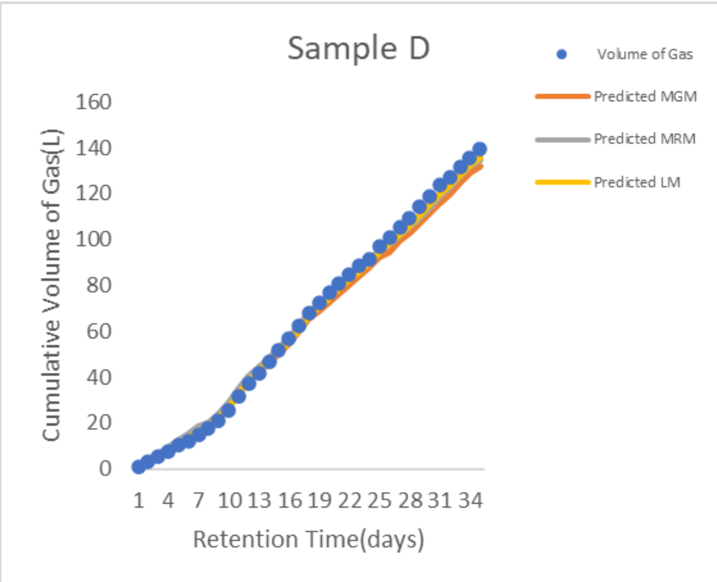
Model	**Modified Richards Model**		**Modified Gompertz Model**		**Logistic function**	
Samples	**SSE**	**RMSE**	**SSE**	**RMSE**	**SSE**	RMSE
A	49.93	1.19	47.38	1.16	52.09	1.22
B	21.54	0.78	22.55	0.80	24.53	0.84
C	81.04	1.52	76.08	1.47	77.28	1.49
D	83.05	1.54	84.21	1.55	87.03	1.58
E	54.3	1.25	51.85	1.22	53.19	1.23
F	96.4	1.66	92.33	1.62	90.11	1.60
G	81.3	1.52	79.93	1.51	84.98	1.56
H	83.5	1.54	81.06	1.52	76.43	1.48
I	99.2	1.68	92.6	1.63	96.25	1.66

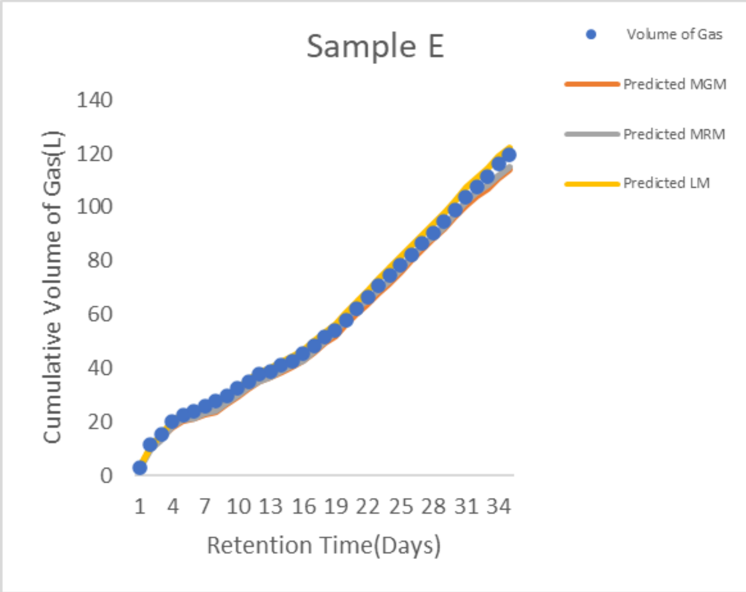
Table 4 provides model fitting parameters for various samples using three different models, the Modified Gompertz, Logistic Function and the Modified Richards Model. The performance of each model was assessed by comparing the SSE (Sum of Squared Errors) and the RMSE (Root Mean Squared Error) for samples with cow dung, cabbage, sandbox and their mixtures with ash. Comparing all models sample B shows the lowest value of SSE (21.54 to 24.53) followed by RMSE (0.78 to 0.84) perfecting the indices in the cabbage sample. The sandbox + ash sample (F) yielded the lowest SSE value of (90.11 - 96.40) and RMSE (1.60-1.66) indicating that it is the poorest fitting model. The Logistic Function Model has the lowest SSE (47.38) and RMSE (1.16), providing the best fit. Overall, the Logistic Function Model gives the overall performance in most samples as it has comparatively lower SSE and RMSE values against the rest of the models.


Sample A
Volume of Gas
Predicted MGM
Predicted MRM
Predicted LM
Cumulative Volume of Gas(L)
0
20
40
60
80
100
120
1 4 7 10 13 16 19 22 25 28 31 34
Retention Time(Days)

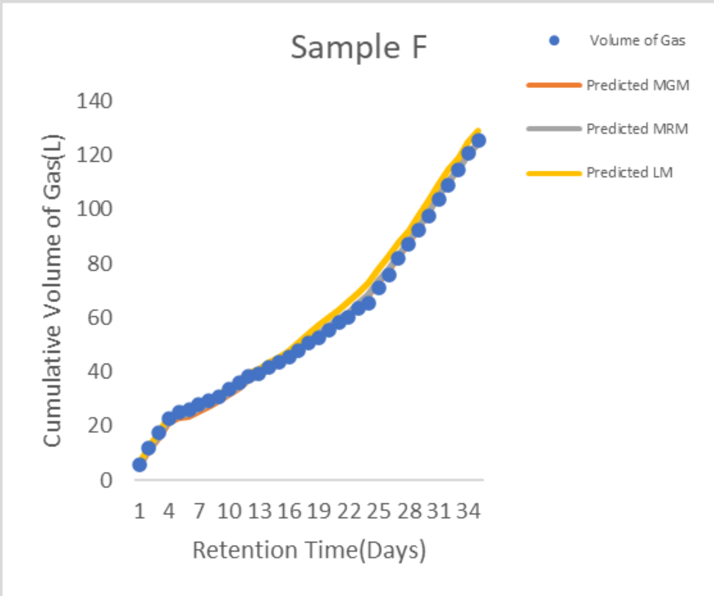

Sample B
Volume of Gas
Predicted MGM
Predicted MRM
Predicted LM
Cumulative volume of gas(L)
0
20
40
60
80
100
120
1 4 7 10 13 16 19 22 25 28 31 34
Retention time (Days)


Sample C
Volume of Gas
Predicted MGM
Predicted MRM
Predicted LM
Cumulative Volume of Gas(L)
0
20
40
60
80
100
120
1 4 7 10 13 16 19 22 25 28 31 34
Retention Time(Days)


Sample D
Volume of Gas
Predicted MGM
Predicted MRM
Predicted LM
Cumulative Volume of Gas(L)
0
20
40
60
80
100
120
140
160
1 4 7 10 13 16 19 22 25 28 31 34
Retention Time(days)



Sample E
Volume of Gas
Predicted MGM
Predicted MRM
Predicted LM
Cumulative Volume of Gas(L)
0
20
40
60
80
100
120
140
1 4 7 10 13 16 19 22 25 28 31 34
Retention Time(Days)



Sample F
Volume of Gas
Predicted MGM
Predicted MRM
Predicted LM
Cumulative Volume of Gas(L)
0
20
40
60
80
100
120
140
1 4 7 10 13 16 19 22 25 28 31 34
Retention Time(Days)

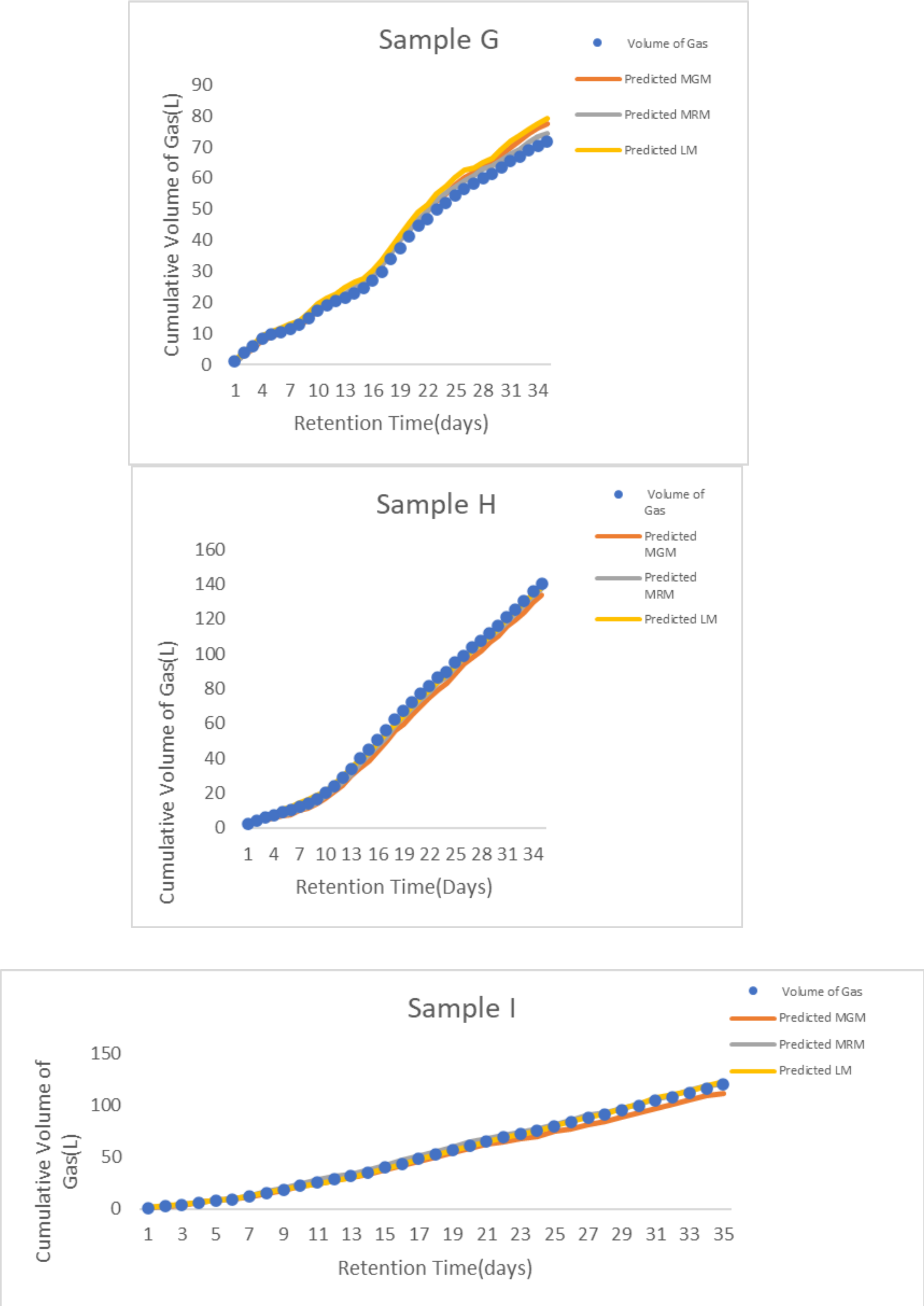


Figure 3: Kinetic model Prediction plots of the cumulative biogas yield fitted with Modified Richards, Modified Gompertz and Logistic function model for Sample A-I.

3.3.4. Machine Learning Model

The performance metrics of a machine learning model that predicts the production of biogas through anaerobic codigestion of cow dung, leaf litter from sandbox and cabbage waste are shown in Figure 4, 5 and 6. The average squared difference between the actual and predicted values of biogas production is measured by the XGBoost model mean square error (MSE) with an average squared error of 0.5518, the predictions of the XGBoost model are relatively close to the actual values, as indicated by the MSE. In general, better model performance is indicated by lower MSE value. The coefficient of determination, or R2 value, shows how well the independent variables account for the variability in biogas production. With an R2 value of 0.71, the model accounts for 71% of the variation in biogas production which suggests that the XGBoost model has an acceptable degree of prediction accuracy. Even though the XGBoost model performed well, its accuracy was not as high as that of the traditional kinetic models (logistic function, modified Richards, and modified Gompertz) used in this study, whose R2 values ranged from 0.79 to 0.82 respectively.

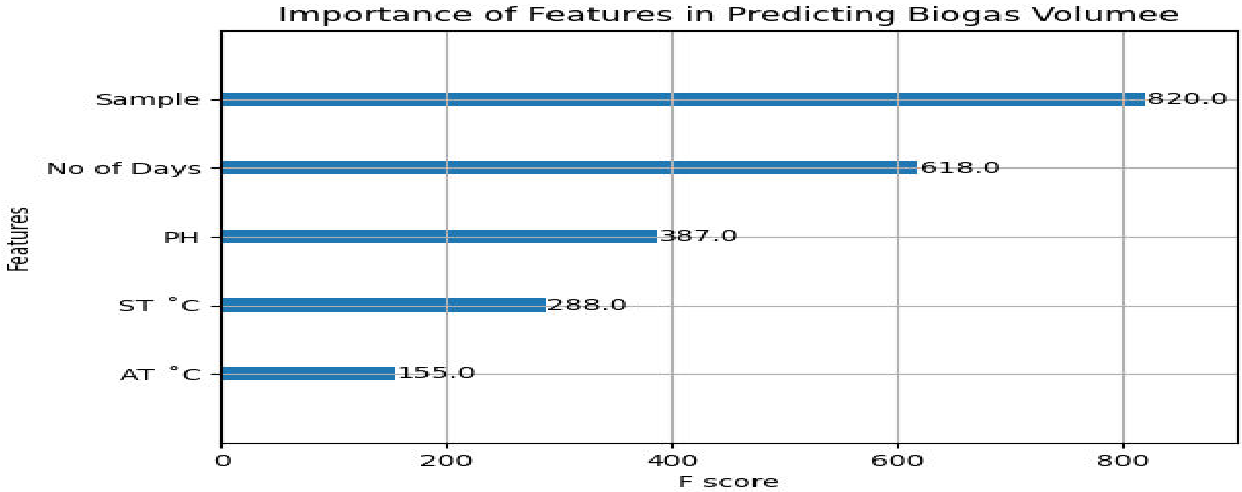


Figure 4: Mutual information gains value of the input variables.

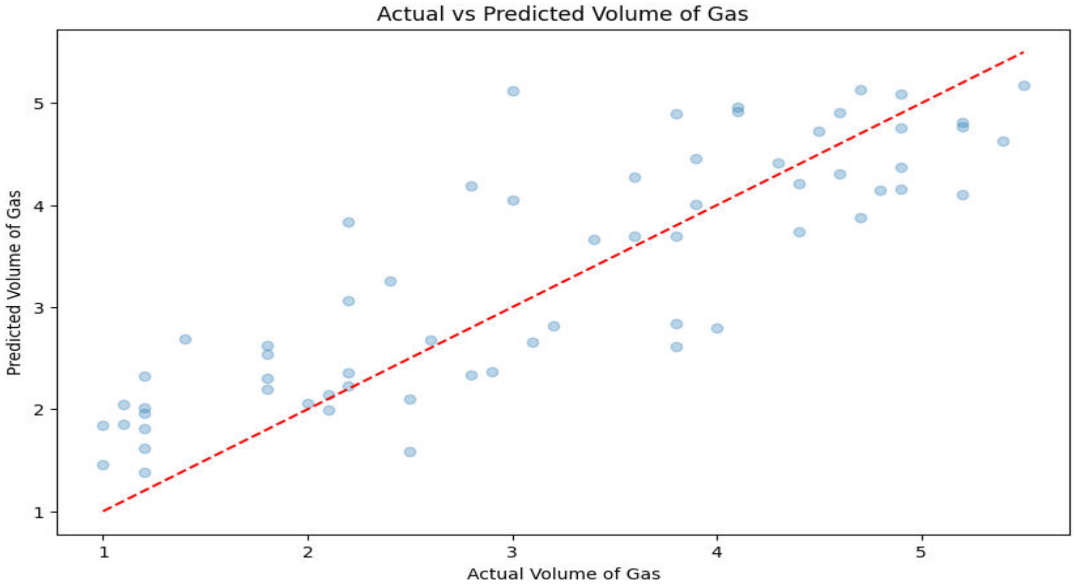


Figure 5: Actual vs predicted values for XGBoost

3.3.3 Performance comparison of the Models

When comparing models, the R2 values and ft statistics of the kinetic and machine learning models across different test scenarios need to be considered (36). These metrics reveal the degree to which the models accurately predict the data and the degree to which they capture data variability. For the kinetic models used in this study, the logistic model showed the highest R2 value of 0.82 and a greater average R2 value of 0.59, indicating that the logistic model curve accurately fits the experimental data. This is followed by the modified Richards model, with an R2 value of 0.81 and an average R2 value of 0.57, and the Gompertz model, with an R2 value of 0.79 and an average value of 0.56. The logistic function model accounts for a sizable portion of the variance in the observed data based on the R2 values. The XGBoost model exhibited a low RMSE value of 0.5518 and a high R2 value of 0.71, demonstrating its great degree of accuracy. Finally, the logistic function model was used to compare the kinetic and machine learning models. Based on a performance comparison using R^2 values, it can be observed that the logistic function model (R^2 = 0.82) accounts for 82% of the variance and is the most successful in biogas yield prediction. The Richards (R2 = 0.81) and Gompertz (R2 = 0.80) models also provided strong fits that were close to those of the logistic model. Overall, for this dataset, the XGBoost model (R2 = 0.71) is a useful tool for predicting biogas production, providing good but not optimal accuracy compared to traditional kinetic models. In this study, the traditional kinetic models predicted the biogas yield better than the XGBoost model.

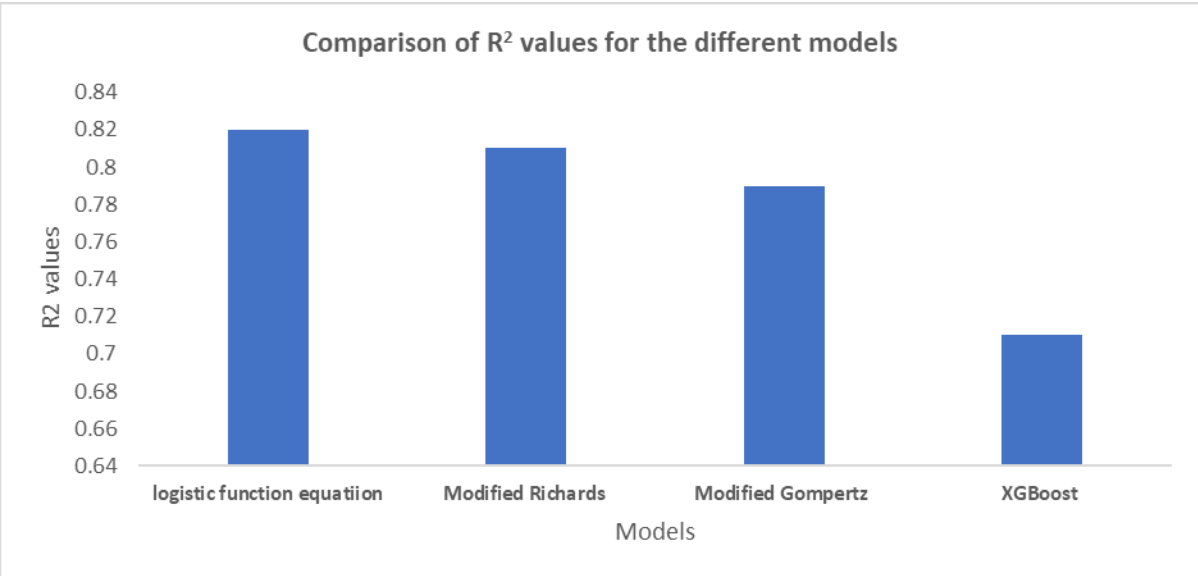


Figure 6: Comparison of the coefficient of determination values for kinetic and machine learning model.

4.0. Discussion

Animal and plant wastes show potential for large-scale biogas generation and are attractive sources of feedstock for anaerobic digestion. The physiochemical properties of the materials used in this research, with respect to minimum to maximum values, are TS (0.25-3.09%) and moisture content (95.11-97.31%), which are lower than the findings obtained for (37), who reported TS of 9% and 9.1%, respectively, for ideal biogas production. Furthermore, (38) recorded a higher methane output for anaerobic digestion with lower TS. This discrepancy was attributed to the various moisture content ranges examined. Thus, the lack of water, which can occur at a greater TS content in the digester can affect the ability of methanogens to bioconvert acids to biomethane. The volatile solids (VS) content obtained for waste samples in this paper (0.21-2.63%) is lower than the VS content of 83.74% reported by (39). This quantity is adequate for biogas production and anaerobic digestion. The quality and biogas yield of digester feeds are significantly influenced by their close composition. The difficulty of fully hydrolyzing the cellulose, hemicellulose, and lignocellulosic components of crop residue and other plant wastes makes them less easily digested than animal wastes. (40). Since no single waste is likely to be suitable for use as biomass to produce biogas on its own, the characteristics of the feedstock are crucial for the quality (methane content) and overall yield of the biogas (41). Optimizing feedstock properties and other digester operational parameters is necessary to increase the efficiency of biogas generation and maximize microbial activity in anaerobic digestion. This was accomplished using co-substrate fermentation and digestion (42). The biogas generation of the waste samples began on day 1 despite its low volume and lack of flammability. The C/N ratios for non-pretreated waste samples (A-C) (21.48-47.07%), pretreated waste samples (D-F) (21.73-46.00%) and blended waste samples (G-I) (26.48-34.69%) are greater than the suggested range of C:N ratios of 25-30:1, which is ideal for biogas production (43). One of the crucial factors to be considered in anaerobic digestion is the carbon to nitrogen (C/N) ratio. There will not be any further reactions on the substrate's residual carbon content when the C/N ratio is very high since the methanogenic bacteria quickly consume nitrogen to satisfy their protein requirements. The maximum methane potential is reached with high C/N ratios. On the other hand, a low C/N ratio will result in an excessive release of nitrogen from microbial metabolism, which will subsequently build up as ammonia (44).

The Gas chromatography results from this study revealed that sample I which represents the digester containing cow dung, cabbage and sandbox blended in ratio (60:20:20) produced the highest quantity of methane (41%), follow closely by sample D (digester containing cowdung and ash) and E (digester containing cabbage and ash) (40% each). Therefore, the blended waste sample produced the highest quantity of methane followed by the pretreated waste sample then the unpretreated(control) waste sample. Therefore digester I containing cowdung, cabbage and sandbox blended in ratio (60:20:20) which produced the highest quantity of methane (41%) can be attributed to the synergistic effect of the waste blends, increased diversity of microorganisms, improved nutrient balance(C/N) ratio and reduced process limitation. Methane levels in this study (41%) were lower than those of (45), which looked at how different organic wastes affected the biogas output from carbonated soft drink sludge. In that study, waste blends were used to create flammable biogas with high methane contents. Methane (76.5%), CO2 (20.1%), and CO2 (18.4%) were present in the CS: POS, CS: SW, and CS: RH, respectively.

Nevertheless, similar studies also showed that codigestion of food wastes with livestock manure, such as sewage sludge, cow dung, and effluent, increases biogas yield, stability and methane concentration (46). This study also aimed to investigate the potential of traditional kinetic and machine learning algorithms in predicting biogas production from anaerobic co-digestion of lignocellulosic biomass. The results show that while biogas output may be accurately predicted using both models, traditional kinetic models performed better in terms of accuracy than the machine learning model (XGBoost). This result is in alignment with other recent research that predicted the output of biogas from anaerobic digestion using machine learning techniques. For instance, XGBoost was utilized in a 2021 study by (47) to forecast the generation of biogas from food waste, and the results showed an R2 value of 0.85. This study's comparison of kinetic and machine learning models demonstrates the predictive ability of machine learning algorithms for biogas generation which contrasts with the result obtained by this present study. Machine learning algorithms are ideally suited for predicting the generation of biogas from anaerobic digestion because they can manage complex relationships between variables and learn from large data sets. On the other hand, kinetic models can be utilized to optimize process conditions and offer a mechanistic knowledge of the anaerobic digestion process (48). These models have intrinsic limitations, even though they offer valuable insights. Machine learning algorithms (MLAs) are prone to overfitting, which can compromise a model's capacity to generalize to new datasets. The quality, representation, and biases of the dataset may have an impact on the model's performance. Overall, the results show that kinetic and machine learning models may both be used to forecast the production of biogas from the anaerobic co-digestion of lignocellulosic biomass. The results of this study can be utilized to improve anaerobic digestion's biogas yield and create more precise prediction models.

5.0. Conclusion

This study results shows that the pretreatment and blending of cabbage waste with the leaf litter of sandbox codigested cow dung enhances biogas yield. Analysis of the models used to predict biogas production during the anaerobic process yielded important insights into the advantages and disadvantages of each model and which model better predicts and fits the experimental data. The strong relationship between the experimental and model-predicted data is demonstrated by the high coefficient of determination (R^2) obtained from the models. The logistic function model's better curve fitting was found to be best suited for the simulation of the experimental process, with a high (R^2) value of 0.82. Strong predictive performance was also demonstrated by the Gompertz model (R2 = 0.80) and the Richards model (R2 = 0.81). Compared with the kinetic models, the XGBoost machine learning model's R^2 value of 0.71 indicates that it is less successful at predicting biogas yield. The results showed that the traditional kinetic models performed better than the machine learning models. The use of a kinetic model in this study has shed important light on the effectiveness and potential of anaerobic digestion (AD) as a waste-to-energy solution while contributing to the improvement of biogas production from typical organic substrates by bridging theoretical modeling with actual implementation. The use of cabbage waste, leaf litter from sandbox trees, and other waste substrates in the production of biogas will greatly reduce national energy demand, improve environmental quality by eliminating cities with nuisances that litter the streets and serve as potential sources of dangerous pathogenic microorganisms, minimize greenhouse gas emissions, rely less on fossil fuels, and advance a circular economy. This study provides useful information to stakeholders, policymakers, and operators of biogas plants who are interested in greenhouse gas reduction techniques and the concepts of the circular economy. The wastes used in this study has shown to be ideal for producing "green energy," following their high carbon-to-nitrogen ratio, which optimizes microbial activity during anaerobic digestion. The models employed in this study successfully simulated the experimental process. Although the XGBoost model shows potential, more optimizations, such as feature engineering or parameter tuning, might be necessary to increase the predictive accuracy of the model. Subsequent investigations may concentrate on augmenting the efficacy of machine learning models or investigating hybrid methodologies that merge the advantages of both kinetic and machine learning models.

Acknowledgement
The authors would like to thank Prof. C.N Madu and Prof. Godwin. E Mbamalu for their encouragement and supervision of this research work. I would want to thank all my lecturers and colleagues at the Centre of Excellence in Environmental Management and Green Energy at the University of Nigeria, Nsukka and the National Centre for Energy Research and Development (NCERD), University of Nigeria, Nsukka, who have had a positive impact on this research.

Declarations.
Ethics approval and consent to participate: Not Applicable

Consent for publication: Not applicable.

Availability of data and materials.
The datasets generated during and/or analyzed during the current study are available from the corresponding author on request.

Competing interests: The authors declare no competing interests.

Funding: Not Applicable.

Author's contribution.
All authors have significantly contributed to the development and writing of this manuscript.

References
1. *G.T.Miller and S.Spoolman (2008). Environmental Science Principles, Connections and Solutions.International student's edition. Brooks/cole. PP.304-315.*
2. *Zamorano M., (2005). "Study of the energy potential of the biogas produced by an urban waste landfill in Southern Spain", Renewable and Sustainable Energy Reviews, 1–15.*
3. *Budiyono., Syafrudin, Dwi Nugraha, W., & Hawali Abdul Matin., (2017). Production of Biogas from Organic Fruit Waste in Anaerobic Digester using Ruminant as Inoculum. MATEC Web of Conferences 156,03053.*

4. *Jianbo Lu, Lei Zhua, Guoliang Hub, Jianguo Wua (2010). Integrating animal manure-based bioenergy production with invasive species control: a case study at Tongren Pig Farm in China. biomass and bioenergy.34 821–827.*
5. *Godi, N. Y., Zhengwuvi, L. B., Salihu, A.,and Kamtu, P.,(2013). Effect of cow dung variety on biogas production. Journal of Mechanical Engineering Research. 5(1): 1-2.*
6. *Sawyerr N, Trois C, Workneh T (2019). Identification and Characterization of Potential Feedstock for Biogas Production in South Africa. J. Ecol. Eng. 20(6):103 - 116.*
7. *Lu X, Jin W, Xue S, Wang X (2017). Effects of waste sources on performance of anaerobic co-digestion of complex organic wastes: taking food waste as an example. Scientific Reports. 17(15702): 1 - 9.*
8. *Zwietering, M.H.; Jongenburger, I.; Rombouts, F.M.; van 't Riet, K. (2017). Modeling of the bacterial growth curve. Appl. Environ. Microbiol. 56, 1875–1881.*
9. *A. Ahmad, A. K. Yadav, and A. Singh(2023). "Biodiesel yield optimization from a third-generation feedstock (microalgae spirulina) using a hybrid statistical approach," International Journal of Ambient Energy, vol. 44, no. 1, pp. 1202–1213, 2023.*
10. *A. Ahmad, A. K. Yadav, and A. Singh,(2023). "Process optimization of spirulina microalgae biodiesel synthesis using RSM coupled GA technique: a performance study of a biogas powered dual-fuel engine," International journal of Environmental Science and Technology, vol. 21.*
11. *W. A. Abera (2021)."Dynamic simulation and modeling of methane production process for habesha beer wastewater treatment process using aspen plus software," American Journal of Chemical Engineering, vol. 9, pp. 91–100.*
12. *R. Rr, A. A, and Z. R,(2020). "Aspen Plus simulation of optimal biogas production in anaerobic digestion process," IOP Conference Series: Materials Science and Engineering, vol. 702.*
13. *A. Inayat, S. F. Ahmed, F. Djavanroodi, F. Al-Ali, M. Alsallani, and S. Mangoosh,(2012). "Process simulation and optimization of anaerobic Co-digestion," Frontiers in Energy Research, vol. 9, pp. 1–7.*
14. *M. K. Ibro, V. R. Ancha, D. B. Lemma, and M. Lenhart,(2023). "Enhancing biogas production from food waste and water hyacinth: effect of co-substrates and inoculum ratios," Biomass Conversion and Bioenergy, vol. 13, 2023.*
15. *D. de la Lama-Calvente, M. J. Fern´andez-Rodr´ıguez, J. Llanos, J. M. Mancilla-Leyt´on, and R. Borja,(2021). "Enhancing methane production from the invasive macroalga Rugulopteryx Okamurae through anaerobic co-digestion with olive mill solid waste: process performance and kinetic analysis," Journal of Applied Phycology, vol. 33, no. 6, pp. 4113–4124.*
16. *Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L.(2020). Interpretability of machine learning-based prediction models in healthcare. Wiley Interdiscip Rev: Data Min Knowl Discov 10(5): e1379.*
17. *Jimenez J, Latrille E, Harmand J, Robles A, Ferrer J, Gaida D, (2020). Steyer JP. Instrumentation and control of anaerobic digestion processes: a review and some research challenges. Rev Environ Sci Bio/Technol 14:615–648.*
18. *Xiang Z, Huang X, Xu Y, Chen H, Liu B, Dong W, Wang H (2024). Impact of thermal hydrolysis time on volatile fatty acids production from sludge fermentation: insights*

into dissolved organic matter transformation and microbial succession. Fuel 357:129733.

19. *Taye MM.(2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers 12(5):91.*
20. *Zhang L, Loh KC, Zhang J (2019). Enhanced biogas production from anaerobic digestion of solid organic wastes: current status and prospects. Bioresour Technol Rep 5:280–296.*
21. *Shaw KM, Poh PE, Ho YK, Chan SK, Chew IML(2022). Predicting volatile fatty acid synthesis from palm oil mill effluent on an industrial scale. Biochem Eng J 187:108671.*
22. *Byliński H, Sobecki A, Gębicki J(2019). The use of artificial neural networks and decision trees to predict the degree of odor nuisance of post-digestion sludge in the sewage treatment plant process. Sustainability 11(16):4407.*
23. *Williamson LD, Scott BE, Laxton M, Illian JB, Todd VL, Miller PI, Brookes KL(2022). Comparing distribution of harbour porpoise using generalized additive models and hierarchical Bayesian models with integrated nested laplace approximation. Ecol Model 470:110011. 2022.*
24. *Narayanan H, Luna M, Sokolov M, Arosio P, Butté A, Morbidelli M (2021). Hybrid models based on machine learning and an increasing degree of process knowledge: application to capture chromatographic step. Ind Eng Chem Res 60(29):10466–10478.*
25. *Zhang Z, Kuang L, Zhao Y, Han Z, Zhou D, Tu J, Ji X (2023). Numerical investigation of the aerodynamic and wake characteristics of a floating twin-rotor wind turbine under surge motion. Energy Convers Manag 283:116957. 2023b.*
26. *Zhu, J., Yi, Z., Fuqing, X., Yebo, L(2014).,Solid-state anaerobic co-digestion of hay and soybean processing waste for biogas production. Biores Technol 154, 240–247.*
27. *AOAC (1990). Official Methods of Analysis: Association of Analytical Chemists. (14th Ed., Washington, USA, 22209).*
28. *D. Pearson (1976). The chemical analysis of foods. (7th Edition. Churchill Livingstone. New York.) 1976, 11-12, 14- 15.*
29. *Meynell PJ (1976). Methane. Planning a digester. Prison Stable Court.Clarington, Dorset. Sochen Books, pp.3.*
30. *A. Walkey, L. A. Black, J. (1934). Soil Sci , 37: 29 – 38.*
31. *Dhamodharan, K.; Kumar, V.; Kalamdhad, A.S(2015). Effect of different livestock dungs as inoculum on food waste anaerobic digestion and its kinetics. Bioresour. Technol.180, 237–241.*
32. *Ware, A.; Power, N (2017). Modelling methane production kinetics of complex poultry slaughterhouse wastes using sigmoidal growth functions.Renew. Energy 104, 50–59.*
33. *Zwietering, M.H.; Jongenburger, I.; Rombouts, F.M.; van 't Riet, K. (1990). Modeling of the bacterial growth curve. Appl. Environ. Microbiol. 56, 1875–1881.*
34. *Mowbray M, Savage T, Wu C, Song Z, Cho BA, Del Rio-Chanona EA, Zhang D(2021). Machine learning for biochemical engineering: a review. Biochem Eng J 172:108054.*
35. *Rahman M, Chen N, Elbeltagi A, Islam MM, Alam M, Pourghasemi HR. Dewan A (2021). Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. J Environ Manag 295:113086. 2021.*

36. *Liotta F, d'Antonio G, Esposito G, Fabbricino M, Frunzo L, Hullebusch EDV, Lens PNL Pirozzi F (2014). Effect of moisture on disintegration kinetics during anaerobic digestion of complex organic substrates. Waste Manage. & Res. 32(1) 40 – 48.*
37. *Patil JH, Antony, Raj M, Gavimath CC, Hooli VR. A (2021). Comparatives Study on Anaerobic Co-Digestion of Water Hyacinth with Poultry Litter and Cow Dung. Int. J. Chem. Sc. and Appl. 2(2): 148-155.*
38. *Bolaji GE, Adebayo A.O (2018). Biogas Production Potentials of Cassava Peels Co-Digested with Yam Peels using a Batch Reactor at Mesophilic Temperature. J. Eng. Res. & Appl. 8(10) Part -3: 09-13.*
39. *Sawyerr N, Trois C, Workneh T, Okudoh V(2017). Codigestion of Animal Manure and Cassava Peel for Biogas Production in South Africa. 9th Int'l Conf. Adv. Sci. Eng. Technol.& Waste Manage. (ASETWM-17) Nov. 27-28, 2017, Parys, South Africa. Pp 165-170.*
40. *Tetteh E, Amano KOA, Denis Asante-Sackey D, Edward AE(2018). Response Surface Optimisation of Biogas Potential in Co-Digestion of Miscanthus Fuscus and Cow Dung. Int. J. Technol. 5: 944-954.*
41. *Yadvika S, Sreekrishnan TR, Kohi S (2004). Enhancement of biogas production from solid substrate using different techniques – a review. Bioresour Technol ;95:1–10.*
42. *Ganiyu OT, Oloke JK (2012). Effects of Organic Nitrogen and Carbon Supplementation on Biomethanation of Rice Bran. Fountain J. Nat. & Appl. Scs. 1(1): 25 – 30.*
43. *Esposito G, Frunzo L, Liotta F, Panico A, Pirozzi F (2012). Bio-Methane Potential Tests to Measure the Biogas Production from the Digestion and Co-Digestion of Complex Organic Substrates. The Open Environ. Eng. J. 5:1 - 8.*
44. *Ofoefule AU, Uzodinma EO (2009). "Biogas production from blends of cassava (Manihot utilissima) peels with some animal wastes. Int. J. Phy. Scs. 4 (7): 398 - 402.*
45. *Ofoefule, Akuzuo, U., Nwankwo, Joseph, I, I. And Cynthia, N., (2010). Biogas production from paper waste and its blend with cowdung. Adv. Appl. Sci. Res., 6: 1-8.*
46. *Zhang W, He Y, Yang S (2022). A multi-step probability density prediction model based on Gaussian approximation of quantiles for offshore wind power. Renew Energy 202:992–1011. 2023a.*
47. *Chen, (2021) 'Machine learning approaches for predicting biogas production from anaerobic digestion: A review', Renewable and Sustainable Energy Reviews, 145, p.*
48. *Demir S, Sahin EK (2023) An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. Neural Comput Appl 35(4):3173–3190.*