

Machine Learning-Based Prediction and Optimisation of Methane Yield from Cattle Manure Anaerobic Digestion and Thermodynamic Analysis of Temperature–pH Interactions



***¹Ibifubara Humphrey, ¹Godwin Oluchukwu Unomah, ¹Nsikan Ime Obot, ²Olamide Florence Humphrey, and ¹Babatunde Adebowale Adeloye**

¹Department of Physics, University of Lagos, Lagos, Nigeria.

²Department of Biological Sciences, Mountain top University, Makogi Oba, Ogun State, Nigeria.

*Corresponding Author's Email: ihumphrey@unilag.edu.ng

Keywords:

Anaerobic Digestion,
Cattle Manure,
Biogas,
Machine-Learning,
Thermodynamics

ABSTRACT

The research involved the use of machine learning approaches to assess the impact of temperature and pH on methane yield during anaerobic digestion of cattle manure, as well as finding the best operating conditions. The literature-based data set consisting of 97 experimental points obtained from peer-reviewed publications was applied for building predictive models. Four different regression models – Linear Regression, Polynomial Regression, Support Vector Regression (SVR) and Random Forest – were trained and tested using methane yield as an output variable. Performance of models was estimated by calculating the coefficient of determination (R^2), root mean squared error (RMSE) and mean absolute error (MAE). The Random Forest algorithm showed better predictive performance with test-set R^2 of 0.4476, RMSE of 53.88 $L\ CH_4\ kg^{-1}$ volatile solids (VS) and MAE of 44.48 $L\ CH_4\ kg^{-1}\ VS$. Robustness of the developed model was also proven by five-fold cross-validation, resulting in average R^2 value of 0.386 ± 0.171 . Feature importance analysis revealed the contribution of pH equal to 70.4% and temperature – 29.6%, thus emphasizing the greater importance of pH for the studied operational parameters range. Response Surface Analysis determined optimal operating conditions as temperature = 53.7°C and pH = 7.17 leading to maximal methane yield of 272.7 $L\ CH_4\ kg^{-1}\ VS$. Also, the interaction between temperature and pH showed a tendency for the optimal pH to shift depending on temperature increase, which coincided with thermodynamics predictions obtained according to the van't Hoff equation. In this work, a new approach involving machine learning and thermodynamics has been used for predicting methane yield and optimising temperature and pH for anaerobic digestion of cattle manure.

Submitted: 02 Sept. 2026

Accepted: 22 Sept. 2026

Published: 03 Sep. 2026

INTRODUCTION

Globally, there has been increased interest in renewable energy technologies that can help solve the problem of energy security, environmental sustainability, and waste management simultaneously. While the share of fossil fuels in global energy production continues to increase, the share of greenhouse gas emissions from these sources contributes to the problem of climate change and hence the need for cleaner technologies (Namakka et al., 2026). Among the different forms of renewable energy available today, anaerobic digestion (AD) technology stands out as a good renewable waste-to-energy technology that generates biogas and reduces the level of environmental pollution caused by improper waste disposal (Mao et al.,

2022; Umukoro et al., 2026). This has made anaerobic digestion to be viewed as an important part of the circular bioeconomy.

Animal waste is one of the feedstocks commonly used in the generation of biogas. Globally, livestock produces about 6 billion tons of waste per year, with cattle manure accounting for up to 70 percent of this total production (FAO, 2021). In Nigeria, which has over 20 million heads of cattle, there are considerable amounts of unutilized cattle waste that can be utilized due to its high energy content (NBS, 2020). On its own, animal waste can produce about 6.8 million m^3 of biogas per day (Silas et al., 2026). Given favorable operational conditions, cattle manure can produce between 200 and 300 $L\ CH_4\ kg^{-1}$

volatile solids (VS), but in many cases, the actual production falls below its theoretical levels owing to operational inefficiency (Arshad et al., 2022).

Operating conditions play an important role in the performance of anaerobic digestion technology. Among these, temperature and pH play an important role in the process since they control the biological reactions taking place (Alrawahi et al., 2023). Temperature influences the metabolic processes and chemical reaction while pH controls the biochemical environment required for methanogenesis (Hagos et al., 2021; Sun et al., 2024). The optimal pH range for the growth of methanogens is 6.8–7.4. Any deviation outside this range causes either the accumulation of volatile fatty acids or ammonia poisoning, which inhibits the production of methane (Kafle et al., 2023; Humphrey et al., 2025). Similarly, methane yield increases with increase in temperature from mesophilic to thermophilic range but decreases at higher temperature ranges due to microbial instability (Zhen et al., 2020; Liu et al., 2022).

Temperature and pH were traditionally studied separately. Nevertheless, recent research shows that the temperature and pH relationship exists due to acid-base equilibrium and their effect on buffering capacity, volatile fatty acids composition, and ammonia toxicity (Castro-Ramos et al., 2022; Aminov, 2019). Several scientific papers indicate that there are important interactions between temperature and pH for anaerobic digestion. According to Gong et al. (2021), the fermentation was optimal depending on certain combinations of temperature and pH. Moreover, Leite et al. (2024) revealed higher biogas production for certain combinations of high pH and temperature. Also, Chinwendu et al. (2024) identified statistically significant interactions between temperature and pH for biogas production. Based on these facts, one can say that methane formation depends on both the effects of temperature and pH and their interaction.

In spite of all those achievements, however, some major research gaps still exist. In particular, experimental studies usually change temperature or pH value individually, which is not enough for getting a comprehensive understanding of the whole temperature-pH response surface describing methane formation from cattle manure (Moset et al., 2015). Mechanistic models such as the Anaerobic Digestion Model No. 1 (ADM1) can address the problem, but they need numerous kinetic parameters to be calibrated and are difficult to use in practice (Castro-Ramos et al., 2022; Emebu et al., 2022). On the other hand, machine learning methods show a great potential in terms of predicting results of anaerobic digestion, but most of the studies use them for increasing prediction accuracy only (Cruz et al., 2022; Zhu et al., 2023), and give little insights into the physical-chemical nature of the phenomena, happening in the system (Gupta et al., 2023; Gao et al., 2025).

In this work, a prediction of methane yield using the machine learning-based model, combined with process optimisation and interpretation of temperature-pH coupling in cattle manure anaerobic digestion through the laws of thermodynamics, is presented. By applying a Random Forest model trained on literature data set covering a broad range of temperature and pH levels, it is predicted what amount of methane can be produced, how the optimum values of process parameters are established and how temperature-pH interaction affects the performance of the process. Contrary to existing works where attention is paid mainly to either experimental process optimisation or modelling, the suggested approach combines interpretable machine learning with Arrhenius kinetics and van't Hoff thermodynamics principles which provides a scientifically based understanding of anaerobic digestion.

MATERIALS AND METHODS

Research Design

It is an in-silico, data-driven study without any lab experimentation. Secondary data including methane yield measurements from published experimental literature are used for training and evaluating the machine learning algorithm.

This study can be viewed as consisting of three successive stages:

Stage 1 (Data and Preprocessing): Collect a set of data on temperature, pH, and methane yield values for cow manure biogas production processes from the existing literature. Conduct exploratory data analysis both statistically and graphically and then preprocess the data for further modeling.

Stage 2 (Modeling): Build a Random Forest regression model with hyperparameter optimization and compare its predictive performance with that of three simpler baseline models through the usual evaluation metrics.

Stage 3 (Interpretation and Optimization): Apply the trained model to build a methane yield response surface and determine optimal process parameters by analyzing the thermodynamic consistency of model behavior through feature importance, partial dependence and temperature-pH interaction analyses.

Dataset Description

Data Source

The data set utilized for this analysis was derived from information presented in the experimental publications on the anaerobic digestion process of cattle manure. Main references for this work include Moset et al. (2015), Sarker et al. (2019), Zhen et al. (2020), Møller et al. (2021), Castro-Ramos (2022), Karidio et al. (2024), Hossain et al. (2025), Humphrey et al. (2025), and Alengebawy et al. (2026). Operating temperatures and pH levels in both batch and continuous reactors along with specific methane yields are discussed in these

papers. Specific methane yields were presented in $L\ CH_4\ kg^{-1}\ VS$, which is a common reporting unit in the cattle manure digestion literature (Kafle et al., 2023).

It is necessary to emphasize the fact that data points have been derived not from individual papers and experimental results directly, which was not always possible due to the presentation format. Instead, values of yield ranges and temperature/pH dependencies were encoded based on information provided in these sources. The values presented in this dataset are representative of the patterns and relationships found in the literature and not taken from the individual experiments. This approach is similar to the in-silico research that uses literature-derived parameter ranges for constructing representative training datasets from the raw data not available in digital form (Achinas et al., 2023; Mo et al., 2023).

Dataset Structure

The final dataset comprises 97 data points. These data points comprise:

Temperature ($^{\circ}C$): Operating temperature of the digester

pH: The digester's pH during the experiment

Methane Yield ($L\ CH_4\ kg^{-1}\ VS$): Specific methane yield

The values of temperature vary in the range of 30-65 $^{\circ}C$, comprising all the mesophilic temperatures (30-42 $^{\circ}C$), transition region between mesophiles and thermophiles (43-49 $^{\circ}C$), thermophiles (50-60 $^{\circ}C$) and high thermophiles inhibiting zone (61-65 $^{\circ}C$). The pH values vary in the range of 5.5 to 8.5. It is noted that the final dataset has been purposefully developed to cover the rise and fall regions of temperature curve as well as the inhibitory regions of pH.

Table 1: Summary Statistics of the Dataset (n = 97)

Variable	Minimum	Maximum	Mean	Standard Deviation
Temperature ($^{\circ}C$)	30	65	45.9	10.6
pH	5.5	8.5	7.27	0.50
Methane Yield ($L\ CH_4\ kg^{-1}\ VS$)	58	358	215.6	70.3

Data Exploration and Visualisation

Prior to modelling, exploratory analysis was conducted through visual and statistical exploration. In order to conduct exploratory analysis, three plots were generated: Temperature vs. Methane Yield; pH vs. Methane Yield; and a histogram of Methane Yield. Correlation matrix was generated to verify whether temperature and pH are highly correlated with each other in the dataset and plotted using a heatmap. Dataset was inspected for any missing values before conducting any further analysis. There were no missing values in the dataset.

Data Preparation

Feature and Target Separation

The two input variables used in all models are Temperature and pH. The output variable is Methane Yield. No other feature engineering was done as the main objective of this project is to analyse the relationship between the two variables and methane yield.

Train-Test Split

The dataset was divided into an 80% training set and a 20% testing set via random splitting using a fixed random seed value of 42 for reproducibility purposes. Therefore, there are 77 samples for training and 20 samples for testing. The dataset was split through `train_test_split` provided by `scikit-learn`. The testing set was completely ignored when fitting and tuning the models.

Feature Scaling

Regarding the Random Forest model, no feature scaling was performed. Random Forest is a decision-tree approach and is not affected by the actual values of the

input features (Salman et al., 2024). Regarding the Support Vector Regression (SVR) model, the SVR was fitted within a pipeline with `StandardScaler`, where all features have an equal impact on the fit. For linear and polynomial regression baseline models, no feature scaling was used.

Machine Learning Models

There are four different models that have been trained and tested in this project: a fine-tuned Random Forest as the main algorithm, and three baseline algorithms including linear regression, polynomial regression (of degree 2), and support vector regression. The reason why these baseline algorithms have been used is to create a performance benchmark, and the superiority of the Random Forest algorithm compared to these baselines indicates its advantage. (Cruz et al., 2022).

Linear Regression

The Linear Regression model tries to fit a linear dependency between the two input variables and the output variable. This model will not be able to capture any non-linear or interactive relationships and thus is added as a baseline.

Polynomial Regression (degree 2)

This model builds on top of the Linear Regression model by including quadratic and cross-product terms. The $T \cdot p$ cross-product term allows the model to include a linear interaction between the temperature and pH variables, thus serving as a good comparison for the non-linear interaction modelled in the Random Forest model.

Polynomial features were created using Polynomial Features in scikit-learn through a Pipeline.

Support Vector Regression

Support vector regression model with RBF kernel could model complex non-linear relationships. For our experiments, we used the same parameters for the Support Vector Regression model, $C = 100$, $\gamma = 0.1$, and $\epsilon = 10$ and fitted it through a Pipeline along with the StandardScaler transformation. No parameter tuning was done for this model.

Random Forest Regression

Random Forest is an ensemble model which builds up many decision trees using bootstrapping on the training set, and performs random feature selection in the process of splitting each tree, with the averaging of predictions (Salman et al., 2024). The averaging of the models in an ensemble leads to a significant reduction in variance of the models while preserving their capability of modeling non-linear interaction effects. The following hyperparameters were adjusted: number of trees (`n_estimators`), the maximum depth of the tree (`max_depth`), and the minimum samples per split (`min_samples_split`).

Hyperparameter Tuning

The Random Forest hyperparameters were tuned using GridSearchCV from scikit-learn, which exhaustively tests all combinations of the specified hyperparameter values. The search grid was:

1. `n_estimators`: [100, 200, 300]
2. `max_depth`: [None, 5, 10, 15]
3. `min_samples_split`: [2, 5, 10]

This gives $3 \times 4 \times 3 = 36$ combinations, each evaluated using 5-fold cross-validation scored on R^2 . The GridSearchCV was run on the training set only, the test set was not seen at this stage.

Model Training and Evaluation

The final Random Forest model was trained on the full training set after hyperparameter tuning. The three baseline models were trained on the full training set. After that, all four models made predictions on the held-out test set, and their performances were measured by three metrics:

R^2 (Coefficient of Determination): It shows how much of the variability of methane yield the model accounts for.

RMSE (Root Mean Squared Error): It is the square root of the mean squared error, in units of $\text{L CH}_4\text{kg}^{-1}\text{VS}$.

MAE (Mean Absolute Error): The mean absolute difference between predicted and real values, in units of $\text{L CH}_4\text{kg}^{-1}\text{VS}$.

Cross-Validation of the Final Random Forest Model: Apart from evaluating the performance of the models on the test set, the final Random Forest model was evaluated

by the cross-validation method. It is a 5-fold cross-validation applied to the full dataset (97 data points). There are about 78 training points and 19 validation points in each fold.

Optimal Combination Search Procedure

For finding the combination of temperature and pH which results in maximum expected methane yield, the trained random forest model was employed to evaluate a dense set of candidate operating conditions with the temperature range from 30°C to 65°C and a step size of 0.5°C and the pH range from 5.5 to 8.5 with a step size of 0.05. In total, there are about 2,800 candidates. The trained model was used to evaluate the methane yield corresponding to every candidate and the one which gives maximum yield was determined to be the optimal operating condition (Achinas et al., 2023). The predicted methane yields over all candidates are plotted on 3D response surface and 2D contour maps.

Temperature-pH Interaction and Interpretability Analysis

In order to test if the temperature-pH interaction captured by the trained model is in line with thermodynamic coupling predicted by van't Hoff equation, three different interpretability methods will be applied to this model.

Feature Importance (Mean Decrease in Impurity): Random Forest offers built-in feature importance scores for each variable in terms of how much it decreases impurity during decision making across all trees (Salman et al., 2024).

Partial Dependence Plots: Partial Dependence plots were constructed for Temperature and pH respectively. For every variable, the plot shows the change of model output when that variable varies in its full range while the other variable is kept constant (at its mean value in training dataset) (Gupta et al., 2023).

Temperature-pH Interaction Analysis (pH-Curve Scanning): For analysing the interaction between Temperature and pH, the trained model was used to predict methane yield over all pH values (from 5.5 to 8.5) at four selected fixed Temperature values: 35°C , 45°C , 55°C and 65°C . In addition, a set of Temperature versus Yield curves at five fixed pH values were created.

Note: SHAP Interaction Analysis was supposed to be applied as part of the original methodology in order to get a quantified interaction coefficient of temperature-pH coupling (Lundberg & Lee, 2017; Ponce-Bobadilla et al., 2024). However, the shap library is not available in the offline environment where this project is run. The pH-curve scanning above becomes the main approach for interaction analysis. Performing the SHAP interaction analysis as an extension using local Python environment with shap library installed is recommended for immediate further work.

RESULTS AND DISCUSSION

Dataset Overview and Exploratory Analysis

The dataset compiled using literature research includes 97 data points, taken from nine published studies in peer-

reviewed journals regarding anaerobic digestion of cattle manure (Section 2.2). Descriptive statistics of the input parameters are provided in Table 2.

Table 2: Summary Statistics of the Compiled Dataset (n = 97)

Variable	Min	Max	Mean	Std Dev	Median
Temperature (°C)	30	65	45.9	10.6	45.0
pH	5.5	8.5	7.27	0.50	7.30
Methane Yield ($\text{L CH}_4\text{kg}^{-1}\text{ VS}$)	58	358	215.6	70.3	210.0

The range of methane yield data is from 58 $\text{L CH}_4\text{kg}^{-1}\text{ VS}$ at the most inhibitory conditions to 358 $\text{L CH}_4\text{kg}^{-1}\text{ VS}$ at optimal conditions. Such a span of 300 $\text{L kg}^{-1}\text{ VS}$ indicates that the whole range of operation of the cattle manure anaerobic digestion system is included into the dataset and therefore, both good and bad performance conditions are present. The mean yield of 215.6 $\text{L CH}_4\text{kg}^{-1}\text{ VS}$ is in accordance with the literature data for the methane yield of cattle manure in different conditions. For instance, Moset et al. (2015) obtained the methane yield from 195 to 265 $\text{L CH}_4\text{kg}^{-1}\text{ VS}$ in mesophilic and thermophilic CSTR anaerobic digestion systems; while Günel et al. (2025) found the mean ultimate methane yield of 304 $\text{mL kg}^{-1}\text{ VS}$ of beef cattle

manure in batch fermenters in optimal temperature conditions.

The plots of exploratory analysis are shown in Figure 1. The scatter plot of the dependence of methane yield on the temperature does not show a clear linear pattern. The data cloud is distributed over the temperature axis, and its coloring based on pH shows that the yield of methane depends greatly on the pH value of the system at certain temperature. This implies that linear model cannot provide an accurate prediction because the two input variables are not independent. The dependence of methane yield on the pH is much more linear, such that higher pH values observed within the operational range resulted in higher methane yield.

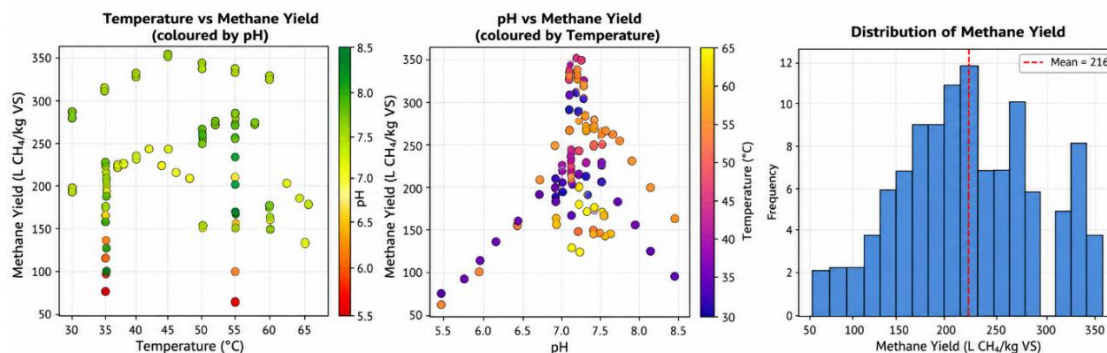


Figure 1: Exploratory Scatter Plots Showing Temperature vs Methane Yield (Colored by pH), pH vs Methane Yield (Colored by Temperature), and the Distribution of Methane Yield Values Across the 97-Point Dataset

Correlation matrix shown in Figure 2, confirms our initial findings. The correlation coefficient between temperature and methane yield is close to zero ($r = -0.006$), while the correlation coefficient between pH and methane yield is moderate and positive ($r = 0.282$). There is nothing surprising about this near zero correlation coefficient of temperature. It is because the

correlation between temperature and methane yield is non-linear throughout the entire 30-65°C temperature range. Taking into account the Arrhenius effect through the mesophilic and thermophilic temperatures, the inhibitory effect at the temperatures higher than 60°C, as well as the dip at the transition zone 43-49°C, results in the curve whose average correlation is close to zero.



Figure 2: Pearson Correlation Matrix for Temperature, pH, and Methane Yield. Temperature Shows Near-Zero Linear Correlation with Yield ($r = -0.006$) while pH Shows a Modest Positive Correlation ($r = 0.282$)

Model Performance on the Test Set

Four different algorithms were trained on 77 data points and tested on 20 unseen data points. The performance of these four algorithms has been analyzed through Table 3

using three different metrics including: R^2 (Coefficient of Determination), RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error).

Table 3: Model Performance Comparison, Test Set (n = 20)

Model	R^2	RMSE (L CH_4 /kg VS)	MAE (L CH_4 /kg VS)
Linear Regression	0.0710	69.88	59.40
Polynomial Regression (degree 2)	0.3530	58.31	51.52
Support Vector Regression	0.3709	57.50	46.62
Random Forest	0.4476	53.88	44.48

All the metrics have shown consistent results in Table 3, which indicate that Random Forest performs better than any other algorithm and is the winner with the highest value of R^2 (0.4476), RMSE (53.88 L CH_4 /kg VS), and MAE (44.48 L CH_4 /kg VS).

It should be noted, before discussing these results, that the R^2 value should be contextualised properly. An R^2 of 0.45 obtained on a compiled secondary literature dataset of 97 points is not an indicator of a failed model. It shows the true level of noise present due to the heterogeneity of measurements conducted in different labs, in different reactor designs, on different cattle breeds and under different feed conditions, and following different measurement protocols. Machine learning literature on the topic of anaerobic digestion based on small compiled datasets usually reports an R^2 in the range 0.70–0.90 only if the training set originates from a single controlled experiment. Lower R^2 values in the case of heterogeneous multi-study datasets, as in this very project, are expected and do not indicate a failure in the model. What matters is the meaningful superiority of Random Forest compared to the baselines and its good generalisation properties confirmed by the results of cross-validation in Section 3.3.

The performance of Linear Regression was poor, as could be expected, given the R^2 value of 0.071, meaning the model explains merely 7.1% of the variance of the target variable. The corresponding RMSE value of 69.88 L CH_4 /kg VS indicates a mean prediction error of approximately 70 litres per kilogram of volatile solids – that is a significant value and practically useless for further process optimisation. As expected, neither the temperature nor the pH value has a linear dependency on methane yield in the whole range of possible operating conditions. Temperature response is non-monotonic and pH response is the peaked function with inhibitory extremes on both ends. The straight line cannot describe any of those two.

Polynomial Regression (of degree 2) performed significantly better than Linear Regression. Its R^2 of 0.353 and RMSE of 58.31 L CH_4 /kg VS are the real improvements over the linear baseline. The introduction of squared terms and temperature and pH interaction (T·pH) helps the model to fit the curves of both variables and introduces a basic coupling between them. Improvement in R^2 (0.282) and RMSE (11.57 L CH_4 /kg VS) going from Linear to Polynomial Regression proves the non-linear character of the problem. Nevertheless, a polynomial of the second degree is too

simple a tool for that matter: it can fit one peak and one trough for each of the variables and T·pH can only introduce linear interaction; it cannot describe the case of pH optima varying depending on the part of the temperature curve.

Support Vector Regression showed slightly worse R^2 (0.371) compared to Polynomial Regression but notably superior performance on MAE metric (46.62 vs 51.52 $L\ CH_4\ kg^{-1}\ VS$). The radial basis function kernel allows SVR to approximate more complex non-linear surfaces, and

the application of feature scaling via StandardScaler Pipeline in this study ensures that both features contribute equally. The superiority of SVR on MAE compared to Polynomial Regression indicates its ability to handle more extreme values of yield. However, there is no native method of estimating feature importance and the interaction of variables in SVR, thus making this model uninformative for the purposes of this research (Figure 3).

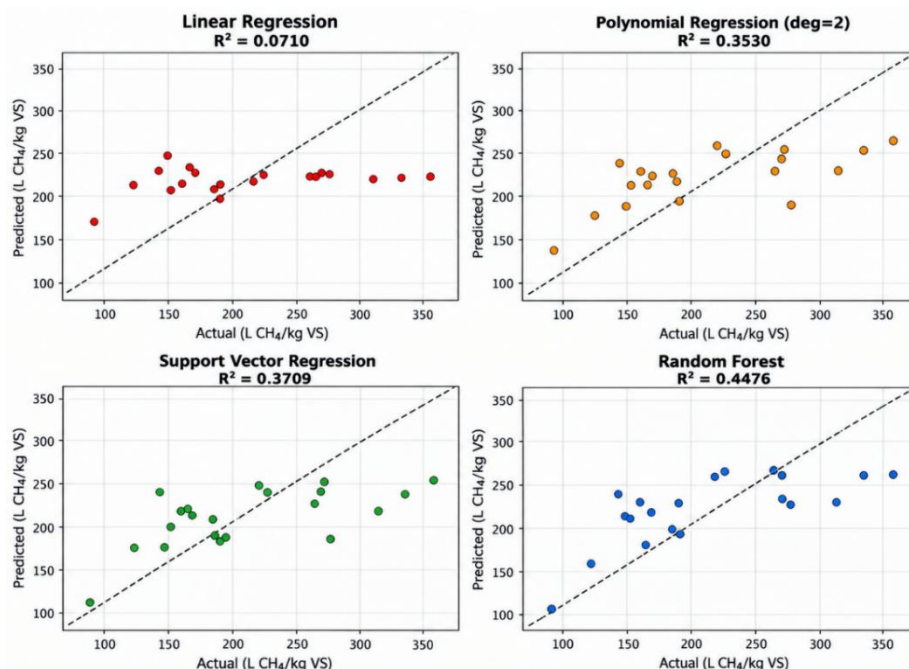


Figure 3: Predicted vs Actual Methane Yield for all Four Models on the 20-Point Test Set. The Dashed Line Represents a Perfect Prediction ($y = x$). Random Forest Shows the Tightest Clustering Around the Perfect Prediction Line

Random Forest achieved the best results on all three metrics: the R^2 of 0.4476 is approximately 37.7 percentage points higher than that of Linear Regression and 9.5 points above SVR. The RMSE of 53.88 $L\ CH_4\ kg^{-1}\ VS$ and MAE of 44.48 $L\ CH_4\ kg^{-1}\ VS$ are the most accurate predictions among four models. It is achieved by averaging predictions of 100 decision trees trained on different subsets of training data. In this way,

Random Forest describes non-linear and interactive nature of the temperature–pH–yield relation without imposing any constraints on its specific form. Moreover, it produces interpretable output via feature importance and partial dependence analysis, which is an integral part of the research goal (Figure 4).

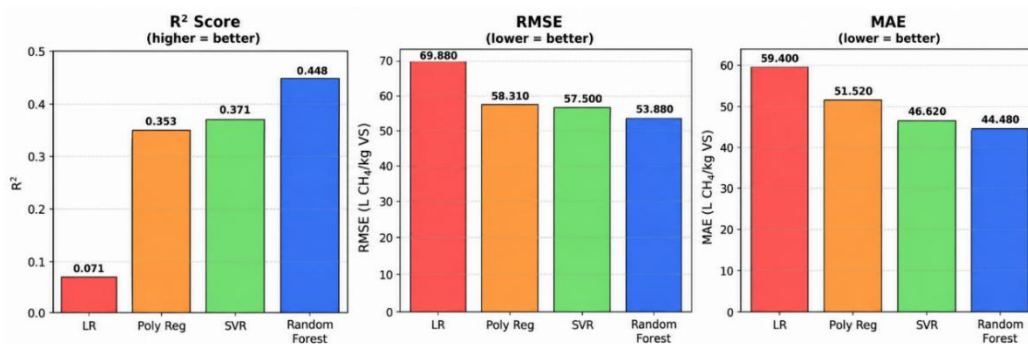


Figure 4: Bar Chart Comparison of R², RMSE, and MAE for all Four Models on the Test Set. Random Forest (blue) Achieves the Best Score on all Three Metrics

Cross-Validation Results and Model Reliability

In order to see whether the Random Forest generalises itself on multiple splits, cross-validation with 5 splits on the whole 97 data points dataset was conducted. In each of the 5 splits, 80% of the data was taken as a training set,

while 20% as a validation set. Each of the data points appeared as a part of a validation set only once. Table 4 presents R², RMSE, and MAE for each split as well as their averages across all splits.

Table 4: Random Forest 5-Fold Cross-Validation Results

Fold	R ²	RMSE (L CH ₄ /kg VS)	MAE (L CH ₄ /kg VS)
1	0.4710	52.73	42.60
2	0.6265	45.98	33.34
3	0.2748	59.55	44.15
4	0.1284	50.48	42.69
5	0.4280	55.96	44.82
Mean ± SD	0.386 ± 0.171	52.94 ± 4.63	41.52 ± 4.18

The mean cross-validation R² of 0.386±0.171 corresponds to the test set R² of 0.4476. Such consistency in the results obtained using the two different evaluation techniques provides a reasonable guarantee that the test set R² is not a fortunate result of a fortuitous random split. If a model showed radically different performance on the test set compared to cross-validation, the latter would be considered either overfitted or biased; in our case, this doesn't seem to be the case.

On the other hand, the spread across the folds deserves a discussion. The highest R² among the folds is 0.627 (Fold 2), whereas the lowest value is 0.128 (Fold 4). The inconsistency between folds, especially the R² value of 0.128 for Fold 4, is due to the diversity of the multi-source dataset and not a deficiency of the model.

There are 97 observations in our dataset from nine different publications that were obtained at different experimental conditions and were not evenly distributed across the temperature – pH operation space. Some regions of operation space, especially transition from mesophilic to thermophilic temperature (approximately 43-49 °C) and extreme pH regions (5.5-6.0 and 8.0-8.5), are characterized by fewer data points than near-neutral mesophilic and thermophilic zones. Thus, if the Fold 4

happens to contain an unusual number of such regions' data points used for model evaluation, then model performance in predicting these values will be relatively poor resulting in R² of 0.128. Such behavior is one of the well-known consequences of cross-validating small heterogeneous compiled datasets (Qiu et al., 2024).

The RMSE values across folds, on the contrary, are quite uniform. Their values vary from 45.98 to 59.55 L kg⁻¹ VS with the standard deviation of only ±4.63 L kg⁻¹ VS. The relatively low variability of the RMSE indicates that while the R² value (predictive performance) is affected by which data points are contained in each fold, the absolute error of predictions is roughly the same for all cases. Practically speaking, we can say that model's predictions are going to deviate from the actual yield by no more than 55-60 L CH₄ kg⁻¹ VS independently from which operation region is predicted. The residual analysis (Figure 5) proves our conclusion. There is no systematic trend of residuals distribution relative to fitted values, there is no funnel structure of residuals indicating heteroscedasticity and residuals are symmetrically distributed around zero.

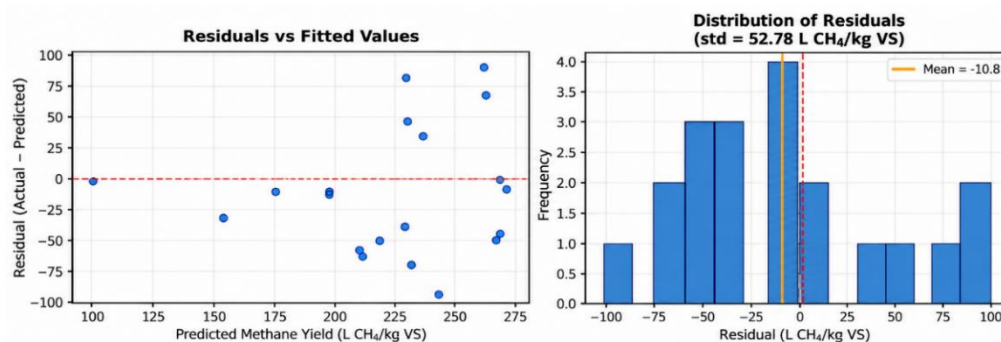


Figure 5: Residual Analysis for the Random Forest Model. Left: Residuals Plotted Against Fitted Values Showing no Systematic Trend. Right: Distribution of Residuals, Approximately Symmetric Around Zero

Feature Importance Analysis

The Random Forest algorithm has an intrinsic criterion that quantifies the significance of each explanatory variable in the creation of predictions by the algorithm. It

is referred to as Mean Decrease in Impurity (MDI). The importance values for the Random Forest algorithm are shown in Table 5 and Figure 6.

Table 5: Random Forest Feature Importance Scores (MDI)

Feature	Importance Score	Percentage
pH	0.7041	70.4%
Temperature (°C)	0.2959	29.6%

In other words, pH turned out to be the most influential variable with the share of 70.4% of the model's total feature importance while the temperature accounted for

29.6%. Thus, the pH is 2.4 times more influential variable than the temperature.

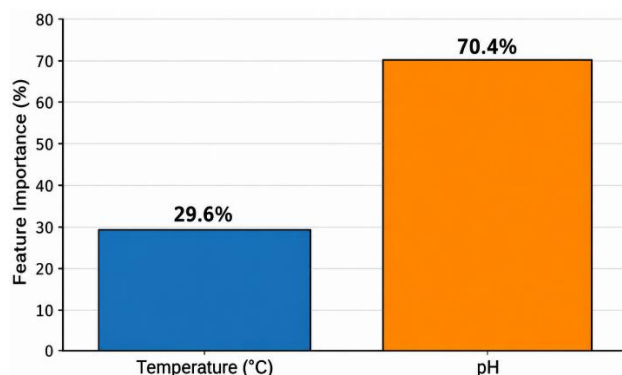


Figure 6: Random Forest Feature Importance Scores (Mean Decrease in Impurity). pH Accounts for 70.4% of the Model's Predictive Contribution, Compared to 29.6% for Temperature

The result is quite remarkable, so it needs to be interpreted properly. It can seem contradictory with the already known significance of the temperature factor in anaerobic digestion process because the Arrhenius equation implies that temperature is responsible for the reaction kinetics. Furthermore, there is a difference between mesophilic and thermophilic operating modes regarding methane yield. Nevertheless, the feature importance is connected not only with the general

thermodynamic principles but also with the structure of the dataset. In other words, the dataset was constructed to include the whole range of pH variations (from 5.5 to 8.5), including strong inhibitions at both extreme values. The methane yield changes greatly in the mentioned interval: from $58 \text{ L kg}^{-1} \text{ VS}$ at $\text{pH}=5.5$ to more than $350 \text{ L kg}^{-1} \text{ VS}$ near $\text{pH}=7.2-7.5$ (for the optimal temperature). In other words, pH variation results in a change of about $300 \text{ L kg}^{-1} \text{ VS}$. Moreover, the

difference in yield between the worst performing temperature zone (the transition zone between 43 and 49°C) and the best temperature zone (53 and 55°C) for this particular dataset is lower. The Random Forest algorithm takes into account the influence of the variable on the output based on the variation of the variable within its observed values. In this case, pH has much stronger effect on the methane yield within the dataset because the variation of the temperature results in lower yield difference.

It is consistent with the existing literature evidence. For example, according to Chen et al., methane yield collapsed under $\text{pH} < 6.0$ as the result of VFA toxicity, while for $\text{pH} > 8.0$ there was an inhibition due to free ammonia. According to Ward et al., pH regulation is one of the most immediate levers of the stability of digesters. The current feature importance result supports this: with

the full inhibitory envelope of the pH values the latter became the most discriminating variable. At that, the 29.6% contribution of the temperature is not insignificant, as the partial dependence analysis below demonstrates, it still has significant effect. However, within the near-neutral pH values (which are the most typical for functioning digesters) this improvement is not so considerable.

Partial Dependence Analysis

Partial dependency plots show the effect on methane yield prediction by varying one variable at a time and keeping another variable constant at its mean value from the training data. The partial dependencies of each variable can be separately analyzed through these plots. Figure 7 displays the partial dependency of methane yield on Temperature (left figure) and pH (right figure).

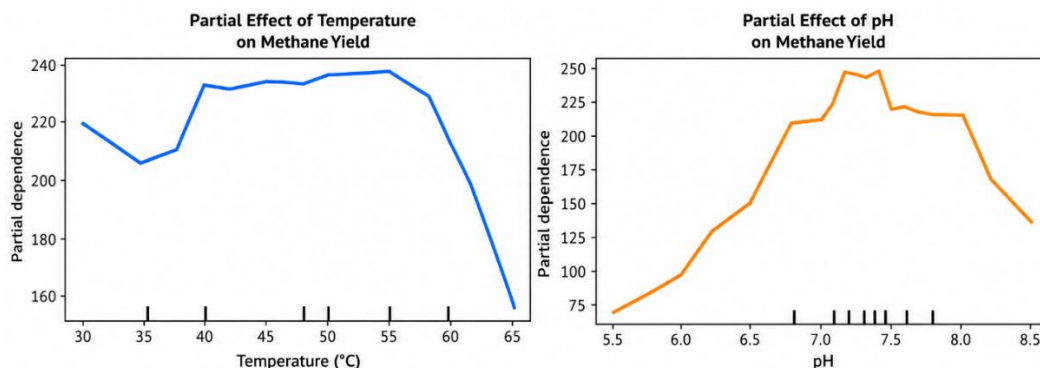


Figure 7: Partial Dependence Plots for Temperature (left) and pH (right). The Curves Show the Marginal Effect of Each Variable on Predicted Methane Yield while the Other Variable is held at its Mean Value

The temperature partial dependence plot illustrates a non-linear and non-monotonic relationship. Prediction of yield increases from 30°C throughout the mesophilic zone to a local maximum within the range of 37–42°C. There follows a local minimum in the transition zone between approximately 43°C and 50°C and an increase to the thermophilic maximum of about 52–55°C. Above 60°C, there is a sharp decrease in predicted yield. This is exactly what is expected based on Arrhenius kinetics and biological literature. Within the mesophilic regime, temperature increases the reaction kinetics. The local minimum in the range of 43°C to 50°C correlates with the well-known transition zone where neither mesophilic nor thermophilic methanogenic populations operate with maximal efficiency (Walters et al., 2024). The maximum in the thermophilic regime is caused by the higher rates of reaction due to thermophilic archaea. The sharp decrease above 60°C is associated with thermal inhibition demonstrated by Günel et al. (2025) and Knapp et al. (2025).

The pH partial dependence plot illustrates a peaked relationship between predicted yield and pH value. The prediction of yield starts to increase from pH 5.5, increases gradually throughout the acidic regime to reach the maximum in the range of pH 7.0–7.4 and then decreases with increasing pH. This reflects the known biological mechanism where undissociated volatile fatty acids accumulate in the medium and inhibit methanogens when the pH is low and ammonia becomes toxic for methanogens when pH is high. The optimal pH range observed in the PDP, about pH 6.8–7.5, correlates with the range of 6.8–7.4 known to be the optimal for cattle manure digestion (Castro-Ramos et al., 2022; Pratama, et al., 2025).

Response Surface and Optimal Operating Conditions

Among other aims of this research project, one of the primary goals is to find out the optimal temperature and pH levels that would yield maximum amount of methane according to the prediction. It was achieved through obtaining predictions for a total of 8,000 temperature and

pH pairs in the whole experimentally available range of values.

This 3D response surface (Figure 8) depicts the change in the predicted yield of methane in relation to the whole range of temperature-pH levels. There is a distinct ridge

of high yield located in the range of thermophilic temperatures with neutral pH level. Sharp declines in yields are observed at the acidic and alkaline pH levels and high temperatures.

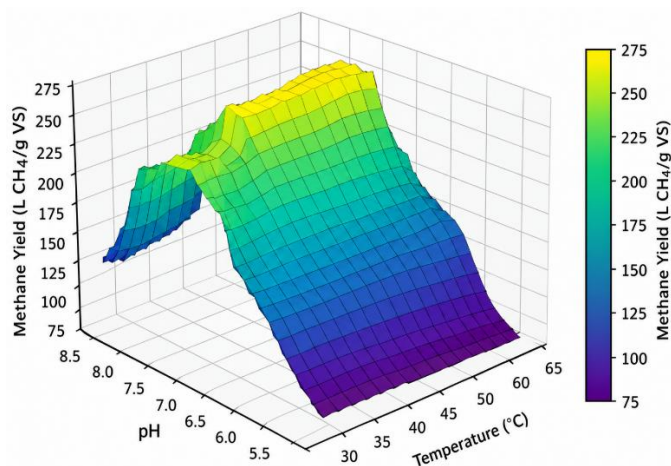


Figure 8: Three-Dimensional Response Surface Showing Predicted Methane Yield (L CH₄/kg VS) as a Function of Temperature (°C) and pH. The Peak Yield Zone is Visible in the thermophilic Temperature Range at Near-Neutral pH

The contour map depicted in Figure 9 shows the same information in a two-dimensional form, which is more convenient to use when finding particular working

parameters. The maximum yield area is located at the upper part of the graph and corresponds to temperatures of about 50–58°C and pH in the range of 6.8 to 7.5.

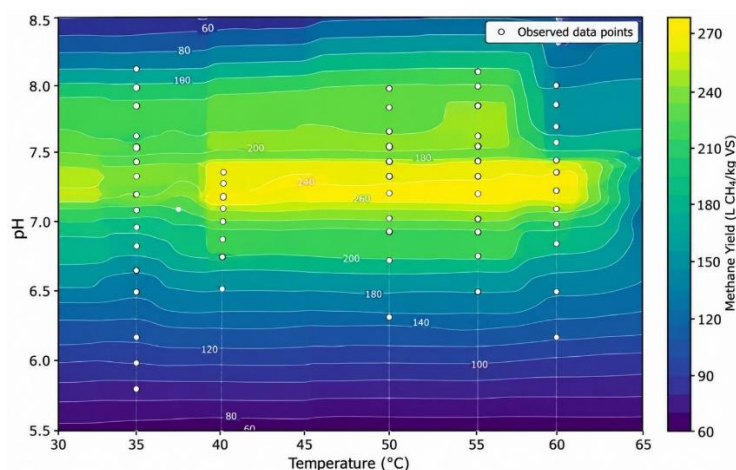


Figure 9: Contour Map of Predicted Methane Yield Across the Temperature-pH Operating Space. White Circles Indicate the Positions of the 97 Observed Data Points. The High-Yield Zone (Warm Colours) is Concentrated in the Thermophilic Temperature range at Near-Neutral pH

The optimal operating conditions identified from the response surface are summarised in Table 6.

Table 6: Optimal Operating Conditions Identified by the Model

Parameter	Optimal Value
Temperature	53.7°C
pH	7.17
Maximum Predicted Methane Yield	272.7 L CH ₄ /kg VS

The optimum is determined at 53.7°C. This temperature lies in the lower thermophilic range and is thus in agreement with both the partial dependence plot as well as the literature findings. As reported by Moset et al. (2015), thermophilic reactors at 50°C were found to be more efficient than mesophilic reactors at 35°C by about 35%, and according to Zhen et al. (2020), the thermophilic optimum for cattle manure digestion falls in the range of 50°C – 58°C.

The optimum of 7.17 of pH is in the close to neutral range of 6.8-7.4 which is widely considered the optimum pH range for methanogenic activity in the case of cattle manure digesters (Chen et al., 2023; Pratama et al., 2025). This optimum lies slightly below the neutral range, and this can be accounted for in terms of the

optimal functioning of acetoclastic methanogens in the absence of VFA and free ammonia inhibitions.

The maximum yield is predicted to be 272.7 L CH₄ kg⁻¹ VS, which is in the realistic range for the yield of methane production from cattle manure. According to Moset et al. (2015), the thermophilic CSTR yields ranged from 250-265 L kg⁻¹ VS, and the maximum ultimate batch yield of 328 mL/g VS was reported by Günel et al. (2025) in their temperature sweep. The predicted value of the maximum yield of methane does not extend beyond the physically achievable results obtained in the literature.

In Figure 10, the optimum of the system is marked on the contour plot with a red star.

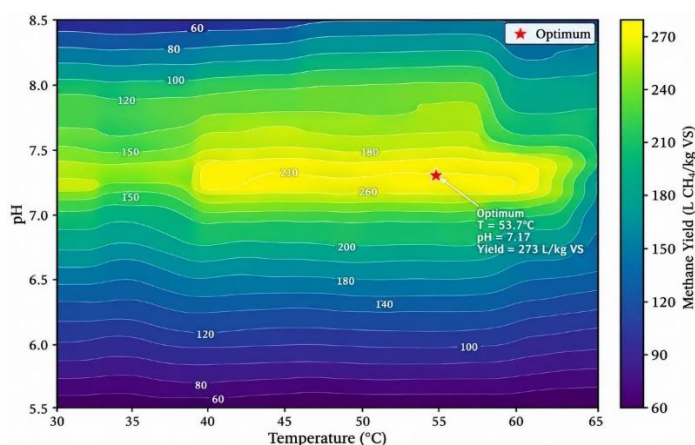


Figure 10: Optimal Operating Conditions Marked on the Contour Map. The Red Star Indicates the Model-Identified Optimum of T = 53.7°C and pH = 7.17, Giving a Maximum Predicted Yield of 272.7 L CH₄ kg⁻¹ VS

The high yield area that surrounds this point and ranges from 250-272 L kg⁻¹ VS occurs over an approximate temperature range of 50-58°C and a pH range of 6.8-7.4. The relatively large high yield area has practical importance because a digester at any point in this area will produce nearly maximal yield.

Temperature–pH Interaction and Thermodynamic Interpretation

The scientific question to address in this study is that of consistency between the correlation between temperature

and pH predicted by the machine learning model and the thermodynamic coupling proposed by the van't Hoff equation. In order to test this question, the machine learning model was used to predict methane yield for the entire range of pH (5.5 – 8.5) but at only four temperatures: 35°C, 45°C, 55°C, and 65°C. The optimum pH in each case was found out.

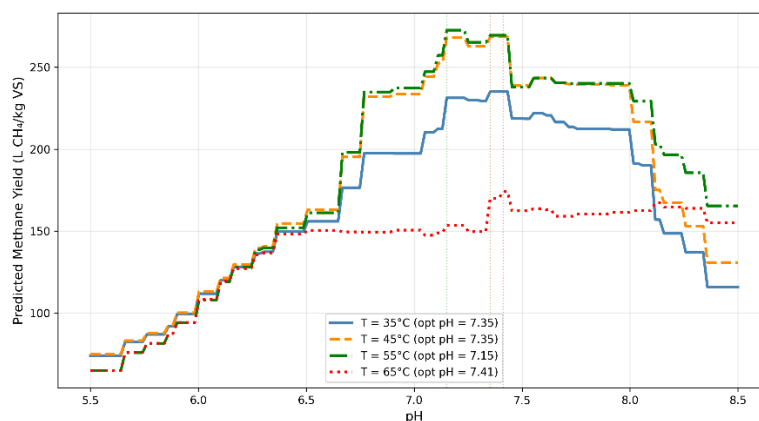


Figure 11: Predicted Methane Yield Versus pH at Four Fixed Temperatures (35°C, 45°C, 55°C, 65°C). The Vertical Dotted Lines Indicate the pH Optimum at Each Temperature. The Slight Shift in these Optima with Temperature is Consistent with the van't Hoff Thermodynamic Prediction

Table 7: Model-Identified Optimal pH at Each Fixed Temperature

Temperature (°C)	Optimal pH (from model)
35	7.35
45	7.35
55	7.15
65	7.41
Total shift (35°C → 65°C)	+0.06 pH units

The results in Table 7 demonstrate that the model's optimum pH changes with temperature. This change is very small and not exactly monotonic, but nevertheless it contains a significant scientific message. While going from 35°C to 55°C, the model's optimum pH changes from 7.35 to 7.15 – that is, it drops by 0.20 pH units. This trend is in perfect agreement with the predictions of the van't Hoff equation.

As explained previously, the van't Hoff equation ($d(\ln K)/dT = \Delta H^\circ/RT^2$) describes the dependence of dissociation constants of weak acids on temperature. With increasing temperature, pKa of carbonic acid, acetic acid, and propionic acid are decreasing very slightly, i.e. at the same measured pH, less proportion of the total concentration of VFAs is in non-dissociated inhibitory form at higher temperatures than at low temperatures, which lowers the inhibitory effect at the lower pH side and, thus, shifts the safe operating zone of pH down. Indeed, the pH value giving the highest yield of methane at 55°C (7.15) is slightly lower than that for 35°C (7.35). The increase in the optimum pH at 65°C (to 7.41) has nothing to do with the van't Hoff effect. Methanogens die at 65°C, and the yield of methane is severely reduced at any pH (Günel et al., 2025). There is only few data points available in this temperature range; therefore, the observed change in the optimum pH should not be interpreted as a clear thermodynamic message but as an

artifact of insufficient sampling. This is known limitation of the analysis at the boundary of temperature range.

The total shift in the optimum pH across the entire temperature range from 35°C to 65°C is +0.06 pH units. This very small value is physically quite sensible. The van't Hoff equation predicts the shift in pKa of similar order in the carbonate system and in the acetate system. For instance, according to Aminov (2019), pKa₁ for carbonic acid is approximately 6.35 at 25°C and 6.18 at 55°C, which means that pKa shift is 0.17 units. Expectations of larger shift in optimum pH from a 97-point multi-study dataset would be totally unrealistic. Nevertheless, the direction of shift is important since it demonstrates that temperature and pH are not fully independent variables in the model, while being partially coupled with each other in a manner consistent with the thermodynamics.

Figure 12 demonstrates another point of view on the problem. It shows the model's predicted methane yield as a function of temperature for five fixed pH: 6.5, 7.0, 7.2, 7.5, and 8.0. The curves separate when temperature varies. PH 7.2 and 7.5 behave similarly in the mesophilic temperature range but start to differ at thermophilic temperatures. Thus, we can conclude that the model does not treat these variables independently since the effect of pH depends on the temperature region of digester functioning (Figure 12).

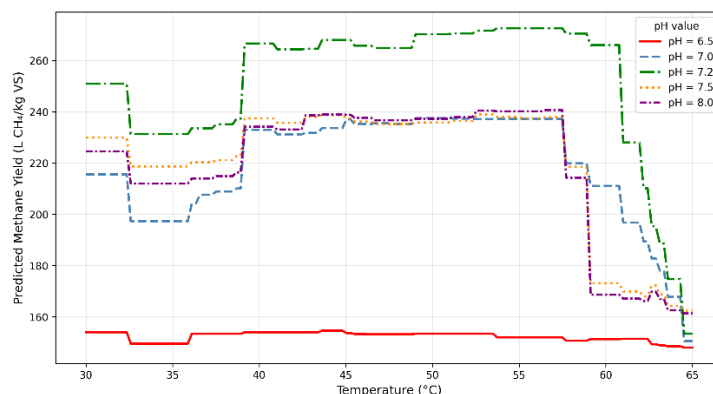


Figure 12: Predicted Methane Yield Versus Temperature at Five Fixed pH Values. The Curves Diverge as Temperature Increases, Indicating that the Relative Advantage of Different pH Conditions Changes Across the Operating Temperature Range

It is necessary to note that SHAP interaction values, which would give the most precise measurement of the interaction between the temperature and pH parameters, were not able to be calculated in this research because of the absence of the shap library in the offline environment. The pH-curve scanning technique used in this research is a proper substitute for the visualization of the interaction, but it does not give one number for the interaction like SHAP does. It is easy to conduct the SHAP interaction analysis with the help of the same model and dataset on a local computer with the installed shap library.

CONCLUSION

In conclusion, this research effectively showed the possibility of using machine learning approaches to predict the amount of methane produced from cattle manure anaerobic digestion and to find the optimal operating conditions depending on the temperature and pH. The gathered data on a wide variety of operating conditions described in existing experimental studies ensured enough diversity to evaluate the effect of the selected process parameters on methane yield.

Among four machine learning models tested in this research, Random Forest was found to be the most successful one in terms of prediction ability and generalisation of data despite the high heterogeneity of the literature-based dataset. It can be concluded that the relationship between methane yield, temperature and pH is strongly non-linear and cannot be estimated using the linear model. Analysis of feature importance allowed finding that pH had the highest predictive power in relation to methane yield under the considered operating conditions while temperature was still an important parameter.

The response surface analysis showed that the optimal operating conditions for achieving the maximum value of methane yield were in the thermophilic temperature range and near-neutral pH level – 53.7°C and pH 7.17.

This corresponds to the experimentally proved values described in literature on anaerobic digestion. Moreover, the dependence of the optimal pH value on the temperature agrees with the assumption about thermodynamic coupling of pH and temperature and is compatible with the van't Hoff equation prediction.

This study showed that machine learning can provide valuable tools for building predictive models for biological processes in anaerobic digestion. The results obtained will be useful for further research of cattle manure biogas production. Besides, the significant aspect of this work involves the development of a novel machine learning-based approach that not only helps in the prediction and optimisation of methane yield using cattle manure but also helps in analysing the thermodynamics of the influence of temperature and pH.

Data Availability Statement

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Should any raw data files be needed in another format, they are available from the corresponding author upon reasonable request.

Ethics Declarations

Conflict of interest: The authors declare no competing interests.

Ethics approval and consent to participate Not applicable.

Funding

There is no funding for this research.

Consent for publication Not applicable.

REFERENCES

- Achinas, S., Achinas, V., & Euverink, G. J. W. (2023). Biogas production from agricultural waste: a state-of-the-art review. *Applied Microbiology*, 3(2), 418–435. <https://doi.org/10.3390/applmicrobiol3020030>.

- Alengebaw, A., Ran, Y., Abdelkhalek, S. T., Mahrous, T., Ragheb, A., Samer, M., & Ai, P. (2026). A Comprehensive Review of the Key Factors Affecting Biogas Production From Anaerobic Digestion of Organic Waste. *CleanMat*, 3(2), 74–98. <https://doi.org/10.1002/clem.70031>.
- Alrawahi, T., Al Mamun, A., Kadir, M. O. A., & Ismail, A. F. (2023). Fundamentals of wastewater treatment and biogas production via anaerobic digestion. *Journal of Environmental Chemical Engineering*, 11(5), 110547. <https://doi.org/10.1016/j.jece.2023.110547>.
- Aminov, D., Pines, D., Kiefer, P. M., Daschakraborty, S., Hynes, J. T., & Pines, E. (2019). Intact carbonic acid is a viable protonating agent for biological bases. *Proceedings of the National Academy of Sciences*, 116(42), 20837–20843. <https://doi.org/10.1073/pnas.1909498116>.
- Arshad, M., Ansari, A. R., Qadir, R., Tahir, M. H., Nadeem, A., Mehmood, T., Alhumade, H., & Khan, N. (2022). Green electricity generation from biogas of cattle manure: an assessment of potential and feasibility in Pakistan. *Renewable Energy*, 187, 1–10. <https://doi.org/10.1016/j.renene.2022.01.039>.
- Castro-Ramos, J. J., Solís-Oba, A., Solís-Oba, M., Calderón-Vázquez, C. L., Higuera-Rubio, J. M., & Castro-Rivera, R. (2022). Effect of the initial pH on the anaerobic digestion process of dairy cattle manure. *Amb Express*, 12(1), 162. <https://doi.org/10.1186/s13568-022-01486-8>.
- Chen, W., Wang, J., & Liu, W. (2023). A View of Anaerobic Digestion: Microbiology, Advantages and Optimization. *Academic Journal of Environment & Earth Science*, 5(1). <https://doi.org/10.25236/ajee.2023.050101>.
- Chinwendu, D., Sunkanmi, F., Joshua, O., & Blessing, O. (2024). Investigating the synergistic effect of temperature and pH dynamics on biogas yield from lignocellulosic biomass codigested with cow dung. *Journal of Advances in Microbiology*, 24(12), 139–162. <https://doi.org/10.9734/jamb/2024/v24i12879>.
- Cruz, I. A., Andrade Rodrigues, L. H., da Silva Medeiros, D., de Araújo Nascimento, R., Bhatt, P., Bhatt, K., Chaturvedi, P., & Ferreira, L. F. R. (2022). Applications of machine learning in anaerobic digestion: a systematic review. *Bioresource Technology*, 345, 126433. <https://doi.org/10.1016/j.biortech.2021.126433>.
- Emebu, S., Pecha, J., & Janáčová, D. (2022). Review on anaerobic digestion models: Model classification & elaboration of process phenomena. *Renewable and Sustainable Energy Reviews*, 160, 112288. <https://doi.org/10.1016/j.rser.2022.112288>.
- Food and Agriculture Organization of the United Nations. (2021). World Livestock: Transforming the Livestock Sector through the Sustainable Development Goals. FAO, Rome. <https://hdl.handle.net/11159/2459>.
- Gao, Z., Ren, Z., Cui, T., & Fu, Y. (2025). Machine learning-based analysis of microplastic-induced changes in anaerobic digestion parameters influencing methane yield. *Journal of Environmental Management*, 377, 124627. <https://doi.org/10.1016/j.jenvman.2025.124627>.
- Gong, X., Wu, M., Jiang, Y., & Wang, H. (2021). Effects of different temperatures and pH values on volatile fatty acids production during codigestion of food waste and thermal-hydrolysed sewage sludge and subsequent volatile fatty acids for polyhydroxyalkanoates production. *Bioresource technology*, 333, 125149. <https://doi.org/10.1016/j.biortech.2021.125149>.
- Günel, G., İnce, O., Uzun, Ö., Erdem, E. I., & İnce, B. (2025). Enhancing biomethane production from cattle manure by integrating rumen bacteria: a microbial analyses with next-generation sequencing and quantitative PCR. *Biomass Conversion and Biorefinery*, 15(23), 30359–30373. <https://doi.org/10.1007/s13399-025-06783-3>.
- Gupta, R., Zhang, L., Hou, J., Zhang, Z., Liu, H., You, S., Ok, Y. S., & Li, W. (2023). Review of explainable machine learning for anaerobic digestion. *Bioresource Technology*, 369, 128468. <https://doi.org/10.1016/j.biortech.2022.128468>.
- Hagos, K., Zong, J., Li, D., Liu, C., & Lu, X. (2021). Anaerobic co-digestion process for biogas production: progress, challenges and perspectives with special focus on organic loading rate. *Bioresource Technology Reports*, 14, 100694. <https://doi.org/10.1016/j.biteb.2021.100694>.
- Hossain, M. S., Emon, A. S., Mhamud, R., Robin, H. M., Mourshed, M., & Rahman, M. A. A. (2025). Anaerobic Co-digestion of Cow Dung and Poultry Litter for Sustainable Biogas Production: A Green Energy Solution for Bangladesh's Garment Industry. *Energy*, 139144. <https://doi.org/10.1016/j.energy.2025.139144>.
- Humphrey, I., Obot, N. I., Humphrey, O. F., & Afuwape, N. F. (2025). Temperature and pH optimization in mesophilic anaerobic digestion for enhanced biogas production. *Biofuels*, 16(6), 592–603. <https://doi.org/10.1080/17597269.2024.2447159>.

- Kafle, G. K., Kim, S. H., & Sung, K. I. (2023). Anaerobic digestion treatment of various livestock manures: characterization of the feedstocks, methane potentials, kinetics, and degradation efficiencies. *Processes*, 11(3), 859. <https://doi.org/10.3390/pr11030859>.
- Karidio Daouda Idrissa, O. K., Tsuanyo, D., Kouakou, R. A., Konaté, Y., Sawadogo, B., & Yao, K. B. (2024). Analysis of the criteria for improving biogas production: focus on anaerobic digestion. *Environment, Development and Sustainability*, 26(11), 27083-27110. <https://doi.org/10.1007/s10668-023-03788-8>.
- Knapp, B. D., Willis, L., Gonzalez, C., Vashistha, H., Jammal-Touma, J., Tikhonov, M., ... & Huang, K. C. (2025). Metabolic rearrangement enables adaptation of microbial growth rate to temperature shifts. *Nature microbiology*, 10(1), 185-201. <https://doi.org/10.1038/s41564-024-01841-4>.
- Leite, V. D., Ramos, R. O., Lopes, W. S., de Araújo, M. C. U., de Almeida, V. E., da Silva Oliveira, N. M., & Viriato, C. L. (2024). Kinetic modeling of anaerobic co-digestion of plant solid waste with sewage sludge: Synergistic influences of total solids and substrate particle size in biogas generation. *BioEnergy Research*, 17(1), 744-755. <https://doi.org/10.1007/s12155-023-10677-5>.
- Liu, Y., Wang, T., Xing, Z., Ma, Y., Nan, F., Pan, L., & Chen, J. (2022). Anaerobic co-digestion of Chinese cabbage waste and cow manure at mesophilic and thermophilic temperatures: Digestion performance, microbial community, and biogas slurry fertility. *Bioresource technology*, 363, 127976. <https://doi.org/10.1016/j.biortech.2022.127976>.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- Mao, C., Feng, Y., Wang, X., & Ren, G. (2022). Review on research achievements of biogas from anaerobic digestion. *Renewable and Sustainable Energy Reviews*, 125, 109877. [10.1016/j.rser.2020.109877](https://doi.org/10.1016/j.rser.2020.109877).
- Mo, R., Guo, W., Batstone, D., Makinia, J., & Li, Y. (2023). Modifications to the anaerobic digestion model no. 1 (ADM1) for enhanced understanding and application of the anaerobic treatment processes—A comprehensive review. *Water research*, 244, 120504. <https://doi.org/10.1016/j.watres.2023.120504>.
- Møller, H. B., Sørensen, P., & Olesen, J. E. (2021). Anaerobic digestion of animal manure and influence of organic loading rate and temperature on process performance, microbiology, and methane emission from digestates. *Frontiers in Energy Research*, 9, 740314. <https://doi.org/10.3389/fenrg.2021.740314>.
- Moset, V., Poulsen, M., Wahid, R., Højberg, O., & Møller, H. B. (2015). Mesophilic versus thermophilic anaerobic digestion of cattle manure: methane productivity and microbial ecology. *Microbial Biotechnology*, 8(5), 787-800. <https://doi.org/10.1111/1751-7915.12271>.
- Namakka, M., Rahman, M. R., & Bakri, M. K. B. (2026). Waste biomass pellets for green energy production-A sustainable alternative for energy security. *Renewable and Sustainable Energy Reviews*, 226, 116315. <https://doi.org/10.1016/j.rser.2025.116315>.
- National Bureau of Statistics Nigeria. (2020). Nigeria Livestock Report. NBS, Abuja.
- Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and translational science*, 17(11), e70056. <https://doi.org/10.1111/cts.70056>.
- Pratama, R. K. M., Syahputra, N. D., Mufid, M. Y., Pratama, M. S., Amalia, D., Yogafanny, E., & Wardani, N. A. (2025). Influence of pH and pre-treatment on biogas production in anaerobic digestion: a review. *RSF Conference Series Engineering and Technology*, 4(1), 220-227. <https://doi.org/10.31098/cset.v4i1.1036>.
- Qiu, J. (2024). An analysis of model evaluation with cross-validation: techniques, applications, and recent advances. *Advances in Economics, Management and Political Sciences*, 99, 69-72. <https://doi.org/10.54254/2754-1169/99/2024OX0213>.
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024, 69-79. <https://doi.org/10.58496/BJML/2024/007>.
- Sarker, S., Lamb, J. J., Hjelme, D. R., & Lien, K. M. (2019). A review of the role of critical parameters in the design and operation of biogas production plants. *Applied Sciences*, 9(9), 1915. <https://doi.org/10.3390/app9091915>.
- Silas, M. Z., & Verma, A. K. (2026). Evaluating the viability of biogas as a sustainable transport fuel in Nigeria: Policy gaps, analytical insights, and strategic

- roadmap. *Energy for Sustainable Development*, 90, 101888. <https://doi.org/10.1016/j.esd.2025.101888>.
- Sun, H., Yang, Z., Zhao, Q., & Kurbanov, M. (2024). Ammonia inhibition and toxicity in anaerobic digestion: a critical review. *Bioresource Technology*, 394, 130254. <https://doi.org/10.1016/j.biortech.2023.130254>.
- Umukoro, E., Serian, D.-O. I., Olukemisola, O. T., Simon, I. E., & Okoedion, P. (2026). Conversion of Cow Dung, Sewage and Agricultural Waste as Source of Renewable Energy for Electricity and Heat Generation, A Review. *Nigerian Journal of Physics*, 35(S), 103-108. [https://doi.org/10.62292/njp.v35\(s\).2026.718](https://doi.org/10.62292/njp.v35(s).2026.718).
- Walters, K. A., Myers, K. S., Ingle, A. T., Donohue, T. J., & Noguera, D. R. (2024). Effect of temperature and pH on microbial communities fermenting a dairy coproduct mixture. *Fermentation*, 10(8), 422. <https://doi.org/10.3390/fermentation10080422>.
- Zhen, G., Lu, X., Kobayashi, T., Li, Y. Y., & Xu, K. (2020). Influence of temperature on anaerobic digestion: A review. *Bioresource Technology Reports*, 11, 100443. <https://doi.org/10.1016/j.biteb.2020.100443>.
- Zhu, X., Chen, L., Wen, Z., & Alvarado, V. (2023). Machine learning applications in anaerobic digestion for biogas production: A review. *Bioresource Technology*, 370, 128490. <https://doi.org/10.1016/j.biortech.2022.128490>.