

Mathematical Modeling and Machine Learning Approaches for Biogas Production from Anaerobic Digestion: A Review

Osama H. Galal,^a Mahmoud M. Abdel-Daiem ^{b,c,*} Hani S. Alharbi,^c and Noha Said ^b

Anaerobic digestion (AD) is a widely recognized method for converting organic waste into biogas, offering a sustainable solution for both waste management and renewable energy generation. This review critically examines recent advancements in mathematical modeling and machine learning (ML) approaches applied to biogas production from AD processes. The study categorizes the models into daily and cumulative biogas production models, kinetic models, and hybrid AI-based predictive techniques. Special attention is given to the comparative evaluation of first-order kinetics, modified Gompertz, and Chen-Hashimoto models, highlighting their applicability and limitations. Furthermore, the integration of artificial neural networks (ANNs) and other ML algorithms is discussed in the context of optimizing biogas yield, understanding system dynamics, and reducing operational uncertainties. Research gaps are identified, including the need for more robust hybrid models, real-time monitoring systems, and studies under diverse feedstock and environmental conditions. The review emphasizes that combining traditional modeling with intelligent systems offers a powerful approach to enhancing AD performance and scaling sustainable energy solutions.

DOI: 10.15376/biores.20.4.Galal

Keywords: Mathematical modeling; Anaerobic digestion; Multi-dimensional models; Machine learning; Parameters uncertainty; Renewable energy

Contact information: a: Engineering Mathematics and Physics Department, College of Engineering, Fayoum University, 63514, Fayoum, Egypt; b: Environmental Engineering Department, Faculty of Engineering, Zagazig University, 44519, Zagazig, Egypt; c: Civil Engineering Department, College of Engineering, Shaqra University, 11911, Duwadmi, Riyadh, Saudi Arabia;

* Corresponding author: mmabdeldaiem@eng.zu.edu.eg

INTRODUCTION

Anaerobic digestion (AD) converts organic waste into biogas (primarily CH₄ and CO₂), delivering simultaneous sanitation and energy recovery, and aligning with circular economy goals (Jameel *et al.* 2024; Alengebawy *et al.* 2024). Across common feedstocks, including sewage sludge, agricultural residues, food waste, and manure co-digestion, as well as process tuning (temperature, pH, organic loading rate (OLR), hydraulic retention time (HRT)), it is possible to enhance yields and stability when the system is properly managed (Adnane *et al.* 2024; Liu *et al.* 2025). Mathematical modeling has emerged as a critical tool in understanding, simulating, and scaling up AD processes across various substrates, including sewage sludge, agricultural residues, and municipal solid waste (Abdel Daiem *et al.* 2021). Recent advancements in kinetic and mechanistic modeling

approaches have significantly improved the predictive accuracy and control of AD systems (See Table 1).

Unlike mathematical models, machine learning (ML) learns patterns from data, enabling flexible prediction and optimization of biogas production. In recent years, the application of ML in renewable energy has gained significant traction, particularly in modelling complex biological processes such as AD for biogas production (Najafi and Ardabili 2018; Beltramo *et al.* 2019; Abdel Daiem *et al.* 2021; Cruz *et al.* 2023; Komarysta *et al.* 2023; Shindell *et al.* 2024; Zhu *et al.* 2025). The nonlinear and dynamic nature of biogas production processes makes conventional modelling approaches less effective. In contrast, artificial neural networks (ANNs) offer high adaptability, pattern recognition, and learning capabilities, making them well-suited for predicting biogas yields from various organic feedstocks (Abdel Daiem *et al.* 2021). This is especially relevant in the context of sewage sludge and biomass residues, which vary in composition and behaviour during digestion. The integration of ANN into biogas research represents a promising direction for optimizing system performance and enhancing energy recovery, aligning with global sustainability and waste-to-energy initiatives.

The ML techniques have become promising alternatives and complement the traditional mathematical models discussed in this paper, especially for dealing with AD processes' non-linear, dynamic, and uncertain characteristics. Unlike deterministic models, such as the modified Gompertz or logistic equations, which depend on specific kinetic assumptions and can have difficulty handling variable feedstocks or operational conditions (Roberts *et al.* 2023; Ling *et al.* 2024), ML methods are data-driven and capable of capturing complex patterns from high-dimensional inputs without predefined mechanisms (Ling *et al.* 2024). This makes them suitable for predicting biogas yields, optimizing co-digestion ratios, estimating uncertain parameters, and supporting monitoring of real-time (models that continuously update predictions and provide actionable outputs during ongoing AD plant operation using live SCADA data streams) in multi-dimensional AD systems (Asadi and McPhedran 2021). Recent studies (2019 to 2025) have utilized ML algorithms, such as ANNs (Cruz *et al.* 2023; Komarysta *et al.* 2023), random forests (RF), support vector machines (SVM), and deep learning models (LSTM) for AD, often comparing their performance favourably to traditional models (Yildirim and Ozkaya 2023). These approaches address research gaps, such as incorporating parameter uncertainty through probabilistic predictions and extending to multi-dimensional inputs *via* feature engineering and hybrid models (Sappl *et al.* 2023).

This review article presents a novel, integrative synthesis of recent advancements in the modelling and optimization of AD processes for biogas production, focusing on the convergence of mathematical modelling and ML techniques. While prior reviews have addressed modelling frameworks in isolation, this work uniquely bridges deterministic kinetic models with data-driven approaches, offering a comparative assessment of their capabilities, limitations, and future trajectories. Thus, the purpose of this study is to evaluate the predictive performance of widely used mathematical models, such as first-order kinetics, modified Gompertz, and Chen–Hashimoto models, alongside ANN and hybrid ML models, including random forests, SVMs, and deep learning architectures. The review highlights how ML algorithms increasingly address the nonlinearities and uncertainties inherent in AD systems, particularly for complex substrates such as sewage sludge, food waste, and co-digested residues. Moreover, it outlines gaps in current modelling practices, including limited real-time adaptability, feature selection, and

parameter sensitivity analysis. It proposes future extensions involving hybrid modelling frameworks and smart digesters. Through integrating insights across computational and engineering domains, this review advances a comprehensive understanding of biogas system optimization, promoting scalable and intelligent waste-to-energy solutions aligned with sustainability goals.

Table 1. Summary of Key Studies on Anaerobic Co-Digestion, Highlighting Substrates, Operating Conditions, Biogas/Methane Yields, and Kinetic/Statistical Model Performance

Substrate(s)	Co-Digestion Ratio / Conditions	Key Findings	Kinetic / Statistical Model	R ² / Accuracy	Study / Reference
Sewage sludge, agricultural residues, municipal solid waste	N/A	Highlighted the importance of mathematical modeling in understanding and scaling AD processes	General kinetic & mechanistic modeling	N/A	Abdel Daïem <i>et al.</i> 2021
Food waste + groundnut shells	50:50:00	32.28% increase in biomethane yield vs mono-digestion	Gompertz, Modified Gompertz, Schunte	0.97–0.99	Olatunji <i>et al.</i> 2025
Sewage sludge + wheat husk	N/A	Chen-Hashimoto showed superior predictive accuracy and robustness; the ultimate biogas potential most sensitive	First-order, Modified Gompertz, Chen-Hashimoto	Highest accuracy (not specified)	Tiwari <i>et al.</i> 2025
Food waste + sewage sludge + poultry litter	2:1:1, ambient, summer	Highest biogas output 640 L/kgVS, 65% CH ₄ ; temperature critical for seasonal optimization	Gompertz, First-order	Confirmed experimental results	Lohani <i>et al.</i> 2025
Sewage sludge + agro-industrial fruit & vegetable waste	70:30, mesophilic	Optimal methane yield 542.88 mL CH ₄ /g VS; improved biodegradability & pathogen reduction	Cone model	>0.98	Pulgarín-Muñoz <i>et al.</i> 2025

Novelty and Distinctiveness

This review differs from others in the following respects:

1. Classical vs. ML Modeling: The review compares classical kinetic models (first-order, Gompertz, Chen-Hashimoto) with ML approaches across daily-rate and cumulative-yield frameworks. Findings highlight where traditional kinetics remain useful and where ML achieves better predictive accuracy.

2. Multidimensional Kinetic Framework: A multidimensional framework is introduced, treating kinetic parameters as functions of operational variables such as temperature and mixing ratio. This enables response surfaces that support scenario mapping and process optimization, which are rarely discussed in prior AD reviews.

3. Stochastic Parameter Uncertainty: Kinetic parameters are modeled as random variables using stochastic methods, including Karhunen–Loève expansions. This generates probabilistic biogas trajectories with means, quantiles, and variances, offering a risk-aware alternative to point estimates.

4. ML Applications: Advanced ML methods (LSTM, TFT, SHAP) are synthesized for forecasting, optimization, and stability control in AD systems. Their performance is benchmarked against kinetic baselines, emphasizing practical deployment guidance.

5. Hybrid Mechanistic–ML Framework: A hybrid framework integrates mechanistic kinetics with ML residual learning, enabling IoT-based smart digesters. Recommendations for dataset standardization and cross-validation strengthen pathways toward real-world implementation.

6. Up-to-Date Coverage: The article emphasizes the most recent advances (2023–2025), including emerging algorithms (LSTM, hybrid ML models) and updated kinetic formulations, which have not been synthesized elsewhere.

MATHEMATICAL MODELS

Daily Biogas Production Models

Table 2 identifies the parameters and their goodness of fit using daily biogas production models (linear, exponential, and Gaussian models). Among the case studies summarized in Table 2, exponential daily-rate functions consistently achieved the highest goodness-of-fit on both rising and falling limbs ($R^2 \approx 0.960$ – 0.999), followed by Gaussian profiles when a single, roughly symmetric peak was present ($R^2 \approx 0.95$). Linear fits were acceptable mainly for descending limbs or simple substrates but tended to underfit peak regions and onset dynamics. Practically, daily-rate forecasting should default to exponential models unless there is strong peak asymmetry or multi-modal behavior; linear fits are best used for quick, conservative screening.

Exponential daily-rate models are the most reliable across substrates and digestion stages, with Gaussian profiles competitive when production exhibits a single, symmetric peak; linear fits chiefly succeed on descending limbs and under simple matrices. Lo *et al.* (2010) and Latinwo and Agarry (2015) illustrate this pattern: exponential fits track both rise and fall with the highest R^2 , Gaussian captures unimodal curves, and linear underestimates peak curvature. Practically, investigators often default to using exponential approaches for short-horizon forecasting and reserve Gaussian approaches for pronounced single-peak shapes; they use linear fitting only for conservative trend screening.

Table 2. Daily Biogas Production Models (Linear, Exponential, and Gaussian) Applied to Diverse Feedstock, with Key Parameters and R^2 values. All models Show Strong Predictive Accuracy ($R^2 > 0.90$), with Exponential Models Excelling in Dynamic Phases and Gaussian Models Performing Well for Heterogeneous Wastes

Model	Main Target	Model Parameters	Goodness of Fit R^2	Sub-target (if it Exists)	References
Linear	Organic fraction of MSW co-digested with MSWI ashes	$a = 0.8360, b = 0.1641$ $a = 16.7085, b = -0.4283$	The best $R^2 = 0.9579$ (for the ascending limb) The best $R^2 = 0.9211$ (for the descending limb)	FA/MSW: 10 g L^{-1} bioreactor BA/MSW: 100 g L^{-1} bioreactor	Lo <i>et al.</i> 2010
	Cow dung only	N/A	$R^2 = 0.8850$ (Ascending limb) $R^2 = 0.9950$ (Descending limb)	—	Latinwo and Agarry 2015
	Mixture of cow dung and plantain peels	N/A	$R^2 = 0.8790$ (Ascending limb) $R^2 = 0.9970$ (Descending limb)		
Exponential	Landfill gas generation of municipal solid waste after mechanical-biological treatment	Eight weeks, rising limb $a = 0, b = e^{0.0113}, c = 0.0803$ Eight weeks, falling limb $a = 0, b = e^{0.0066}, c = -0.0348$ 15 weeks, rising limb $a = 0, b = e^{0.0108}, c = 0.0773$ 15 weeks, falling limb $a = 0, b = e^{0.0061}, c = -0.0347$	$R^2 = 0.84$ $R^2 = 0.90$ $R^2 = 0.81$ $R^2 = 0.95$	—	De Giannis <i>et al.</i> 2009
	Organic fraction of MSW co-digested with MSWI ashes	Rising limb $a = 20160, b = 20160, c = 8.135 \times 10^{-6}$ Falling limb $a = 0.0047, b = 172.65, c = -0.0936$	$R^2 = 0.9579$ $R^2 = 0.9288$	FA/MSW: 10 g L^{-1} FA/MSW: 10 g L^{-1}	Lo <i>et al.</i> 2010
	Cow dung only	N/A	$R^2 = 0.9988$ (rising limb) $R^2 = 0.9969$ (falling limb)	—	Latinwo and Agarry 2015
	Mixture of cow dung and plantain peels	N/A	$R^2 = 0.9951$ (rising limb) $R^2 = 0.9969$ (falling limb)		

Model	Main Target	Model Parameters	Goodness of Fit R^2	Sub-target (if it Exists)	References
Gaussian Model	Organic fraction of MSW co-digested with MSWI ashes	$a = 5.312, t_m = 32.68, b = 3.05696$	The best $R^2 = 0.9486$	FA/MSW: 20 g L ⁻¹	Lo <i>et al.</i> 2010
	Heterogeneous organic and inorganic wastes with the organic fraction of municipal solid waste (OFMSW)	$a = 57, t_m = N/A, b = 17.24$	The best $R^2 = 0.95$	Garden wastes (9 – 11%VS) mixture with the OFMSW (2.5%VS)	Nielfa <i>et al.</i> 2015

Linear model

The linear model has been used to simulate and predict the daily biogas production resulting from AD (Rossi *et al.* 2022). This model assumes that the biogas production starts at an initial time, t_o with a value P_o , and then increases linearly up to a maximum value P_{max} at time t_m , after which it decreases linearly to a final value, P_f at time t_f . This plot has two limbs, an ascending limb for $t_o \leq t \leq t_m$ and a descending one for $t_m \leq t \leq t_f$. Assuming the plot similarity about the maximum value, the model equation can be written as,

$$P_{bg} = \begin{cases} a + b(t - t_o), t_o \leq t \leq t_m \\ a + b(t - t_m), t_m \leq t \leq t_f \end{cases} \quad (1)$$

where a and b are two dimensionless constants to be determined for the best fitting of the experimental data. They may be expressed as some other constants multiplied by P_o and $\left(\frac{P_{max}-P_o}{t_m-t_o}\right)$, respectively. Generally, this model is considered the simplest one, but its statistical indices are not as satisfying as those of some other models. However, this model, along with the exponential one, was shown by Lo *et al.* (2010) to have a better plot for the descending limb for the BA/MSW 100 g L⁻¹ bioreactor in the process of biogas production from the organic fraction of MSW co-digested with MSWI ashes. Moreover, this model was employed to simulate the biogas production resulting from cow dung only and cow dung with plantain peels (Latinwo and Agarry 2015). It showed an R² of 0.885 for the ascending limb and 0.995 for the descending one in the first case, while it was 0.879 and 0.997 for the ascending and descending limbs, respectively, in the second case. These correlation values are not that satisfying in comparison with the other models used in the same study. Nevertheless, linear models can still be valuable for first-cut assessments or when computational simplicity is paramount.

Exponential model

This model proposes an exponential increase in the daily biogas production with time up to an inevitable climax, and then it would decrease exponentially to zero (De Gioannis *et al.* 2009; Lo *et al.* 2010; Latinwo and Agarry 2015). The model equation is given by Eq. 2,

$$P_{bg} = a + b \exp(ct) \quad (2)$$

where a and b are two constants (L Kg⁻¹d⁻¹) while c is another constant (d⁻¹), the latter has a positive value for the rising limb and a negative value for the falling one. De Gioannis *et al.* (2009) used this model in its differential form to simulate Municipal Solid Waste (MSW) landfill gas generation after mechanical biological treatment. Their study aimed to estimate the model constants after 8 and 15 weeks. Regarding R², the model accuracy showed 0.84 and 0.90 for the rising and falling limbs, respectively, in the case of eight weeks of gasification, while it was 0.81 and 0.95 for 15 weeks. Moreover, Lo *et al.* (2010) utilized the exponential model in their work mentioned above, where the best R² values were 0.9579 and 0.9288 for the rising and falling limbs, respectively, and both were achieved in the case of FA/MSW 10 g L⁻¹. Furthermore, Latinwo and Agarry (2015) have employed this model to simulate biogas production resulting from both cow dung and cow dung activated by plantain peels, showing outstanding representation in both cases. The R²

for the ascending and descending limb was 0.9988 and 0.9969 in the first case, while 0.9951 and 0.9969 for the second.

Gaussian model

The Gaussian distribution is usually used to plot numerous natural phenomena (Simon 2002; Lo *et al.* 2010). It has also been used to describe bacterial growth, resulting in biogas production during AD. Therefore, this model and some other models for growth and decay can be used to simulate the daily production process. The Gaussian model is given as Eq. 3,

$$P_{bg} = a \exp(-(t - t_m)^2/2b^2) \quad (3)$$

where a is a constant ($L\ Kg^{-1}d^{-1}$), while t_m and b are the mean and standard deviation, respectively, in (d), this model has been investigated by Tonner *et al.* (2017) to simulate the differential effects of media, genetics, and stress on microbial population growth. Moreover, it was utilized to simulate and predict the biogas production evaluated by Lo *et al.* (2010), where the best R^2 was 0.9486 in the case of FA/MSW 20 g L^{-1} . In addition, Nielfa *et al.* (2015) used this model to simulate methane production resulting from the composition of heterogeneous organic and inorganic wastes with OFMSW. The highest R^2 was achieved in the case of a garden waste mixture with the OFMSW, where it was 0.95.

However, AD operational monitoring and management depend heavily on daily biogas output models. Data in Table 2, together with Eqs. 1 to 3, indicate that although basic models such as Gaussian, exponential, and linear can fit the ascending and descending limbs of daily production, their accuracy is strongly influenced by the substrate and process conditions. For example, the exponential model can achieve excellent fits (R^2 up to 0.9988) for certain organic fractions and waste combinations, while the linear model performs reasonably well (R^2 up to 0.96) but is often outperformed. The Gaussian model, with good fits ($R^2 = 0.95$) for heterogeneous organic wastes, also demonstrates robustness and usefulness in simulating the symmetric rise and fall of daily production rates in specific systems.

For operations, daily-rate models are most useful for short-term scheduling, diagnosing inhibition or overload patterns, and checking whether a feeding change alters rise or fall constants as expected. Exponential forms are a sensible default for forecasting both sides; Gaussian is informative when production shows a single, symmetric peak, while linear fits act as conservative trend indicators rather than control-relevant predictors. These choices help operators prioritize sampling frequency and decide if a perturbation requires adjusting the OLR or mixing strategy in the next cycle.

Linear, exponential, and Gaussian daily-rate forms implicitly assume a unimodal production curve under a stable operating regime over the day, with negligible gas-holding/back-pressure effects. In continuous or semi-batch operation, feed pulses, temperature swings, transient inhibition (*e.g.*, ammonia, sulfide, long-chain fatty acids), foaming, or mixing disruptions can create asymmetric or multi-peak profiles that a single exponential or Gaussian cannot reproduce, biasing rise/fall constants and peak timing (Lo *et al.* 2010; Altaş 2009). In such cases, segmented fits or multi-population kinetics are preferable; at minimum, re-fit pre-/post-perturbation windows and avoid extrapolating across regime shifts (Ling *et al.* 2024).

Cumulative Biogas Production Models

Table 3 summarizes cumulative-yield models and reveals a clear pattern: the modified Gompertz consistently achieves near-perfect fits across various substrates and operating conditions ($R^2 \approx 0.98$ to 1.00), often outperforming the logistic and modified-logistic models. The exponential rise-to-maximum model performs exceptionally well in landfill BMP contexts ($R^2 \approx 0.99$ to 0.996), while simple logistic models are mainly competitive for more homogeneous feedstocks (e.g., manure). In practice, A (ultimate potential) and λ (lag) are the most influential parameters in modified-Gompertz fits, emphasizing the importance of accurate estimation or uncertainty ranges.

Engineering interpretation of cumulative-yield parameters directly supports design and start-up. The ultimate potential A informs gasholder/CHP sizing and energy contracts; the lag λ frames warm-up and acclimation windows; and the maximal rate D_m or kinetic constant k links to target HRT and expected time to plateau. Sensitivity analyses around A and λ are therefore recommended before committing to co-digestion ratios or pre-treatment choices, especially where substrate supply is seasonal or heterogeneous.

Logistic kinetic model

The model assumes an exponential increase up to a maximum value and remains constant (Latinwo and Agarry 2015). It has three parameters: A , which is the biogas production potential ($L\ Kg^{-1}d^{-1}$); b , a dimensionless constant; and k , another constant (d^{-1}). Equation 4 expresses this model:

$$P_{bg} = A / (1 + b \exp(-k t)) \quad (4)$$

The modified Gompertz model most consistently attains near-perfect cumulative fits across feedstocks and operating regimes, with A (ultimate potential) and λ (lag) dominating sensitivity; exponential rise-to-maximum excels in landfill/BMP contexts; while logistic/modified-logistic forms are competitive for homogeneous manures. Lo *et al.* (2010), Nielfa *et al.* (2015), and Deepanraj *et al.* (2017) embody these trends, modified Gompertz captures lag and plateau robustly, exponential rise-to-maximum performs in mid-range, and simple logistic is adequate when variability is low. Design-wise, use A for gasholder/CHP sizing, λ for start-up windows, and D_m or K to inform HRT and time-to-plateau.

Modified logistic model

This model is based on the bacterial population growth, which leads to the biogas production during the AD process using Eq. 5 (Amleh and Al-Freihat 2025),

$$P_{bg} = A / \left(1 + \exp\left(\frac{4\mu}{A}(\lambda - t) + 2\right) \right) \quad (5)$$

where A is as defined before, μ is the maximum rate of cumulative biogas production, and λ is the lag (delay) time for the start of biogas production. This model was studied by Jafari-Sejahrood *et al.* (2019) to plot and predict the biogas production from cow manure, where its R^2 was 0.993. Moreover, the inhibitory effect of four heavy metals on the methane-producing anaerobic granular sludge was studied using the same model by Altaş (2009).

Table 3. Biogas Production Kinetic Models (Exponential, Logistic, Modified Gompertz, Modified Richards) Showing High Predictive Accuracy across Substrates, with Modified Gompertz Achieving $R^2 > 0.99$ in Most Cases

Model	Main Target	Model Parameters	Goodness of Fit R^2	Sub-target (if it Exists)	References
Exponential rise-to-maximum	The biochemical methane potential of landfilled solid waste	$A = 0.2327, k = 0.0823$	$R^2 = 0.9961$ for R1	--	Bilgili <i>et al.</i> 2009
		$A = 0.2768, k = 0.0759$	$R^2 = 0.9942$ for R2		
	Organic fraction of MSW co-digested with MSWI ashes	$A = 241.9, k = 0.0112$	$R^2 = 0.9907$	In a case-control bioreactor without ash addition	Lo <i>et al.</i> 2010
	Cow dung only	$A = 7.616 \times 10^5, k = 1.15 \times 10^{-7}$	$R^2 = 0.8543$	A in (dm ³ /gm)	Latinwo and Agarry 2015
	Mixture of cow dung and plantain peels	$A = 8.26 \times 10^5, k = 1.247 \times 10^{-7}$	$R^2 = 0.8561$		
	Heterogeneous organic and inorganic wastes with the OFMSW	$A = 140, k = 0.24$	$R^2 = 0.99$	OFMSW only	Nielfa <i>et al.</i> 2015
Logistic kinetic	Cow dung only	$A = 10.55, b = 80.15, k = 0.1249$	$R^2 = 0.9859$	--	Latinwo and Agarry 2015
	Mixture of cow dung and plantain peels	$A = 4.918, b = 55.12, k = 0.1766$	$R^2 = 0.9775$		
Modified logistic	Biogas production from cow manure	$A = 14.489, \mu = 0.326, \lambda = 12.099$	$R^2 = 0.9930$	--	Jafari-Sejahrood <i>et al.</i> 2019
	The inhibitory effect of four of these metals on the on methane-producing anaerobic granular sludge	$A = 0.65$ to $44.14, \mu = 0.12$ to $3.21, \lambda = 6.40$ to 68.66	--	--	Altas 2009
	The biogas production of food wastes co-digested with poultry manure	$A = 9764.9, \mu = 841.3, \lambda = 3.1$	$R^2 = 0.9991$	Ultrasonication (US)	Deepanraj <i>et al.</i> 2017
Modified Gompertz	Cow dung only	$A = 4.733, D_m = 0.0059, \lambda = 7.178$	$R^2 = 0.9834$	--	Latinwo and Agarry 2015
	Mixture of cow dung and plantain peels	$A = 5.660, D_m = 0.0134, \lambda = 6.11$	$R^2 = 0.9895$		
	Inhibition of heavy metals on fermentative hydrogen production by granular sludge	$A = 171$ to $10, D_m = 4.9$ to $0.1, \lambda = 4.6$ to 29.9	$R^2 > 0.95$ (in all cases)	For Zn concentrations (0 to 5000)	Li and Fang 2007

Model	Main Target	Model Parameters	Goodness of Fit R^2	Sub-target (if it Exists)	References
Modified Gompertz	The effects of ionic Cr, Cu, and Zn on the fermentative hydrogen production of sewage sludge	$A = 12.8$ to 177 , $D_m = 0.8$ to 19.2 , $\lambda = 3$ to 33.7	$R^2 = 0.9912$ to 0.9998	Investigating the Cu effect	Lin and Shei 2008
	Bio-hydrogen production from food waste and sewage sludge in the presence of aged refuse excavated from the refuse landfill	$A = 193.85$, $D_m = 94.35$, $\lambda = 15.28$	$R^2 = 0.9821$	--	Li <i>et al.</i> 2008
	The inhibitory effect of four of these metals on the on methane-producing anaerobic granular sludge	$A = 0.65$ to 38.73 , $D_m = 0.11$ to 2.94 , $\lambda = 5.57$ to 66.45	$R^2 > 0.99$ (in all cases except that for Cr)	--	Altaş 2009
	Kinetics of hydrogen production from sucrose by mixed anaerobic cultures	N/A	$R^2 = 0.994$	--	(Mu <i>et al.</i> 2007)
	Mixture of manure and rumen (ratio = 1:1), (MR 11)	$A = 172.51 \pm 6.64$, $D_m = 3.89 \pm 0.28$, $\lambda = 7.25 \pm 1.65$	$R^2 = 0.9983226$	--	(Budiyono <i>et al.</i> 2010)
	Mixture of manure and water (ratio = 1:1), (MR 11)	$A = 73.81 \pm 4.01$, $D_m = 1.74 \pm 0.13$, $\lambda = 14.75 \pm 2.87$	$R^2 = 0.9987334$		
	The co-digestion of horse and cow dung	$A = 360$, $D_m = 36.99$, $\lambda = 8.07$	$R^2 = 0.998$	Case of 75% horse dung and 25% cow dung	(Yusuf <i>et al.</i> 2011)
	Organic fraction of MSW co-digested with MSWI ashes	$A = 165.4$, $D_m = 4.507$, $\lambda = 5.67$	The best $R^2 = 0.9977$	In case of FA/MSW 10 g L^{-1}	(Lo <i>et al.</i> 2010)
	Modeling biogas production kinetics of various heavy metals exposed anaerobic fermentation process	$A = 34.18$, $D_m = 2.05$, $\lambda = 3.99$	The best $R^2 = 0.9989$	In case of Cu with concentration 500 mg L^{-1}	(Tian <i>et al.</i> 2020)
	Heterogeneous organic and inorganic wastes with OFMSW	$A = 299$, $D_m = 21.42$, $\lambda = 2.54$	The best $R^2 = 1.00$	In case of meat/fish wastes	(Nielfa <i>et al.</i> 2015)
	The biogas production of food wastes co-digested with poultry manure	$A = 8964.3$, $D_m = 712.6$, $\lambda = 2.8$	The best $R^2 = 0.9995$	In case of NT	(Deepanraj <i>et al.</i> 2017)

Model	Main Target	Model Parameters	Goodness of Fit R^2	Sub-target (if it Exists)	References
Modified Richards	The inhibitory effect of four of these metals on the on methane-producing anaerobic granular sludge	$A = 0.65$ to 44.83 , $\mu_m = 0.11$ to 2.94 , $\lambda = 5.57$ to 67.24 , $\nu - 1, 0, 1$	$R^2 > 0.99$ (in all cases except that for Cr)	--	(Altaş 2009)
	The kinetics of hydrogen production from sucrose by mixed cultures	N/A	$R^2 = 0.994$	--	Mu <i>et al.</i> 2007

These studied metals were zinc, nickel, cadmium, and chromium, where the correlation coefficient R^2 was greater than 0.99 for all metals except chromium. In addition, Mu *et al.* (2007) investigated the kinetics of hydrogen production from sucrose by mixed anaerobic cultures. They used this model, which shows an R^2 of 0.9916. Concerning the food waste, Deepanraj *et al.* (2017) studied the biogas production of food waste co-digested with poultry manure. They considered four types of digestate pre-treatment: autoclave (AC), microwave (MW), ultrasonication (US), and a no-pre-treatment case (NT). The best fit of this model for the US is where $R^2 = 0.9991$.

Exponential rise-to-maximum model

The exponential rise to maximum model describes many physical phenomena in various fields, including biology, physics, economics, and finance. The model has two parameters: A and k . The first one, A , is the biogas production potential ($L Kg^{-1}d^{-1}$), while k is another constant (d^{-1}), and is given as the following equation (Bilgili *et al.* 2009):

$$P_{bg} = A (1 - \exp(-k t)) \quad (6)$$

Bilgili *et al.* (2009) investigated the exponential rise to maximum model for predicting the biochemical methane potential of landfilled solid waste. They designed two landfill reactors; R1 operated with leachate recirculation and R² without it. The best R^2 was 0.9961 for R1 and 0.9942 for R² after 400 days of operation for both reactors. For the same problem treated above by Lo *et al.* (2010), this model was applied, where the best R^2 was 0.9907 in the case of the control bioreactor without ash addition. Moreover, Latinwo and Agarry (2015) studied it for the two instances of cow dung only and cow dung with plantain peels where it showed less R^2 of 0.9907 in the first case and 0.8543 for the second case.

Gompertz model

The Gompertz model equation contains three constants, A , b , and c . The constant A is the biogas production potential ($L Kg^{-1}d^{-1}$), while b is a dimensionless constant, and c is another constant in (d^{-1}) (Zwietering *et al.* 1990; Mueller *et al.* 1995; Lo *et al.* 2010; Peleg and Corradini 2011):

$$P_{bg} = A \exp(-\exp(b - ct)) \quad (7)$$

Modified Gompertz model

The modified Gompertz model is one of the most notable models and presented by Eq. 8 (Zwietering *et al.* 1990; Li and Fang 2007; Budiyo *et al.* 2010; Lo *et al.* 2010):

$$P_{bg} = A \exp\left(-\exp\left(\frac{D_m e}{A}(\lambda - t) + 1\right)\right) \quad (8)$$

This model equation has the constant A as defined before, D_m is the maximal daily biogas production rate ($L Kg^{-1}d^{-1}$), λ is the lag phase (d), and e is Euler's number. This model was extensively applied in many AD problems because of its high correlation. Li and Fang (2007) used this model to simulate the inhibition of H_2 production potential due to the effect of six heavy metals on the activity of a granular sludge. They calculated the model constants for different concentrations of these metals, where $R^2 > 0.95$ in all cases. Moreover, Lin and Shei (2008) studied the effects of ionic Cr, Cu, and Zn on the fermentative hydrogen production of sewage sludge. They used different dosages for each

metal and estimated the model constants and correlation in all cases. The model was nearly perfect for the experimental data, with the best R^2 values of 0.9981, 0.9998, and 0.9923 for investigating the effects. Combined with emerging data analytics, these extensions promise to bridge the gap between theoretical modelling and practical implementation in diverse operational contexts of Cr, Cu, and Zn, respectively. In addition, Altaş (2009) studied the inhibitory effect of four of these metals as mentioned above, where it was shown that R^2 was greater than 0.99 for all metals except Cr. Additionally, Tian *et al.* (2020) studied the kinetic evaluation of the biogas potential from a heavy-metal-stressed anaerobic fermentation process. The model showed good correlation for most studied metals with different concentrations, where the best R^2 was 0.9989. Furthermore, Li *et al.* (2008) investigated the enhancement of bio-hydrogen production from food waste and sewage sludge in the presence of aged refuse excavated from a refuse landfill. They applied the modified Gompertz model to plot the biogas production, which showed a relatively high correlation with R^2 of 0.9820. In another work concerning food waste, Deepanraj *et al.* (2017) used this model to simulate the four cases of digestate, as mentioned before. The best was 0.9995 in the case of NT. Moreover, Mu *et al.* (2007) used this model in the problem mentioned, showing an R^2 of 0.9940. Budiyo *et al.* (2010) predicted the biogas production rate from cattle manure. They employed this model for two substrates to investigate the effect of liquid rumen to cumulative biogas production. The first substrate consisted of 100 g manure and 100 mL rumen (MR 11), while the second one consisted of manure and water in equal weight ratio (MW 11). The biogas production from both substrates was studied, the model parameters were estimated and R^2 was 0.9983 for MR 11 and 0.9987 for MW 11. In addition, they have performed further experiments in room temperature and 38.5 °C to investigate the temperature effect on the biogas production from both substrates. Furthermore, this model has been used to plot the biogas production resulted from the co-digestion of horse and cow dung (Yusuf *et al.* 2011), where they designed five different mixtures of these dungs based on weight. The maximum biogas production potential and the best R^2 were achieved for the ratio of 75% horse dung and 25% cow dung, where R^2 was 0.998. Moreover, it was utilized to simulate and predict the biogas production evaluated by Lo *et al.* (2010), where the best R^2 was 0.9977 in case of FA/MSW 10 g L⁻¹. Furthermore, concerning the MSW, Nielfa *et al.* (2015) used this model to simulate the methane production as mentioned before. The best R^2 was achieved for the meat/fish mixture with the OFMSW, which was 1.00.

Furthermore, cumulative biogas production models are critical for estimating total biogas yield, which is an essential parameter for system design and economic viability. Data in Table 3, together with Equations (4–8), indicate that the Logistic, Modified Gompertz, and Exponential Rise-to-Maximum models consistently achieve high prediction accuracy (R^2 typically >0.98–0.99) across a variety of substrates, including cow manure and complex industrial wastes. The Modified Gompertz model stands out for its broad applicability and reliability, successfully fitting data even under inhibitory conditions such as heavy metal exposure. This consistently high performance underscores its prominence as the preferred kinetic model for comprehensively understanding the digestion process and predicting ultimate gas potential.

Less-used models

Some models are rarely used to plot the biogas production resulting from the AD process. This may be due to their complicated formulas, which may contain more than one

constant, and hence their difficulty in application. This group of models includes Richard, Stannard, Schunte, and their modified versions. This model used the equation of Richards's model which is represented by the following equation (Hsieh 2009),

$$P_{bg} = A \left(1 + \nu \exp(-k(\tau - t)) \right)^{(-1/\nu)} \quad (9)$$

where A , k , and τ are the biogas production potential, delay time, and a constant, respectively, while ν is an additional constant that provides more flexibility for the biogas production simulation, as shown by Eq. 10:

$$P_{bg} = A \left(1 + \nu \exp(I + \nu) \cdot \exp\left(\frac{\mu_m}{A}(I + \nu)\left(1 + \frac{I}{\nu}\right)(\lambda - t)\right) \right)^{(-1/\nu)} \quad (10)$$

Consider $\nu = m - 1$, and depending on the value of m , Eq. 10 will be reduced to: the Gompertz equation if $m \rightarrow 1$, monomolecular equation if $m = 1$, logistic equation if $m = 2$, or the von Bertalanffy if $m = 2/3$ (Fan *et al.* 2004).

This model was used by Mu *et al.* (2007) to investigate the kinetics of hydrogen production from sucrose by mixed cultures, where it showed a good correlation to the experimental data, as R^2 was 0.994. In addition, it was utilized to investigate the inhibitory effect of four heavy metals on the methane-producing anaerobic granular sludge by Altaş (2009), and R^2 was greater than 0.99.

In contrast, the Stannard model equation is represented in Eq. 11 (Zwietering *et al.* 1990),

$$P_{bg} = A \left(1 + \exp\left(-\frac{(l + kt)}{p}\right) \right)^{-p} \quad (11)$$

where l , k , and p are constants. The modified version of the Stannard equation is the same as the modified Richards' equation, which is given by Eq. 12:

$$P_{bg} = A \left(1 + \nu \exp(I + \nu) \cdot \exp\left(\frac{\mu_m}{A}(I + \nu)\left(1 + \frac{I}{\nu}\right)(\lambda - t)\right) \right)^{(-1/\nu)} \quad (12)$$

One more model that belongs to this section is the Schunte model, which is represented by Eq. 13 (Zwietering *et al.* 1990),

$$P_{bg} = \left(y_1^b + (y_2^b - y_1^b) \cdot \frac{1 - \exp(-a(t - \tau_1))}{1 - \exp(-a(\tau_2 - \tau_1))} \right)^{(1/b)} \quad (13)$$

and its modified version equation is given by Eq. 14 (Zwietering *et al.* 1990):

$$P_{bg} = \left(\mu_m \frac{(1-b)}{a} \right) \left(\frac{1 - b \exp(a\lambda + 1 - b - at)}{1-b} \right)^{(1/b)} \quad (14)$$

However, no key works were addressed in the literature using both Stannard and Schunte models and their modified versions.

Sigmoidal equations (logistic/modified Gompertz, Richards/Schunte) presume a single dominant population and constant biodegradability; co-digestion, pre-treatment, or staged hydrolysis–acidogenesis–methanogenesis often produce shoulders or long tails (multiple inflections) that a one-sigmoid curve cannot capture (Nielfa *et al.* 2015; Deepanraj *et al.* 2017). Parameter equifinality is common: λ often trades off with D_m or K

when sampling is sparse (*e.g.*, < daily), and *A* can absorb gas losses, leakage, or incomplete degassing, inflating uncertainty (Bilgili *et al.* 2009; Lo *et al.* 2010). Inhibition episodes flatten mid-slope and shift apparent lag (Altaş 2009; Tian *et al.* 2020). Mitigations include higher early-phase sampling, consistent methane normalization (STP, dry gas, per gVS), mass-balance checks, and reporting parameter CIs or Bayesian posteriors rather than single best fits.

Machine Learning Approaches

Table 4 summarizes key peer-reviewed studies emphasizing ML applications related to biogas production in AD. Table 4 details the algorithms used, data sources, performance metrics (*e.g.*, correlation coefficient (R^2) and root mean square error (RMSE)), comparisons with traditional models when available, and specific AD contexts. The studies reviewed show a shift from mechanistic to data-driven modelling, with ML consistently achieving higher accuracy (R^2 often above 0.90) than traditional kinetic models like Gompertz or logistic, especially in co-digestion scenarios involving sewage sludge, agricultural waste, or food waste (Asadi and McPhedran 2021; Ling *et al.* 2024). For example, tree-based models (RF, XGBoost) perform well in full-scale systems because they handle non-linearity and feature importance through SHAP, highlighting key variables like OLR, pH, and biomass input (Zou *et al.* 2024). Deep learning methods, such as LSTM with attention or TFT, provide probabilistic forecasts and capture long-term dependencies, addressing parameter uncertainty with quantile regression and data augmentation (Jeong *et al.* 2021). The regression-based models can be updated with new data, but they usually require explicit recalibration or retraining, whereas ML (especially online learning or adaptive ML). Hybrid techniques incorporating GA or PSO for optimization improve biogas yield and stability management (Salamattalab *et al.* 2024).

Feature engineering and data quality are crucial, as high-frequency SCADA data or derived indices (*e.g.*, VFA/ALK) improve predictions without needing extensive lab measurements (Zou *et al.* 2024). Incorporating genomics or pre-treatment data expands the input space, connecting microbial communities to performance (Adeleke *et al.* 2025). Explainable AI tools address the "black-box" issue, building trust and enabling integration with biokinetic equations for physics-informed hybrids (Gupta *et al.* 2023).

This extension fills gaps in traditional models by enabling multi-dimensional simulations, such as with variable selection networks, and managing stochastic parameters, for instance, through ensembles. Future research should focus on creating standardized datasets, facilitating real-time IoT integration, and developing hybrid ML-mechanistic frameworks to deploy robust AD systems on a large scale.

Therefore, Table 4 shows that ML methods consistently outperform traditional kinetic models in predicting biogas production, particularly for the co-digestion of diverse wastes. Tree-based models (RF, XGBoost) and deep learning approaches (LSTM, TFT) effectively handle non-linearity, probabilistic forecasting, and feature importance. Hybrid optimization techniques (GA, PSO) further improve biogas yield and process stability. High-frequency SCADA data, feature engineering, and genomics enhance prediction accuracy, while explainable AI tools (*e.g.*, SHAP) increase operational trust and allow integration with biokinetic models. These advancements fill gaps in traditional approaches and enable multi-dimensional simulations.

Table 4. Applications of ML in AD (ANN, LSTM, TFT, RF, etc.) Showing High Predictive Performance across Substrates and Processes, often Surpassing Classical Kinetic Models and Enabling Real-time Optimization and Decision Support

AD Application	ML Algorithms and Data	Key Outcomes (Metrics, Comparison)	References
Combined microbial electrolysis–AD system; real-time optimization	ANN, ANFIS; pH, oxidation-reduction potential, solids, HRT, OLR, voltage, current	$R^2 \approx 0.984$ (ANN); RMSE ≈ 188 mL day ⁻¹ (ANFIS); outperforms deterministic models	Tufaner <i>et al.</i> 2025
Lignocellulosic biomass pre-treatment; biomethane yield optimization	ANN, RF, SVM, DT; NaOH-pretreated <i>Xyris capensis</i> biomass; SHAP, k-means	RF: RMSE ≈ 3.15 , MAPE $\approx 5.75\%$; superior to kinetic models; exposure time key	Adeleke <i>et al.</i> 2025
Global AD review; perturbation detection, parameter estimation	ANN, FL, ANFIS, SVM, RF, GA, PSO; hybrid GA-ANN	$R^2 \approx 0.9986$ (GA-ANN); ML outperforms deterministic models; SCADA integration	Ling <i>et al.</i> 2024
Dry AD of kitchen waste; real-time stability control	Eight algorithms (CatBoost); data from four dry AD plants; SHAP	CatBoost: $R^2 = 0.604$ – 0.915 (biogas), 0.618 – 0.768 (VFA/ALK); soft sensors	Zou <i>et al.</i> 2024
Municipal wastewater AD; emission prediction	Non-linear regression vs. ANN, ANFIS, GA, VFA, solids, pH, flow	$R \approx 0.81$ (regression); ANN/ANFIS higher but uncertain; rates 22.0 to 28.6 m ³ min ⁻¹	Asadi and McPhedran 2021
Full-scale co-digestion; raw wastewater influence	LSTM with GA feature selection; raw wastewater, sludge data; BOD ₅ , COD, TSS	$R^2 \approx 0.84$ to 0.90 ; GA improves LSTM; HRT essential	Salamattalab <i>et al.</i> 2024
Food-waste AD; feedstock configuration	Mixup augmentation + global-attention LSTM; food-waste data	Accuracy ≈ 0.988 ; prevents overfitting; better than classical models	Geng <i>et al.</i> 2024
Time-series quantile prediction; operational planning	Temporal Fusion Transformer (TFT); high-frequency data, categorical features	MAPE $< 8\%$ (7-day); probabilistic quantiles; interpretability via attention	Sappl <i>et al.</i> 2023
WWTP biogas prediction; decision support under limited input data	Eight ML models tested; 3-model voting ensemble; full-scale WWTP sludge data; SHAP for feature importance	$R^2 = 0.778$, RMSE = 0.306 ; return sludge and influent temperature key features	Sun <i>et al.</i> 2023
Municipal co-digestion; short-term forecasting	MLP; daily lab + minute-SCADA data; 11 derived features	Adjusted $R^2 \approx 0.78$, MAPE $\approx 13.4\%$; SCADA nearly as good as lab; outperforms others	Schroer and Just 2023
Industrial AD; continuous monitoring and stability	RF, ANN, KNN, SVR, XGBoost; industrial-scale AD data	RF best: $R^2 \approx 0.924$; suggests IoT integration; tree-based superior for large datasets	Yildirim and Ozkaya 2023
Municipal wastewater AcoD; process optimization and prediction under missing data	Hybrid DL: DA-LSTM + Variable Selection Network (VSN); 2-year AcoD data	R^2 : LSTM = $0.38 \rightarrow$ DA-LSTM = $0.68 \rightarrow$ DA-LSTM-VSN = 0.76 ; VSN improves interpretability via feature importance	Jeong <i>et al.</i> 2021

Across pilot and full-scale settings, ML methods generally outperform classical kinetic baselines for short-term forecasting and stability proxies, with many studies reporting usable accuracy (often $R^2 \geq 0.80$) for operational decision-making. Tree-based ensembles (RF, XGBoost/CatBoost) are the most reliable with tabular SCADA inputs, while sequence models (LSTM/TFT) capture temporal dependencies and enable probabilistic (quantile) forecasts.

Explainability tools (SHAP/attention) consistently identify OLR, pH, temperature, and feed configuration as primary levers, and soft-sensor surrogates (e.g., VFA/ALK) enhance early warning. Practically, plants can retain modified-Gompertz-type fits for design/batch contexts and layer ML for online supervision, provided basic hygiene (outlier handling, rolling/external validation) is in place to limit overfitting and improve transferability. In practice, ANN models may overfit small datasets and fail to generalize to new substrates or variable operating conditions. Industrial deployment is further constrained by the high cost of sensors, limited data availability, and the complexity of integrating ML models into real-time control systems.

Comparative Performance of Models

To evaluate the relative strengths of different modelling approaches, a comparative analysis was conducted between mathematical models and ML by using ANN techniques applied to biogas production from co-digestion systems. This comparison assessed predictive accuracy using statistical indicators such as R^2 and RMSE. The results provide insights into the trade-offs between classical kinetic formulations and advanced data-driven methods.

Table 5 presents a comparative analysis between classical and ML models' performance metrics for predicting biogas production from co-digestion systems for the same dataset (Abdel Daiem *et al.* 2021). The comparative analysis highlights the performance of both traditional TDMMs and ANN approaches in predicting biogas production from co-digestion systems.

Among the mathematical models, the logistic kinetic formulation emerged as the most accurate, with an R^2 value of 0.9879, although all mathematical models achieved strong correlations ($R^2 > 0.97$). Nevertheless, their relatively large $RMSE > 1000$ indicates limited predictive precision when applied to dynamic and nonlinear digestion processes, underscoring their inability to capture the complexity of anaerobic digestion fully. In contrast, ANN-based approaches demonstrated considerably lower error margins ($RMSE < 10$), highlighting their superior capacity to model process variability and nonlinear relationships.

Conventional ANN training methods such as back-propagation, Marquardt–Levenberg, and ant colony optimization yielded moderate-to-high predictive accuracy (R^2 between 0.89 and 0.92); however, the integration of metaheuristic optimization techniques substantially improved performance. Specifically, the MFFNN-MFO model achieved near-perfect predictive accuracy ($R^2 = 0.9994$; $RMSE = 3.86$), clearly outperforming both conventional ANN structures and mathematical models. These findings illustrate the value of ANN models, particularly when coupled with advanced optimization algorithms, in addressing the complexity of anaerobic digestion systems and emphasize the potential of hybrid ANN–optimization frameworks as robust and reliable predictive tools for biogas production modelling.

Table 5. Comparative Analysis between Classical and ML Models' Performance Metrics for Predicting Biogas Production from Co-digestion Systems (Abdel Daïem *et al.* 2021)

Model	R ² (Correlation Coefficient)	RMSE (Root Mean Square Error)	Notes
Logistic Kinetic Model	0.9879	1079.00	Best-performing mathematical model
Exponential Rise-to-Maximum	0.9753	1540.00	Lower accuracy than logistic
Modified Gompertz	0.9815	1334.20	Good fit but less robust
Modified Logistic	0.9845	1221.00	Reliable, close to a logistic model
ANN-BP (Back Propagation)	0.8990	7.20	Acceptable, but weaker
ANN-ML (Marquardt-Levenberg)	0.9200	3.94	Improved ANN performance
ANN-ACO (Ant Colony Optimization)	0.900	7.50	Like BP
MFFNN-MFO (Proposed ANN with Moth Flame Optimization)	0.9994	3.86	Highest predictive accuracy

RESEARCH GAPS AND AVAILABLE FUTURE EXTENSIONS

Following the previous review of the mathematical modelling of the AD process, some research gaps have arisen, which can be considered promising candidates for future extensions. These gaps may be concluded as follows.

Future Extensions: Actionable Directions AD

Recent practice in AD has introduced dosing of conductive materials (*e.g.*, biochar, Fe₃O₄) to stimulate direct interspecies electron transfer (DIET) (Lo *et al.* 2010). A natural extension is to augment cumulative kinetic models (*e.g.*, Chen–Hashimoto, modified Gompertz) with a conductivity/DIET factor,

$$k_{\text{eff}} = k_0 [1 + \alpha \phi^\beta / (1 + \gamma d)] \quad (15)$$

where ϕ denotes the mass fraction of conductive additive and d a representative particle size, this formulation preserves parameter interpretability while explicitly linking additive dosing to performance. Calibration requires only routine operational data (biogas rate, temperature) supplemented with two readily available proxies: oxidation, reduction potential, and slurry conductivity. Toxic inhibition (*e.g.*, free NH₃, sulfide, LCFA) can be included multiplicatively *via* Haldane-type terms, allowing operators to evaluate when inhibitory effects offset DIET benefits and to adjust set-points accordingly (Lo *et al.* 2010).

For control-oriented applications, the process can be represented by two coupled states, hydrolysis/acidogenesis and methanogenesis, driven by measurable or soft-sensed variables. The following equations define a minimal state-space model,

$$x = [\text{S_VFA}, \text{X_meth}] \quad (16)$$

$$\dot{x} = f(x, \text{OLR}, T, \text{pH}), \quad (17)$$

with outputs including biogas flow and a soft VFA/ALK indicator derived from pH, alkalinity, and gas rate. An extended Kalman filter or moving-horizon estimator can integrate SCADA data with the soft sensor to reconstruct unmeasured states and provide (1 to 3) day acidification risk bands, enabling operators to connect forecasts to actionable levers (e.g., OLR ramping, temporary set-point changes, co-substrate throttling) (Schroer and Just 2023).

Given the prevalence of small, noisy datasets, plant-level kinetic parameters should be treated as random effects, e.g., $(A, \lambda, D_m)_j \sim N(\mu, \Sigma)$ for plant j . Partial pooling stabilizes estimates in data-scarce settings while retaining site-specific behaviour. Multi-facility fitting with leave-one-plant-out validation quantifies transferability, producing plant-specific posterior distributions with credible intervals. These can be propagated into risk-aware dashboards and sustainability KPIs (e.g., GWP per kWh, LCOE), ensuring that uncertainty is explicitly visible in decision-making (Gala 2021).

For forecasting with tree- or sequence-based ML models, embedding domain constraints is essential: monotonicity of biogas rate with OLR (within safe ranges), positive correlation of VFA with OLR, and soft penalties for mass-balance violations. Residual-based change-point detection (e.g., CUSUM, Bayesian online methods) can flag operational regime shifts (feedstock change, mixer outage). These triggers initiate lightweight re-tuning and widen predictive intervals, transforming ML from a static predictor into an operator-safe assistant (Ling *et al.* 2024).

Finally, the experimental design can be optimized to reduce the cost of BMP and pilot trials. Starting from a Latin-hypercube of feed ratios and pre-treatments, cumulative or hybrid models are fitted, and the next experimental point is selected by maximizing expected reduction in parameter uncertainty under safety constraints (e.g., VFA/ALK \leq threshold). This adaptive loop accelerates the development of decision-quality models for novel feedstock mixtures while minimizing resource requirements (Tiwari *et al.* 2025).

Incorporating Parameter Uncertainty

Estimating the model parameters is one of the main objectives when simulating biogas production over the AD process using mathematical modelling. However, if the same AD process has been repeated enough times, these parameters are expected to vary slightly from time to time. Few studies estimated the ranges of some model parameters to investigate their variations. For example, Kumar *et al.* (2004) achieved a qualitative assessment study of different methane emission data using municipal solid waste disposal sites; Danner (2006) considered the parameter uncertainty for some of the growth models; Budiyo *et al.* (2010) estimated the parameters' ranges in the modified Gompertz equation that was used to simulate the biogas production resulting from cattle manure.

Mathematically, to express these parameters more accurately, they may be described as random variables rather than deterministic ones. In such a case, a general parameter, γ , can be expressed by the following equation (Ghanem and Spanos 2003),

$$\gamma(\theta) = \gamma(1 + \varepsilon_\gamma \xi(\theta)) \quad (18)$$

where ε_γ is a controlling factor for the random part and $\xi(\theta)$ is a random variable that describes the expected uncertainty in the deterministic value of γ . The random variable $\xi(\theta)$ is a real-valued measurable function defined on a probability space as $\xi(\theta): \Omega \rightarrow R$, defined on the triple probability space (Ω, \mathcal{F}, P) . This random variable can be assigned entirely by repeating the AD process a relatively large number of times, then estimating the model parameters in each time. For each parameter, the obtained values can

then be plotted to determine its probability distribution and its statistical moments such as mean, variance, skewness, and kurtosis, so that a complete definition for this uncertain parameter will be available. The repetition of the AD process several times to determine the parameter uncertainty requires short-time processes and many reactors working simultaneously. Moreover, when the parameter uncertainty is more complicated and expected to have higher fluctuations with time, the random part can be expressed as a random process as Eq. 16,

$$\gamma(t, \theta) = \gamma(1 + \varepsilon \Phi(t; \theta)) \quad (19)$$

where $\Phi(t; \theta)$ is a second-order random process with a finite variance. This random process can be expanded into random variables multiplied by deterministic constants using K-L expansion, as Eq. 17 (Ghanem and Spanos 2003),

$$\gamma(t; \theta) = \bar{\gamma}(t) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} f_i(t) \xi_i(\theta) \quad (20)$$

where $\gamma(t)$ is the mean value of $\gamma(t; \theta)$, $\xi_i(\theta)$ is a set of uncorrelated random variables, $\lambda_i, f_i(t)$ are the eigenvalues and Eigen functions, respectively. Both $\lambda_i, f_i(t)$ can be evaluated by solving the integral Eq. 18,

$$\int_D C_{\gamma\gamma}(t_1, t_2) f_i(t_1) dt_1 = \lambda_i f_i(t_2) \quad (21)$$

where D is the time domain over which $\gamma(t; \theta)$ is defined and $t_1, t_2 \in D$.

Including these parameters, uncertainty in the model equation yields a probability distribution curve for the biogas production every time. This provides the expected value (mean), variance, different quartiles, required threshold values, and statistical moments for the biogas production. This probably gives a clear vision of the AD process. Such stochastic approaches could also incorporate sensitivity analysis to identify dominant parameters influencing biogas yield variability. This concept has been applied successfully in many fields (Galal 2013, 2021) and could provide the designers with the system's random response due to these uncertain parameters.

Multidimensional Mathematical Models

The existing models usually plot the biogas production with time under certain conditions, such as the operating temperature, mixing ratio, heavy metal concentration, *etc.* This yields a single plot for the biogas production *versus* time for each realization of these conditions. However, these models can be extended to cases with two or more dimensions. This extension to multidimensional modelling can be conducted through an equal number of curve-fitting steps. To implement this extension to a multi-dimensional case, consider a mathematical model with three parameters A , b , and k , then consider several variables such as the time, which is defined as $t \in \{t_1, t_2, \dots, t_l\}$, the mixing ratio defined as $r \in \{r_1, r_2, \dots, r_m\}$, the operating temperature defined as $T \in \{T_1, T_2, \dots, T_n\}$, and so on. First, the biogas production is plotted *versus* all time values, $t \in \{t_1, t_2, \dots, t_l\}$ at r_1 and T_1 . Then, the A , b , and k values are estimated for the best correlation with the experimental data. This will be repeated for $(r_2, T_1), \dots, (r_m, T_1)$. This yields a set of m values for each parameter varying with r . A second step of curve fitting is then performed to determine the best function with the highest correlation for each parameter in r . Using the MATLAB program (2022), many functions are available to plot the model parameters *versus* r , such as the exponential, rational, power, spline, Gaussian, Weibull, Fourier, and

sum of sine, and polynomial functions with different degrees. The function selection is based on the best curve fitting results determined by R^2 and RMSE.

In some cases, a function shows a higher R^2 value, but it is excluded if its curve does not match the expected behaviour of the experimental data in particular intervals. The previous curve fitting step will be repeated for T_2, T_3, \dots, T_n . This yields another n set of parameter values varying with T , which will be plotted, through a third step of curve fitting for the best correlation, obtaining another function for these parameters in both r and T . Finally, the model parameters are obtained as functions; i-e: $= f(r, T)$, $= g(r, T)$, and $k = h(r, T)$. Substituting these obtained functions in the model equation provides a multidimensional version of this model.

This technique was applied successfully in the case of anaerobic co-digestion process of waste activated sludge with wheat straw by Abdel Daiem *et al.* (2021). They considered time as the first variable and mixing ratio as the second one, and then biogas production was expressed as a function of both variables. This was applied to a group of models that contains a logistic kinetic model, a modified logistic model, an exponential rise-to-maximum model, and a modified Gompertz model. The introduced two-dimensional models were highly correlated to the experimental data, as the R^2 ranged from 0.9753 to 0.9879. Extending this strategy to hybrid mechanistic-machine-learning surrogates could reduce data requirements while maintaining physical interpretability. However, the same concept explained above can be applied to include more variables as inputs and be extended to all the known models.

Machine Learning

Despite the growing use of ANNs in modelling biogas production, several research gaps remain. Most existing studies are based on small-scale, laboratory, or pilot datasets, which may not accurately reflect the variability and complexity of full-scale AD systems. Moreover, many models lack external validation, limiting their generalizability across different feedstocks, climates, and reactor types. Few studies have addressed temporal dynamics in biogas production, such as seasonality or real-time operational fluctuations. Additionally, the integration of ANN with other advanced methods—such as hybrid ML models (e.g., ANN-GA, ANN-PSO), deep learning frameworks (e.g., LSTM, CNN), and Internet of Things (IoT)-based sensor networks—is still in its early stages. There is also a need for explainable AI techniques to enhance the interpretability of ANN predictions for plant operators and decision-makers. Future extensions should focus on developing adaptive, self-learning ANN models capable of real-time prediction and control and trained on diverse and large-scale datasets. Furthermore, coupling ANN models with life cycle assessment (LCA) and techno-economic analysis (TEA) tools can provide a more holistic understanding of sustainability and system performance. These improvements would significantly enhance the operational reliability, economic viability, and environmental benefits of biogas systems, especially in decentralized rural and urban applications.

Future research should focus on developing hybrid models that combine the strengths of multiple ML techniques (e.g., ANN-GA, RF-PSO, or LSTM-CNN) to enhance robustness and generalizability. There is also significant potential in integrating ML with IoT sensors for real-time monitoring, as well as with LCA or TEA to evaluate sustainability and economic performance. Additionally, explainable AI (XAI) can improve model transparency and stakeholder confidence. Through addressing these gaps, ML can play a transformative role in optimizing biogas systems, improving resource efficiency, and

supporting climate-resilient waste management strategies, particularly in countries with abundant biomass resources.

Beyond laboratory datasets, several full-scale and pilot studies demonstrate operational value from ML in live operating biogas plants. In an industrial-scale AD facility, tree-based models achieved high forecasting accuracy (RF, $R^2 \approx 0.924$), supporting routine set-point decisions (Yildirim and Ozkaya 2023). In a full-scale WWTP digester, an ensemble approach delivered usable accuracy ($R^2 = 0.778$; RMSE = 0.306), with temperature and return sludge emerging as key levers (Sun *et al.* 2023). Across four dry-AD plants processing kitchen waste, CatBoost models reached $R^2 = 0.604$ – 0.915 for biogas and enabled a VFA/ALK soft sensor to anticipate instability (Zou *et al.* 2024). A large-scale study coupling LSTM with genetic algorithms improved short-term prediction ($R^2 \approx 0.84$ – 0.90) and highlighted HRT sensitivity (Salamattalab *et al.* 2024). For municipal co-digestion, deep models with data-augmentation and variable-selection networks increased robustness under missing data (e.g., LSTM \rightarrow DA-LSTM-VSN, R^2 from 0.38 to 0.76), clarifying driver importance for operators (Jeong *et al.* 2021). Similarly, feature-engineered MLPs using minute-rate SCADA achieved an adjusted $R^2 \approx$ of 0.78 (MAPE \approx 13.4%), showing that soft-sensor surrogates can approach lab-assisted baselines (Schroer and Just 2023).

From Prediction to Sustainability Metrics (LCA/TEA)

Linking ML outputs to sustainability assessment enhances the relevance of predictive modelling in decision-making by translating results into policy and financial metrics. In practice, probabilistic forecasts of methane production rates, $r_{\text{CH}_4}(t)$, and biogas composition can be transformed into environmental and economic key performance indicators (KPIs), such as global warming potential (GWP, expressed as kg CO₂-eq per kWh delivered) and the levelized cost of energy/biogas (LCOE/LCBG) (Said *et al.* 2020). By defining a clear functional unit (e.g., “per kWh of electricity exported” or “per tonne VS fed”) and system boundary, ML predictions can be mapped to life-cycle inventory flows (electricity and heat generated through CHP efficiency, auxiliary energy for heating and mixing, flaring episodes or CH₄ slip, digestate mass and nutrient proxies) as well as to financial cash flows (CAPEX annualization; OPEX for energy, chemicals, and labour; tipping fees; and revenues from energy and fertilizer products). The resulting KPIs are computed through straightforward transformations of predicted flows.

This coupling enables direct scenario testing on operational levers identified by explainable ML (e.g., organic loading rate, hydraulic retention time, temperature set-point, or co-substrate ratio). Operators can explore feasible parameter sets, propagate forecast uncertainty through quantile or bootstrap ensembles to generate 5 to 95% confidence bands for GWP and LCOE, and then identify Pareto-efficient operating points (e.g., minimizing GWP while keeping LCOE below a defined threshold). Additional credits and burdens, such as displacement of grid electricity, heat recovery, avoided landfill emissions, or nutrient substitution from digestate, can be incorporated modularly, provided assumptions and units are transparently reported for transferability.

Field-Scale Evidence and Engineering Implications

From a model-selection perspective, cumulative-yield kinetics remain the most practical option when only batch/BMP tests or limited monitoring data are available. The modified Gompertz typically provides the most accurate fit across substrates, with parameters A (ultimate potential) and λ (lag) being highly sensitive and directly guiding

gasholder sizing and start-up expectations; the exponential rise-to-maximum is particularly effective in landfill/BMP contexts (Bilgili *et al.* 2009; Lo *et al.* 2010; Latinwo and Agarry 2015; Nielfa *et al.* 2015). When digester operation is affected by syntrophic interactions or direct interspecies electron transfer, Chen-Hashimoto-type models can outperform simpler kinetics and should be considered during scenario screening (Li *et al.* 2018).

In continuously fed plants with SCADA data, forecasting and stability control benefit from ML pipelines that use 5 to 15-minute aggregates of standard sensors (biogas flow, temperature, pH, influent characteristics), supplemented by soft sensors such as VFA/ALK estimators (Zou *et al.* 2024). Practical deployment involves routine outlier handling, rolling cross-validation to account for seasonal shifts, and external validation on unseen weeks. Probabilistic time-series models (e.g., LSTM/TFT with quantiles) transform predictions into risk bands that operators can map to actions such as moderating OLR ramps, adjusting HRT, or temporarily reducing recalcitrant co-substrates before acidification escalates (Sappl *et al.* 2023; Jeong *et al.* 2021; Salamattalab *et al.* 2024).

Hybrid mechanistic-ML approaches provide a balanced solution when both interpretability and accuracy are needed: a kinetic core captures the mass-balance structure, while ML learns residuals and context-specific effects (Gupta *et al.* 2023; Ling *et al.* 2024; Geng *et al.* 2024). This framework naturally fits with IoT-enabled “smart digesters,” where uncertainty-aware forecasts, explainable features, and control heuristics are integrated into operator dashboards to increase energy yield, reduce downtime, and support TEA/LCA decision-making for co-digestion and pre-treatment options.

In full-scale digesters, non-ideal hydraulics (dead zones, short-circuiting), variable RTD, and intermittent sensors violate the homogeneity and stationarity assumed by both kinetic and ML models. SCADA streams are irregularly sampled, exhibit drift, and are frequently unsynchronized with gas-quality measurements; without resampling, calibration, and basic QC, models learn artefacts (Sun *et al.* 2023; Zou *et al.* 2024). Seasonal substrate shifts and co-substrate swings cause distribution shift that degrades accuracy unless rolling validation and periodic recalibration are used (Yildirim and Ozkaya 2023; Ling *et al.* 2024). Standardizing units (e.g., mL CH₄ gVS⁻¹ at STP, dry gas) and documenting feed configuration are prerequisites for model transfer across sites.

Full-scale and pilot experiences increasingly show that data-driven models can directly improve operations when used with routine plant instrumentation. In industrial and municipal environments, tree-based ensembles and sequence models have provided reliable short-term forecasts and soft-sensor surrogates, with performance generally ranging from $R^2 \approx 0.60$ to 0.99 depending on horizon, inputs, and plant variability (Schroer *et al.* 2023; Yildirim and Ozkaya 2023; Zou *et al.* 2024; Sun *et al.* 2023; Salamattalab *et al.* 2024; Jeong *et al.* 2021). Feature attribution methods (e.g., SHAP, attention) consistently identify OLR, pH, temperature, and feed configuration as the main factors, supporting targeted set-point tuning and early-warning dashboards (Ling *et al.* 2024; Zou *et al.* 2024; Gupta *et al.* 2023).

From an economic perspective, the adoption of advanced ML models requires substantial investment in sensors, automated data acquisition systems, and skilled personnel for calibration and maintenance. Operational costs for energy, data storage, and software infrastructure may limit uptake, particularly in resource-constrained contexts. From an engineering standpoint, integrating predictive models into real-time plant control is complex, as biogas systems are subject to fluctuations in feedstock supply, microbial community dynamics, and environmental conditions. The operational reliability of IoT-enabled monitoring, communication latency, and data quality further constrain

implementation. These challenges highlight the importance of coupling modelling studies with TEA and LCA to ensure decision relevance. Beyond established frameworks, future research should prioritize the development of hybrid mechanistic–machine learning models tailored specifically for anaerobic digestion. For example, coupling ANN architectures with AD-specific kinetic equations (*e.g.*, hydrolysis–acidogenesis–methanogenesis dynamics) could combine predictive accuracy with mechanistic interpretability. Such models, trained on large-scale, multi-site datasets, would enable adaptive real-time control strategies unique to AD. This approach moves beyond general ML challenges, offering concrete, novel pathways for advancing AD modelling.

CONCLUSIONS

This review has presented a comprehensive analysis of the recent developments in mathematical modeling and ML applications for biogas production through anaerobic digestion. The findings indicate that while classical kinetic models like the first order and Gompertz provide proper baseline estimations, their assumptions such as limit performance under complex and dynamic AD conditions. In contrast, ANN and ML techniques demonstrate superior predictive accuracy, adaptability, and capability in managing nonlinear and multivariate systems. Nonetheless, the absence of standardized datasets, model interpretability issues, and lack of integration with real-time control systems remain challenges. Future research should focus on hybrid modeling approaches that leverage the strengths of both deterministic and data-driven methods, supported by advanced sensing technologies and cross-disciplinary collaboration. By addressing these gaps, the AD process can be optimized for enhanced energy recovery, system stability, and environmental sustainability, which will contribute significantly to circular economy strategies and global clean energy goals.

This review has integrated deterministic kinetics with modern ML for AD, providing a side-by-side appraisal of daily-rate *vs.* cumulative-yield families and clarifying when first-order, modified Gompertz, or Chen–Hashimoto formulations are most defensible. It advances a multidimensional parameterization that elevates kinetic parameters to functions of operating variables, and it frames parameter uncertainty using stochastic (random-variable/process) treatments to yield probabilistic production envelopes. By consolidating study-level metrics and sensitivity emphases (notably A and λ), the work offers a reproducible basis for model selection, benchmarking, and future hybrid mechanistic–ML development.

For practitioners, the review distills field-scale evidence that ML (ensembles and sequence models) can provide short-horizon forecasts and soft-sensor proxies at accuracy suitable for day-to-day control, while retaining modified-Gompertz-type kinetics for design and batch/BMP contexts. It maps explainable features (OLR, pH, temperature, feed configuration) to actionable levers, outlines a pragmatic deployment recipe (clean SCADA ingestion, rolling/external validation, probabilistic outputs), and proposes a hybrid mechanistic–ML blueprint compatible with IoT “smart digester” dashboards. These guidance points translate model choice into concrete decisions on OLR ramps, HRT adjustments, co-substrate scheduling, and risk-aware operations.

Finally, we recommend reporting sustainability KPIs (*e.g.*, GWP per kWh, LCOE/LCBG) alongside predictive accuracy and using probabilistic ML outputs to

propagate uncertainty into LCA/TEA, enabling Pareto-based selection of operating set-points and co-digestion strategies.

ACKNOWLEDGMENTS

The author would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

REFERENCES CITED

- Abdel Daiem, M. M., Hatata, A., Galal, O. H., Said, N., and Ahmed, D. (2021). "Prediction of biogas production from anaerobic co-digestion of waste activated sludge and wheat straw using two-dimensional mathematical models and an artificial neural network," *Renewable Energy* 178, 226-240. DOI: 10.1016/j.renene.2021.06.050
- Adeleke, O., Olatunji, K. O., Madyira, D. M., and Jen, T.-C. (2025). "Application of multimodal machine learning-based analysis for the biomethane yields of NaOH-pretreated biomass," *Scientific Reports* 15(1), article 24372. DOI: 10.1038/s41598-025-09527-5
- Adnane, I., Taoumi, H., Elouahabi, K., Lahrech, K., and Oulmekki, A. (2024). "Valorization of crop residues and animal wastes: Anaerobic co-digestion technology," *Heliyon* 10(5), article e26440. DOI: 10.1016/j.heliyon.2024.e26440
- Alengebawy, A., Ran, Y., Osman, A. I., Jin, K., Samer, M., and Ai, P. (2024). "Anaerobic digestion of agricultural waste for biogas production and sustainable bioenergy recovery: A review," *Environmental Chemistry Letters* 22, 2641-2668. DOI: 10.1007/s10311-024-01789-1
- Altaş, L. (2009). "Inhibitory effect of heavy metals on methane-producing anaerobic granular sludge," *Journal of Hazardous Materials* 162(2-3), 1551-1556. DOI: 10.1016/j.jhazmat.2008.06.048
- Amleh, M. A., and Al-Freihat, I. F. (2025). "Prediction of new lifetimes of a step-stress test using cumulative exposure model with censored Gompertz data," *Statistics, Optimization & Information Computing* 13(4), 1368-1387. DOI: 10.19139/soic-2310-5070-1852
- Asadi, M., and McPhedran, K. (2021). "Biogas maximization using data-driven modelling with uncertainty analysis and genetic algorithm for municipal wastewater anaerobic digestion," *Journal of Environmental Management* 293, article 112875. DOI: 10.1016/j.jenvman.2021.112875
- Beltramo, T., Klocke, M., and Hitzmann, B. (2019). "Prediction of the biogas production using GA and ACO input features selection method for ANN model," *Information Processing in Agriculture* 6(3), 349-356. DOI: 10.1016/j.inpa.2019.01.002
- Bilgili, M. S., Demir, A., and Varank, G. (2009). "Evaluation and modeling of biochemical methane potential (BMP) of landfilled solid waste: A pilot scale study," *Bioresource Technology* 100(21), 4976-4980. DOI: 10.1016/j.biortech.2009.05.012
- Budiyono, B., Widiyasa, I., Sunarso, S., and Johari, S. (2010). "The kinetic of biogas production rate from cattle manure in batch mode," *International Journal of Chemical and Molecular Engineering* 3(1), 39-44.

- Cruz, I. A., Nascimento, V. R. S., Felisardo, R. J. A., dos Santos, A. M. G., de Jesus, A. A., de Vasconcelos, B. R., Kumar, V., Cavalcanti, E. B., de Souza, R. L., and Ferreira, L. F. R. (2023). "Evaluation of artificial neural network models for predictive monitoring of biogas production from cassava wastewater: A training algorithms approach," *Biomass and Bioenergy* 175, article 106869. DOI: 10.1016/j.biombioe.2023.106869
- Danner, T. W. (2006). *A Formulation of Multidimensional Growth Models for the Assessment and Forecast of Technology Attributes*, Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, USA.
- Deepanraj, B., Sivasubramanian, V., and Jayaraj, S. (2017). "Effect of substrate pretreatment on biogas production through anaerobic digestion of food waste," *International Journal of Hydrogen Energy* 42(42), 26522-26528. DOI: 10.1016/j.ijhydene.2017.06.178
- Fan, Y., Wang, Y., Qian, P.-Y., and Gu, J.-D. (2004). "Optimization of phthalic acid batch biodegradation and the use of modified Richards model for modelling degradation," *International Biodeterioration & Biodegradation* 53(1), 57-63. DOI: 10.1016/j.ibiod.2003.10.001
- Galal, O. H. (2013). "A proposed stochastic finite difference approach based on homogenous chaos expansion," *Journal of Applied Mathematics* 2013, article 950469. DOI: 10.1155/2013/950469
- Galal, O. H. (2021). "Stochastic velocity modeling of Magneto-Hydrodynamics Non-Darcy flow between two stationary parallel plates," *Alexandria Engineering Journal* 60(4), 4191-4201. DOI: 10.1016/j.aej.2021.03.010
- Geng, Z., Shi, X., Ma, B., Chu, C., and Han, Y. (2024). "Biogas production prediction model of food waste anaerobic digestion for energy optimization using mixup data augmentation-based global attention mechanism," *Environmental Science and Pollution Research* 31(6), 9121-9134. DOI: 10.1007/s11356-023-31653-8
- Ghanem, R. G., and Spanos, P. D. (2003). *Stochastic Finite Elements: A Spectral Approach*, Springer New York, NY, USA.
- De Gioannis, G., Muntoni, A., Cappai, G., and Milia, S. (2009). "Landfill gas generation after mechanical biological treatment of municipal solid waste. Estimation of gas generation rate constants," *Waste Management* 29(3), 1026-1034. DOI: 10.1016/j.wasman.2008.08.016
- Gupta, R., Zhang, L., Hou, J., Zhang, Z., Liu, H., You, S., Ok, Y. S., and Li, W. (2023). "Review of explainable machine learning for anaerobic digestion," *Bioresource Technology* 369, article 128468. DOI: 10.1016/j.biortech.2022.128468
- Hsieh, Y.-H. (2009). "Richards model: A simple procedure for real-time prediction of outbreak severity," *Modeling and Dynamics of Infectious Diseases* 1, 216-236. DOI: 10.1142/97898142612650009
- Jafari-Sejehrood, A., Najafi, B., Faizollahzadeh Ardabili, S., Shamshirband, S., Mosavi, A., and Chau, K. (2019). "Limiting factors for biogas production from cow manure: Energo-environmental approach," *Engineering Applications of Computational Fluid Mechanics* 13(1), 954-966. DOI: 10.1080/19942060.2019.1654411
- Jameel, M. K., Mustafa, M. A., Ahmed, H. S., Jassim Mohammed, A., Ghazy, H., Shakir, M. N., Lawas, A. M., Khudhur Mohammed, S., Idan, A. H., and Mahmoud, Z. H. (2024). "Biogas: Production, properties, applications, economic and challenges: A review," *Results in Chemistry* 7, article 101549. DOI: 10.1016/j.rechem.2024.101549
- Jeong, K., Abbas, A., Shin, J., Son, M., Kim, Y. M., and Cho, K. H. (2021). "Prediction

- of biogas production in anaerobic co-digestion of organic wastes using deep learning models,” *Water Research* 205, article 117697. DOI: 10.1016/j.watres.2021.117697
- Komarysta, B., Dzhygyrey, I., Bendiuh, V., Yavorovska, O., Andreeva, A., Berezenko, K., Meshcheriakova, I., Vovk, O., Dokshyna, S., and Maidanskyi, I. (2023). “Optimizing biogas production using artificial neural network,” *Eastern-European Journal of Enterprise Technologies* 2(8(122)), 53-64. DOI: 10.15587/1729-4061.2023.276431
- Kumar, S., Mondal, A. N., Gaikwad, S. A., Devotta, S., and Singh, R. N. (2004). “Qualitative assessment of methane emission inventory from municipal solid waste disposal sites: A case study,” *Atmospheric Environment* 38(29), 4921-4929. DOI: 10.1016/j.atmosenv.2004.05.052
- Latinwo, G. K., and Agarry, S. E. (2015). “Modelling the kinetics of biogas production from mesophilic anaerobic co-digestion of cow dung with plantain peels,” *International Journal of Renewable Energy Development* 4(1), 55-63. DOI: 10.14710/ijred.4.1.55-63
- Li, C., and Fang, H. H. P. (2007). “Inhibition of heavy metals on fermentative hydrogen production by granular sludge,” *Chemosphere* 67(4), 668-673. DOI: 10.1016/j.chemosphere.2006.11.005
- Li, M., Zhao, Y., Guo, Q., Qian, X., and Niu, D. (2008). “Bio-hydrogen production from food waste and sewage sludge in the presence of aged refuse excavated from refuse landfill,” *Renewable Energy* 33(12), 2573-2579. DOI: 10.1016/j.renene.2008.02.018
- Lin, C.Y., and Shei, S.H. (2008). “Heavy metal effects on fermentative hydrogen production using natural mixed microflora,” *International Journal of Hydrogen Energy* 33(2), 587-593. DOI: 10.1016/j.ijhydene.2007.09.030
- Ling, J. Y. X., Chan, Y. J., Chen, J. W., Chong, D. J. S., Tan, A. L. L., Arumugasamy, S. K., and Lau, P. L. (2024). “Machine learning methods for the modelling and optimisation of biogas production from anaerobic digestion: A review,” *Environmental Science and Pollution Research* 31(13), 19085-19104. DOI: 10.1007/s11356-024-32435-6
- Liu, Y., Watanabe, R., Li, Q., Luo, Y., Tsuzuki, N., Ren, Y., Qin, Y., and Li, Y.-Y. (2025). “Enhanced biomethane production by thermophilic high-solid anaerobic co-digestion of rice straw and food waste: Cellulose degradation and microbial structure,” *Chemical Engineering Journal* 503, article 158088. DOI: 10.1016/j.cej.2024.158088
- Lo, H. M., Kurniawan, T. A., Sillanpää, M. E. T., Pai, T. Y., Chiang, C. F., Chao, K. P., Liu, M. H., Chuang, S. H., Banks, C. J., and Wang, S. C. (2010). “Modeling biogas production from organic fraction of MSW co-digested with MSWI ashes in anaerobic bioreactors,” *Bioresource Technology* 101(16), 6329-6335. DOI: 10.1016/j.biortech.2010.03.048
- Lohani, S. P., Shakya, S., Gurung, P., Dhungana, B., Paudel, D., and Mainali, B. (2025). “Anaerobic co-digestion of food waste, poultry litter and sewage sludge: Seasonal performance under ambient condition and model evaluation,” *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 47(2), article 1887976. DOI: 10.1080/15567036.2021.1887976
- Mu, Y., Yu, H.-Q., and Wang, G. (2007). “A kinetic approach to anaerobic hydrogen-producing process,” *Water Research* 41(5), 1152-1160. DOI: 10.1016/j.watres.2006.11.047
- Mueller, L. D., Nusbaum, T. J., and Rose, M. R. (1995). “The Gompertz equation as a

- predictive tool in demography,” *Experimental Gerontology* 30(6), 553-569. DOI: 10.1016/0531-5565(95)00029-1
- Najafi, B., and Ardabili, S. F. (2018). “Application of ANFIS, ANN, and logistic methods in estimating biogas production from spent mushroom compost (SMC),” *Resources, Conservation and Recycling* 133, 169-178. DOI: 10.1016/j.resconrec.2018.02.025
- Nielfa, A., Cano, R., Vinot, M., Fernández, E., and Fdz-Polanco, M. (2015). “Anaerobic digestion modeling of the main components of organic fraction of municipal solid waste,” *Process Safety and Environmental Protection* 94, 180-187. DOI: 10.1016/j.psep.2015.02.002
- Olatunji, K. O., Mootswi, K. D., Olatunji, O. O., Zwane, M. I., van Rensburg, N. J., and Madyira, D. M. (2025). “Anaerobic co-digestion of food waste and groundnut shells: Synergistic impact assessment and kinetic modeling,” *Waste and Biomass Valorization* 16, 3745-3760. DOI: 10.1007/s12649-025-02904-1
- Peleg, M., and Corradini, M. G. (2011). “Microbial growth curves: what the models tell us and what they cannot,” *Critical Reviews in Food Science and Nutrition* 51(10), 917-945. DOI: 10.1080/10408398.2011.570463
- Pulgarín-Muñoz, C. E., Saldarriaga-Molina, J. C., Correa-Ochoa, M. A., and Castro-Valencia, J. C. (2025). “Effect of cosubstrate ratio and temperature on sewage sludge and agro-industrial fruit and vegetable waste anaerobic co-digestion,” *Waste and Biomass Valorization* 2025, available online. DOI: 10.1007/s12649-025-03129-y
- Roberts, S., Mathaka, N., Zeleke, M. A., and Nwaigwe, K. N. (2023). “Comparative analysis of five kinetic models for prediction of methane yield,” *Journal of The Institution of Engineers (India) Series A* 104, 335-342. DOI: 10.1007/s40030-023-00715-y
- Rossi, E., Pecorini, I., and Iannelli, R. (2022). “Multilinear regression model for biogas production prediction from dry anaerobic digestion of OFMSW,” *Sustainability* 14(8), article 4393. DOI: 10.3390/su14084393
- Said, N., Alblawi, A., Hendy, I. A., and Abdel Daiem, M. M. (2020). “Analysis of energy and greenhouse gas emissions of rice straw to energy chain in Egypt,” *BioResources* 15(1), 1510-1520. DOI: 10.15376/biores.15.1.1510-1520
- Salamattalab, M. M., Zonoozi, M. H., and Molavi-Arabshahi, M. (2024). “Innovative approach for predicting biogas production from large-scale anaerobic digester using long-short term memory (LSTM) coupled with genetic algorithm (GA),” *Waste Management* 175, 30-41. DOI: 10.1016/j.wasman.2023.12.046
- Sapfl, J., Harders, M., and Rauch, W. (2023). “Machine learning for quantile regression of biogas production rates in anaerobic digesters,” *Science of The Total Environment* 872, article 161923. DOI: 10.1016/j.scitotenv.2023.161923
- Schroer, H. W., and Just, C. L. (2023). “Feature engineering and supervised machine learning to forecast biogas production during municipal anaerobic co-digestion,” *ACS ES&T Engineering* 4(3), 660-672. DOI: 10.1021/acsestengg.3c00435
- Shindell, D., Sadavarte, P., Aben, I., Bredariol, T. de O., Dreyfus, G., Höglund-Isaksson, L., Poulter, B., Saunois, M., Schmidt, G. A., and Szopa, S. (2024). “The methane imperative,” *Frontiers in Science* 2, article 1349770. DOI: 10.3389/fsci.2024.1349770
- Simon, M. K. (2002). *Probability Distributions Involving Gaussian Random Variables*, Springer, New York, NY, USA.
- Sun, J., Xu, Y., Nairat, S., Zhou, J., and He, Z. (2023). “Prediction of biogas production

- in anaerobic digestion of a full-scale wastewater treatment plant using ensembled machine learning models,” *Water Environment Research* 95(6), article e10893. DOI: 10.1002/wer.10893
- Tian, Y., Yang, K., Zheng, L., Han, X., Xu, Y., Li, Y., Li, S., Xu, X., Zhang, H., and Zhao, L. (2020). “Modelling biogas production kinetics of various heavy metals exposed anaerobic fermentation process using sigmoidal growth functions,” *Waste and Biomass Valorization* 11(9), 4837-4848. DOI: 10.1007/s12649-019-00810-x
- Tiwari, S. B., Dixit, S., Tyagi, V. K., Veksha, A., Lim, T.-T., and Kazmi, A. A. (2025). “Comparative evaluation of mechanistic models for biogas production in DIET-enhanced anaerobic digestion,” *Journal of Environmental Management* 391, article ID 126614. DOI: 10.1016/j.jenvman.2025.126614
- Tonner, P. D., Darnell, C. L., Engelhardt, B. E., and Schmid, A. K. (2017). “Detecting differential growth of microbial populations with Gaussian process regression,” *Genome Research* 27(2), 320-333. DOI: 10.1101/gr.210286.116
- Tufaner, F., Dalkılıç, K., and Uğurlu, A. (2025). “Artificial intelligence-based modeling of biogas production in a combined microbial electrolysis cell-anaerobic digestion system using artificial neural networks and adaptive neuro-fuzzy inference system,” *Environmental Science and Pollution Research* 32, 12524-12546. DOI: 10.1007/s11356-025-36467-4
- Yildirim, O., and Ozkaya, B. (2023). “Prediction of biogas production of industrial scale anaerobic digestion plant by machine learning algorithms,” *Chemosphere* 335, article 138976. DOI: 10.1016/j.chemosphere.2023.138976
- Yusuf, M. O. L., Debora, A., and Ogheneruona, D. E. (2011). “Ambient temperature kinetic assessment of biogas production from co-digestion of horse and cow dung,” *Research in Agricultural Engineering* 57(3), 97-104. DOI: 10.17221/25/2010-RAE
- Zhu, T., Zhou, Y., Chen, J. M., Ju, W., Yan, R., Xie, R., and Mao, Y. (2025). “Divergent responses of CH₄ emissions and uptake to global change drivers,” *Global Biogeochemical Cycles* 39(3), article e2024GB008183. DOI: 10.1029/2024GB008183
- Zou, J., Lü, F., Chen, L., Zhang, H., and He, P. (2024). “Machine learning for enhancing prediction of biogas production and building a VFA/ALK soft sensor in full-scale dry anaerobic digestion of kitchen food waste,” *Journal of Environmental Management* 371, article 123190. DOI: 10.1016/j.jenvman.2024.123190
- Zwietering, M. H., Jongenburger, I., Rombouts, F. M., and Van’t Riet, K. (1990). “Modeling of the bacterial growth curve,” *Applied and Environmental Microbiology* 56(6), 1875-1881. DOI: 10.1128/aem.56.6.1875-1881.1990

Article submitted: August 5, 2025; Peer review completed: September 5, 2025; Revised version received: September 14, 2025; Accepted: September 15, 2025; Published: September 17, 2025.

DOI: 10.15376/biores.20.4.Galal