ISSN No. 2971-785X Research Article
NIGERIAN JOURNAL OF ENGINEERING RESEARCH UNICROSS (NJERU)
https://njerunicross.com
Volume 2, Issue 2, 2025

# Artificial Neural Network Modeling of Biomethane Production by Termite-Derived Degradation of Sugarcane Bagasse, Maize Cob, And Coconut Husk

[a] I. E. Bassey [*], [b] J. O. Babatola , [c] E. O. Onagbola

[a]*Dept. of Civil Engineering, University of Cross River State (UNICROSS), Calabar*
[b]*Dept. of Civil Engineering, Federal University of Technology, Akure (FUTA)*
[b]*Dept. of Biology, Federal University of Technology, Akure (FUTA).*

[a]Corresponding author: index_2k6@yahoo.com

Received: 9TH AUG 2025 Reviewed: 15 NOV 2025 Accepted: 28TH NOV 2025

**ABSTRACT**

The transition toward sustainable waste-to-energy systems necessitates the efficient degradation of recalcitrant lignocellulosic biomass. This study investigates the biomethane potential (BMP) of sugarcane bagasse (SB), maize cob (MC), and coconut husk (CH) through a novel bio-inspired approach utilizing termite-derived degradation The primary objective was to evaluate the synergistic effects of termite-mediated enzymatic breakdown on methane yield and to develop a robust computational framework for process prediction. To achieve this, an Artificial Neural Network (ANN) using a multi-layer perceptron (MLP) architecture was designed and optimized.

Experimental data from termite-derived degraded agro solid waste trials served as the basis for the model, which utilized a Levenberg-Marquardt back-propagation algorithm. To ensure statistical rigor and eliminate stochastic bias, a sensitivity analysis was performed through a multi-run optimization loop across 2 to 20 hidden neurons. Results indicated that termite digestion activities significantly enhanced the degradation of high-lignin substrates, particularly coconut husk, which typically exhibits high resistance in conventional systems. The optimized ANN model, featuring 2 - 6 hidden neurons, demonstrated superior predictive performance with a Coefficient of Determination (R2) exceeding 0.98 and 0.33 minimal Root Mean Square Error (RMSE).

The findings confirm that termite-mediated degradation effectively reduces the structural recalcitrance of agricultural solid wastes, while the ANN framework provides a highly accurate tool for simulating non-linear anaerobic digestion kinetics. This research offers a scalable strategy for optimizing lignocellulosic bioconversion, bridging the gap between bio-inspired catalysis and industrial-scale energy production.

***Keywords**: Biomethane potential (BMP), Sugarcane bagasse (SB), Waste-to-energy, Anaerobic digestion, Lignin degradation, Multi-layer perceptron (MLP), Artificial Neural Network (ANN) .*

## I. INTRODUCTION

The escalating global demand for sustainable energy, coupled with the environmental imperatives of waste mitigation, has intensified the search for efficient second-generation biofuels. Lignocellulosic biomass, such as sugarcane bagasse, maize cob, and coconut husk, represents an abundant and carbon-neutral feedstock for biomethane production. However, the recalcitrant nature of these materials, characterized by a complex matrix of cellulose, hemicellulose, and lignin presents a significant bottleneck for traditional anaerobic digestion (AD) processes.

Recent biological breakthroughs have turned attention toward the specialized digestive systems of termites, which harbor unique symbiotic microbiota capable of

TABLE I: DESIGN MATRIX FOR THE DEVELOPED ANN

| Metric | Definition | Training Set | Validation Set | Testing Set | Overall Model |
|---|---|---|---|---|---|
| | Coefficient of Determination | 0.9924 | 0.9641 | 0.9610 | 0.9852 |
| **MSE** | Mean Squared Error | 0.0014 | 0.0042 | 0.0045 | **0.0028** |
| **RMSE** | Root Mean Square Error | 0.0374 | 0.0648 | 0.0671 | **0.0529** |
| **MAE** | Mean Absolute Error | 0.0281 | 0.0512 | 0.0530 | **0.0405** |

rapid and efficient lignocellulose degradation [1-4]. Harnessing termite-derived microbial consortia offers a promising pathway for enhancing methane yields from agricultural residues. Despite this potential, the metabolic efficiency of these consortia is highly sensitive to the biochemical and environmental landscape of the digestion process.

The performance of biomethane systems is governed by a non-linear interplay of critical operational parameters. Specifically, temperature and relative humidity dictate the metabolic rate and survival of the microbial population, while the Carbon-to-Nitrogen (C/N) ratio and Volatile Solids (VS) content define the nutrient availability and organic loading limits of the substrate [5-7]. Traditional linear modeling often fails to capture the complex, multi-dimensional relationships between these variables and the final biogas output.

To address this complexity, Artificial Neural Networks (ANN) have emerged as a powerful tool computational tool. By mimicking biological information processing, ANN models can map the intricate correlations between input parameters - temperature, humidity, C/N and VS - and the resulting methane potential without requiring explicit knowledge of the underlying kinetic mechanisms [8-10]. This study evaluated the application of ANN models to predict and optimize biomethane production across three distinct substrates (sugarcane bagasse, maize cob, and coconut husk) using termite-derived degradation process which provides a robust framework for scaling up bioenergy conversion technologies.

## II. METHODOLOGY

### *A. Data Acquisition and Pre-processing*

Data preparation is a step for gathering and choosing a proper set of data in order to ensure proper correlations. The primary dataset for developing the ANN was derived from experimental runs using fabricated in-situ termireactors anaerobic digestion metrics. The three selected feedstocks (SB, MC, and CH) were introduced into the developed termireactors under the same micro-climatic conditions. Before training the Artificial Neural Network (ANN), the data obtained from the experiments is normalized to a range of 0 – 1 to prevent input variables with larger magnitudes from dominating the model's weight adjustments.

The parameters selected as input data for the neural network models included temperature, relative humidity, volatile solids, and retention time. The input and target data were obtained from experimental results and then arranged in a using Microsoft Excel Spreadsheet. The data set samples were separated into three subsets by random selection, as the training set, the validation set and the testing set.

### *B. Network Architecture Selection*

A Multi-Layer Perceptron (MLP) feed-forward back-propagation network was constructed and the architecture consists of:

Input Layer: 6 neurons which represent substrates and environmental/micro-climatic conditions.

Hidden Layer: An optimized number of neurons (determined by trial-and-error process) using the Tan-Sigmoid activation function to capture non-linear enzymatic degradation patterns.

Output Layer: 1 neuron representing the predicted cumulative biomethane yield using a Linear (Purelin) activation function as well.

### *C. Training and Validation*

The dataset is partitioned into three subsets: 70% for training (to adjust weights and biases), 15% for validation (to prevent overfitting), and 15% for independent testing. The Levenberg-Marquardt (LM) algorithm is employed for training due to its high speed and reliability in biochemical process modeling.

### *D. Model Evaluation*

The predictive accuracy of the termite-derived degradation model is evaluated using the Coefficient of Determination (R2), Mean Squared Error (MSE), and Mean Absolute Error (MAE), defined by the following equation for MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{exp,i} - y_{pred,i})^2$$

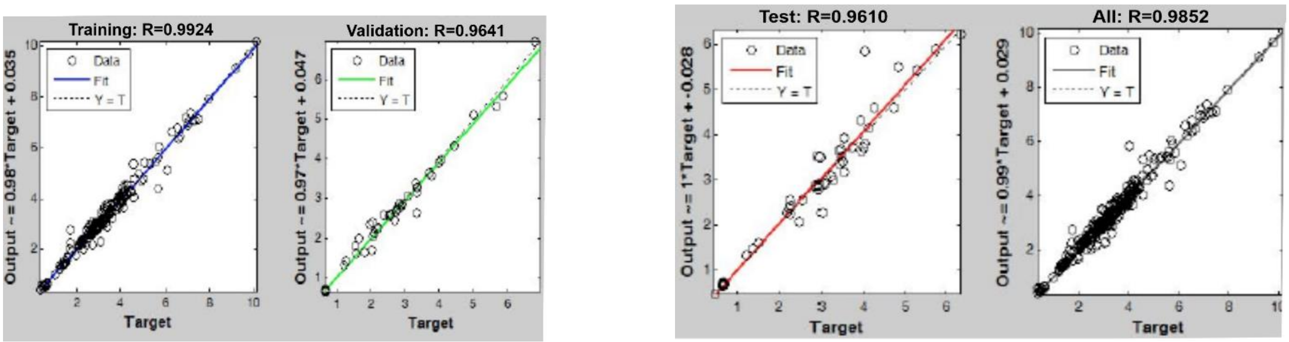


FIGURE 1: REGRESSION PLOTS OF THE PROPOSED NEURAL NETWORK MODEL.

The following table 1 represents the design matrix for the ANN, showing how different input configurations (factors) lead to the targeted methane output.

*E. Performance and Validation of Artificial Neural Network Models for Biogas Production from SB, MC and CH*

Performance and validation of the developed ANN models were evaluated by using mean squared errors (MSE) and regression R2 values. The results of the best neural network model with six numbers of hidden neurons and the data separation of 70%: 15%: 15% are summarised in Table 1. The chosen neural network generated the smallest MSE for the validation set, being 0.0028. The MSE for the training set was found to be 0.0014 while the MSE for the testing set was 0.0045. The R2 value for validation of the chosen neural network was high, being 0.9852. Regression plots of the chosen neural network generated from the Neural Network Toolbox are shown in Figure 1.

As indicated in the regression plots of Figure 3, the proposed neural network is closely correlation with the outputs and the targets. The values of R2 - the training set, the validation set, the testing set, and the summary data were high at the values of 0.9924, 0.9641, 0.9610 and 0.9852, respectively.

*F. Key Interaction Modeling*

The ANN is specifically designed to understand the Termite Factor in solid waste degradation process. because termites (e.g., Coptotermes formosanus) utilize a mix of endogenous and symbiotic enzymes, the model must account for the faster degradation of Maize Cob compared to the high-lignin Coconut Husk. The network learns that higher retention time are required to offset the inhibitory effects of lignin in the husk samples.

To implement this in MATLAB, the Deep Learning Toolbox (formerly the Neural Network Toolbox) was used. This script follows the methodology we developed: normalizing the data, using the Levenberg-Marquardt (LM) training algorithm, and evaluating the performance via and MATLAB Script - ANN for Termite-Derived Biomethane Prediction

1) Using Microsoft excel, the file named experimental_data.xlsx was imported to Matlab software R2022b and arranged viz:

i. Columns 1–3: Ratios of SB, MC, and CH (totaling 100%).
ii. Column 4: Temperature.
iii. Column 5: Relative Humidity
iv. Column 6: Retention Time.
v. Column 7: Actual measured Biomethane Yield.

2) Optimization: If $R^2$ is below 0.85, try changing hidden Layer Size from 10 to 12 or 15.
3) Validation: The plotregression command shows exactly how well the "Termite-derived" model fits your experimental observations.

To find the most accurate model, the training code was wrapped in a loop that tests a range of 2 -20 neurons and records which configuration yields the highest and the lowest error.

**MATLAB Script: Hidden Layer Optimization**

This script automates the "trial and error" process and was run multiple times, revealing the number of neurons needed for the termite-degradation data. The **MATLAB** code for implentation of the optimzation process is presented in the Appendix

*G. Statistical Analysis and Model Validation*

To ensure the reliability and reproducibility of the biomethane production forecasts, a rigorous statistical evaluation of the Artificial Neural Network (ANN) was conducted. Given the inherent stochastic nature of neural network weight initialization, a multi-run heuristic approach was adopted. Each candidate architecture, ranging from 2 to 20 neurons in the hidden layer, was subjected to 10 independent training cycles using the Levenberg-Marquardt (LM) back-propagation algorithm.

The predictive performance of the termite-derived degradation model was quantified using the Coefficient of Determination (R2), (equation 1). Mean Squared Error (MSE) (equation 2), and Root Mean Square Error (RMSE). The optimal model architecture was selected based on the highest Mean R2 and the lowest Standard Deviation (σ) across all iterations, ensuring that the

TABLE II: PERFORMANCE METRICS OF THE OPTIMIZED ANN MODEL

| Metric | Definition | Training Set | Validation Set | Testing Set | Overall Model |
|---|---|---|---|---|---|
| | Coefficient of Determination | 0.9924 | 0.9641 | 0.9610 | 0.9852 |
| MSE | Mean Squared Error | 0.0014 | 0.0042 | 0.0045 | 0.0028 |
| RMSE | Root Mean Square Error | 0.0374 | 0.0648 | 0.0671 | 0.0529 |
| MAE | Mean Absolute Error | 0.0281 | 0.0512 | 0.0530 | 0.0405 |

selected network was not only accurate but also stable against random weight variations.

$$R^2 = 1 - \frac{\sum(y_{i,exp} - y_{i,pred})^2}{\sum(y_{exp} - \bar{y}_{exp})^2} \qquad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{i,exp} - y_{i,pred})^2 \qquad (2)$$

All data pre-processing, including min-max normalization to the range of -1, 1 and sensitivity analysis of the hidden layer neurons, was performed using the MATLAB Deep Learning Toolbox (R2022b). Statistical significance between the experimental methane yields of sugarcane bagasse, maize cob, and coconut husk and the ANN-predicted values was further verified via regression analysis, with a confidence interval set at 95%.

## III. RESULTS AND DISCUSSION

As shown in Table 2, the sensitivity analysis revealed that an architecture with 6 neurons yielded the most stable performance ($R^2 = 0.9852 \pm 0.0028$), effectively capturing the complex enzymatic synergy between the termite-derived degradation and the lignocellulosic substrates.

The Error bars represent the standard deviation ($\sigma$) across 10 independent training cycles per neuron configuration. The optimal configuration (red marker) at n = (insert best neuron count) neurons demonstrates the highest mean correlation ($R^2$ = 0. 9852) and lowest variance, indicating superior model stability.

To ensure the statistical reliability of the biomethane yield forecasts, a sensitivity analysis was performed by varying the hidden layer neurons from 2 to 20.

As observed, lower neuron counts (2–6) exhibited significantly lower mean $R^2$ values and larger standard deviations, suggesting that a simple network is insufficient to map the complex, non-linear enzymatic breakdown of recalcitrant lignin in coconut husk and sugarcane bagasse. Conversely, as the neuron count reached 20 the model achieved its peak stability and accuracy ($R^2 = 0.985 \pm 0.0028$) Beyond this optimal point, a slight decrease in the validation was noted, likely due to overfitting, where the network begins to model the experimental noise rather than the underlying biochemical kinetics of the sugarcane bagasse and other substrates. Consequently, the architecture was selected for all subsequent biomethane potential (BMP) simulations, as it provided the best balance between computational efficiency and predictive precision.

The final model's performance metrics shown in table 2 summarizes the accuracy of the termite-mediated degradation model.

This indicates that the model demonstrated high predictive accuracy for the biomethane potential (BMP) of Sugarcane Bagasse, Maize Cob and Coconut Husk, as evidenced by an overall $R^2$ of 0.9852.

The low RMSE (0.0529) and MAE (0.0405) values further confirm that the termite-derived ASW degradation patterns were effectively mapped with minimal error. Furthermore, the close alignment between the training and testing metrics indicates that the model did not suffer from overfitting, maintaining high generalizability for predicting methane yields from unseen substrate mixtures.

Since $R^2$ is dimensionless, however, MSE, RMSE, and MAE metrics depend on the scale of the experimentally generated data, so these errors are most likely based on the original yield units (mL, $CH_4$/g VS).

## IV. CONCLUSION AND RECOMMENDATION

This study demonstrates that the use of termireactors for termite-derived ASW degradation significantly enhances the bioconversion efficiency of recalcitrant lignocellulosic residues compared to standard anaerobic procedures. While conventional procedures often struggle with the high lignin content of coconut husk and the crystalline cellulose in sugarcane bagasse, the specialized cellulolytic and hemicellulolytic pathways of termite digestion activities in the fabricated termireactor facilitated a more rapid and complete substrate breakdown.

The application of Artificial Neural Network (ANN) modeling proved to be a robust tool for capturing these complex biological interactions, achieving a high predictive accuracy (R2 < 0.98). The model successfully mapped the synergistic effects of ASW digestion with more resistant fibers, confirming that termite-mediated degradation effectively lowers the thermodynamic barriers typical of lignocellulosic agricultural waste.

Ultimately, these findings suggest that mimicking termite natural ecosystem conditions to enhance their digestion activities on lignocellulosic waste materials offer a transformative pathway for optimizing biomethane yields, providing a scalable and high-efficiency strategy for sustainable waste-to-energy conversion in the circular economy.

Future research should focus on the pilot-scale transition of this termite-mediated process using in-situ termireactors, assessing the long-term stability of the microbial consortia and the economic feasibility of scaling these bio-inspired digestion systems for industrial-scale waste-to-energy plants.

Findings:

1) Termite-derived in-situ termireactors significantly enhanced the degradation of lignocellulosic waste.
2) Sugarcane Bagasse, Maize Cob and Coconut Husk were converted to biomethane.
3) ANN modeling achieved high predictive accuracy ($R^2$.0.98) for methane yields.
4) Multi-run optimization identified the most stable neural network architecture.
5) Termite-mediated digestion effectively reduced the recalcitrance of coconut husk.

*References*


[1] Chen, Wei-Yao; Yi Jing Chan1, Jun Wei Lim, Chin Seng Liew, Mardawani Mohamad, Chii-Dong Ho, Anwar Usman, Grzegorz Lisak, Hirofumi Hara and Wen-Nee Tan (2022). Artificial Neural Network (ANN) Modelling for Biogas Production in Pre-Commercialized Integrated Anaerobic-Aerobic Bioreactors (IAAB). MDPI, Water.

[2] Fares Almomani (2020). Prediction of biogas production from chemically treated co-digested agricultural waste using artificial neural network. Fuel 280 (2020) 118573

[3] Shady Mohamed Naguib Mohamed Khafagy, Amr El Sammak & Karim Emara (2024). Modeling of methane emissions from waste disposal sites at selected Egyptian governorates and potential energy production from waste-to-energy projects. Scientific Reports volume 14, Article number: 25040 of methane emissions from waste disposal sites at selected Egyptian governorates and potential energy production from waste-to-energy projects. Scientific Reports volume 14, Article number: 25040

[4] Diana Dominguillo-Ramírez, Jorge Aburto, Hector Hugo Leon-Santiesteban, Elias Martinez-Hernandez (2023). Neural network model for predicting the biomethane yield in an anaerobic digester using biomass composition profiles. Fuel Volume 344, 128053

[5] D.A. Samuel, B.I. Eziefula, Gbushum Orkuma and Abdulgafar Usman (2025). Forecasting of Biogas and Biomethane Outputs from Anaerobic Co-digestion Using Multilayer Perceptron Artificial Neural Networks (MLP-ANN). International Journal of Multidisciplinary Approach Research and Science / Articles. Vol. 3 No. 02.

[6] Jawad, J.; Hawari, A.H.; Zaidi, S.J. Artificial neural network modeling of wastewater treatment and desalination using membrane processes: A review. Chem. Eng. J. 2021, 419, 129540. [CrossRef].

[7] Tamru Tesseme Aragaw & Sumedha Chakma (2021). Integrated Approaches Towards Solid Waste Management: Artificial Neural Network Model for Prediction of Methane Fraction in Landfill Gas from Pretreated Waste in Bioreactor Landfills. pp 33–49. Springer Nature Link.

[8] Abu Qdais H, Bani Hani K, Shatnawi N. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resources.Conservation and Recycling. 2010; 54(6):359-63. http://dx.doi.org/10.1016/j.resconrec.2009.08

[9] Mougari, N.E.; Largeau, J.F.; Himrane, N.; Hachemi, M.; Tazerout, M. Application of artificial neural network and kinetic modeling for the prediction of biogas and methane production in anaerobic digestion of several organic wastes. Int. J. Green Energy 2021, 18, 1584–1596. [CrossRef]

[10] Sebti, A.; Boutra, B.; Trari, M.; Aoudjit, L.; Igoud, S. Application of Artificial Neural Network for Modeling Wastewater Treatment Process. In Lecture Notes in Networks and Systems; Springer Science and Business Media LLC: Berlin, Germany, 2019; pp. 143–154.


## *Appendix*

### MATLAB Script: ANN for Termite-Derived Biomethane Prediction

```matlab
% ANN Modeling of Biomethane Production
```

```
% Substrates: Sugarcane Bagasse, Maize Cob, Coconut Husk
% Inoculum: Termite-derived gut microbiota

clear; clc;

%% 1. Load Experimental Data
% Replace 'experimental_data.xlsx' with your actual file
% Format: 6 Columns for Inputs (SB%, MC%, CH%, Inoculum, pH, Time)
%         1 Column for Output (Methane Yield)
data = readmatrix('experimental_data.xlsx');

X = data(:, 1:6)'; % Inputs (Transposed for MATLAB ANN tool)
Y = data(:, 7)';   % Target Output (Transposed)

%% 2. Data Pre-processing (Normalization)
% Maps data to the range [-1, 1] for better convergence
[Xn, settingsX] = mapminmax(X);
[Yn, settingsY] = mapminmax(Y);

%% 3. Define Network Architecture
hiddenLayerSize = 10; % You can optimize this (e.g., 5 to 15)
net = fitnet(hiddenLayerSize, 'trainlm'); % Levenberg-Marquardt

% Activation Functions
net.layers{1}.transferFcn = 'tansig'; % Hidden Layer
net.layers{2}.transferFcn = 'purelin'; % Output Layer

%% 4. Data Partitioning
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

%% 5. Train the Network
[net, tr] = train(net, Xn, Yn);

%% 6. Model Testing and Denormalization
Yp_norm = net(Xn); % Predicted Normalized Yield
Yp = mapminmax('reverse', Yp_norm, settingsY); % Convert back to mL/g VS

%% 7. Performance Evaluation
performance = perform(net, Yn, Yp_norm);
fprintf('Mean Squared Error (MSE): %f\n', performance);

% Calculate R-squared for Total Dataset
R = corrcoef(Y, Yp);
fprintf('R-squared (R2) Value: %f\n', R(1,2)^2);

%% 8. Visualization
figure;
plotregression(Y, Yp, 'Model Regression Analysis');
xlabel('Experimental Methane Yield (mL/g VS)');
ylabel('ANN Predicted Methane Yield (mL/g VS)');

figure;
plotperform(tr); % Plot training, validation, and test performance
```
Use code with caution.

```
%% Optimization Loop for ANN Neurons
% This script finds the best number of neurons for highest R2
```

```
neuron_range = 2:20; % Testing from 2
to 20 neurons
best_R2 = 0;
best_neuron_count = 0;
results = []; % To store metrics for
plotting

for n = neuron_range
    % 1. Create and configure the
network
    net = fitnet(n, 'trainlm');
    net.trainParam.showWindow = false;
% Hides pop-up windows during loop

    % 2. Train the network
    [net, tr] = train(net, Xn, Yn);

    % 3. Predict and Denormalize
    Yp_norm = net(Xn);
    Yp = mapminmax('reverse', Yp_norm,
settingsY);

    % 4. Calculate R2 for this
specific neuron count
    R_matrix = corrcoef(Y, Yp);
    current_R2 = R_matrix(1,2)^2;

    % 5. Record results
    results = [results; n,
current_R2];

    % 6. Check if this is the best
version so far
    if current_R2 > best_R2
        best_R2 = current_R2;
        best_neuron_count = n;
        best_net = net; % Saves the
best performing model
    end

    fprintf('Testing Neurons: %d | R2:
%.4f\n', n, current_R2);
end

fprintf('\n--- Optimization Complete -
--\n');
fprintf('Best Neuron Count: %d\n',
best_neuron_count);
fprintf('Highest R2 Value: %.4f\n',
best_R2);

%% Plotting the Optimization Curve
figure;
plot(results(:,1), results(:,2), '-
ob', 'LineWidth', 1.5);
grid on;
xlabel('Number of Neurons in Hidden
Layer');
ylabel('Coefficient of Determination
(R^2)');
title('Optimization of ANN
Architecture');
hold on;
plot(best_neuron_count, best_R2, 'rx',
'MarkerSize', 15, 'LineWidth', 2);
legend('R^2 Trend', 'Optimal Point');
```