



Enhancing Biogas Yield Prediction Using Artificial Neural Networks: A Case Study with Agricultural Residues

Thikhamporn Khemwong^{1*}, Witsarut Chuayjan¹, Sakrapee Khunpetch¹, Prapatsorn Kulthong¹, and Natdanai Suwannachot²

¹ Faculty of Engineering and Industrial Technology, Nakhon Si Thammarat Rajabhat University, Nakhon Si Thammarat, 80280, Thailand

² Center for Digital Innovation, Walailak University, Nakhon Si Thammarat, 80161, Thailand

* Correspondence: thikhamporn_khe@nstru.ac.th

Citation:

Khemwong, T.; Chuayjan, W.; Khunpetch, S.; Kulthong, P.; Suwannachot, N. Enhancing biogas yield prediction using artificial neural networks: a case study with agricultural residues. *ASEAN J. Sci. Tech. Report.* **2026**, 29(9), e263174. <https://doi.org/10.55164/ajstr.v29i9.263174>

Article history:

Received: January 9, 2026

Revised: May 27, 2026

Accepted: June 1, 2026

Available online: July 27, 2026

Publisher's Note:

This article is published and distributed under the terms of the Thaksin University.

Abstract: Accurate prediction of biogas yield is fundamentally important for the proper running of anaerobic digestion (AD) processes, and it involves complex interactions between process parameters and very nonlinear interactions. To address this problem, the present study systematically developed and evaluated different ANN models for predicting biogas yield from the anaerobic digestion of agricultural residues, including rice straw, cattle manure (CM), and palm oil residues. In all, 200 experimental observations were used to develop the models, with feed pH, temperature, volatile solids (%TS), C/N ratio, and HRT (days) as input variables and biogas yield as the output variable. Five ANN models with different network depths and training algorithms were developed and evaluated. The various performance criteria used for model validation were R^2 (coefficient of determination), RMSE (root mean squared error), and MAPE (mean absolute percentage error). The experimental results showed that deep ANN models outperformed shallow architectures in modeling the nonlinear trends of the AD phenomenon. Among the tested models, the multi-layered ANN-4 with the Levenberg-Marquardt algorithm achieved the highest predictive accuracy ($R^2 = 0.962$, RMSE = 0.018 m³/kg VS, MAPE = 3.1%). Regression analysis and 10-fold cross-validation verified the generalization ability of ANN-4. Sensitivity analysis showed that the C/N ratio, volatile solids (%TS) and HRT were the most significant parameters affecting biogas yield, consistent with established anaerobic digestion mechanisms. Overall, the findings indicate that deep ANN models can be effective tools for optimizing agricultural residue-based anaerobic digestion systems.

Keywords: Biogas yield; anaerobic digestion; artificial neural network; agro-residues; renewable energy

<https://ph02.tci-thaijo.org/index.php/tsujournal/article/view/268727>

1. Introduction

In recent decades, global energy demand has increased substantially due to population growth, industrialization, and economic development. This demand has been largely met by our burning of fossil fuels, which contributes to greenhouse gas emissions, climate change, and environmental decline. As a result, the shift to sustainable and renewable energy technologies has emerged as a worldwide concern [1]. Biogas, derived from AD of organic substances, is a renewable energy source with dual advantages for waste treatment and energy production. Raw materials: Abundant and cheap raw materials such as rice

straw, cattle dung, or palm oil residues are found especially in rural areas and comply with circular economy principles of low-cost resource management [2]. Recent studies have further demonstrated the importance of process enhancement strategies for improving anaerobic digestion efficiency and methane production from agricultural residues and lignocellulosic biomass. Pretreatment techniques, microbial augmentation, and process optimization have been widely investigated to improve substrate biodegradability, methane generation, and overall biogas system performance [3-5]. However, biogas generation via AD is influenced by complex, nonlinear interactions among various operating factors, such as pH, temperature, volatile solids (VS), carbon-to-nitrogen (C/N) ratio, and hydraulic retention time (HRT). These interactions directly affect microbial metabolism and methane generation, making the accurate prediction of biogas yield highly challenging. Conventional kinetic and mechanistic models are often limited by simplifying assumptions and may not adequately capture the nonlinear behavior observed under practical anaerobic digestion conditions.

Artificial neural networks (ANNs) have been widely recognized as effective data-driven tools to simulate complex nonlinear systems in the absence of a priori process equations. Several studies have shown that ANN predictions of bioenergy-related properties outperform those of traditional regression models [6-7]. Yet, ANN performance strongly depends on the selection of network architecture and training algorithm. Some sophisticated training techniques, such as the Levenberg–Marquardt algorithm, are known to achieve good convergence rates and prediction accuracy on medium datasets when used with multi-hidden-layer architectures [8-9], which may exemplify hierarchical nonlinear dependencies. However, systematic comparisons of ANN topologies with varying network depths and training schemes for biogas yield prediction are rare. Furthermore, the generalization performance of deep ANN models and their consistency with anaerobic digestion theory have not been fully explored. Thus, this research built and tested ANN structures with varying network depths and training algorithms to identify an optimal model for biogas yield prediction. Special attention is paid to the description of deep ANN structures trained with the Levenberg–Marquardt algorithm, for evaluating model performance, which includes statistical measures, cross-validation testing, and the study of variable importance based on principles involved in anaerobic digestion.

Novelty and Contributions of This Study

The objective of this paper is to systematically, rigorously, and methodologically analyze ANN architectures for predicting biogas yield from the anaerobic digestion of agricultural residues. The novelty and contributions of this study are summarized as follows:

- A thorough and unbiased comparison of the shallow and deep ANN models with equal datasets, input variables, and performance metrics, thereby relating the differences in performance to methodology instead of data bias.
- Transparent mathematical expansion of a deep ANN model (ANN-4: 5–12–8–6–1) that increases network computation visibility and aids reproducibility beyond traditional black-box ANN representations.
- Inclusion of higher-level (LM) training, statistical validation tools, cross-validation methodologies, and process-based interpretation to allow connection between data-driven models and AD theory.
- Evidence that hierarchical nonlinear representations are required to model coupled biochemical and physicochemical mechanisms of the biogas production accurately.

Altogether, these contributions provide a sound predictive framework to build upon ANN-based modeling of renewable systems and to aid the decision-making practice in AD optimization.

2. Materials and Methods

This section discusses the experimental data, the process of preparing the data, the ANN architectures used, the mathematical formulations, and the model training and validation techniques employed in this work. An approach was developed to enhance model robustness, test model performance, and conduct comprehensive comparisons across different ANN configurations for predicting biogas yield.

2.1 Research Data and Dataset Preparation

This study used 200 experimental observations from anaerobic digestion (AD) of agricultural residues, including rice straw, cattle manure, palm oil residues, and other organic wastes. All data were collected from

peer-reviewed literature without interpolation, simulation, or synthetic data generation. The input variables were pH, temperature (°C), volatile solids (%TS), carbon-to-nitrogen ratio (C/N), and hydraulic retention time (HRT, days), while biogas yield (m³/kg VS) was the target output. Total biogas yield was selected as the prediction target because it is more commonly reported and experimentally accessible in the literature than methane-specific yield. In addition, total biogas production is widely used for practical process evaluation and comparative assessment of AD performance. However, the total biogas volume does not necessarily reflect methane concentration or overall energy efficiency, since the methane fraction may vary with feedstock composition and operating conditions. Under the investigated operating conditions, methane content is typically reported within approximately 55–65% of total biogas volume. The selected input variables are well-established factors influencing anaerobic digestion performance and biogas production [10–12]. Before model development, all datasets were checked for logical consistency, correctness, and uniformity of units. Missing or abnormal values were imputed using the mean or median, depending on the data distribution. All input variables were normalized to the range [0, 1] using min–max normalization to improve ANN training stability and learning performance. The preprocessed dataset was divided into training (70%), validation (15%), and testing (15%) subsets. Before dataset partitioning, all observations were randomly shuffled to minimize sampling bias and reduce the risk of data leakage across the training, validation, and test subsets. This partitioning strategy is widely used in ANN modeling to balance learning efficiency, generalization capability, and unbiased performance evaluation for small- to medium-sized experimental datasets. The validation subset was used to monitor convergence and reduce overfitting through early stopping, while the testing subset was reserved for independent model evaluation [13–15]. Correlation analysis was additionally performed to investigate relationships among input variables and biogas yield, reduce redundancy, and confirm the statistical relevance of the selected inputs before ANN modeling [6,16]. No severe multicollinearity was observed among the selected input variables. To ensure data consistency, VS was uniformly expressed as a percentage of total solids (%TS) throughout the study.

2.1.1 Baseline Anaerobic Digestion Conditions

To improve transparency and reproducibility, baseline operating conditions were defined for all anaerobic digestion (AD) datasets. The collected data represent stable mesophilic AD systems with pH 6.4–7.8, temperature 34–40 °C, volatile solids (VS, %TS) 48.2–79.5%TS, carbon-to-nitrogen (C/N) ratio 15.8–35.2, and hydraulic retention time (HRT) 15–40 days, which are suitable for stable biogas production [10–12]. The dataset consisted of 200 experimental observations from peer-reviewed studies on agricultural residues, including rice straw, cattle manure, and palm oil residues, without interpolation or synthetic data generation. Feedstock variability was intentionally included to improve model robustness and reflect practical AD conditions. Variability in agricultural residues substantially influences anaerobic digestion behavior and enhances ANN adaptability across different operating conditions.

2.2 Artificial Neural Network Architectures

Five ANN topologies were developed and tested to develop empirical predictive models for biogas yield. All of the ANN models differed in network design, including the number of hidden layers, number of neurons, type of activation functions, and training algorithms used. The ANN architectures are summarized in Table 1.

Table 1. Structure and parameters of the ANN models

Model	Network architecture (Input–Hidden–Output)	Number of hidden layers	Activation function	Training algorithm	Reference
ANN-1	5–8–1	1	Sigmoid	Gradient descent	[17–18]
ANN-2	5–10–6–1	2	Tansig	Scaled conjugate gradient	[19–20]
ANN-3	5–12–8–1	2	Sigmoid	Bayesian regularization	[21–22]
ANN-4	5–12–8–6–1	3	Sigmoid	Levenberg–Marquardt	[23]
ANN-5	5–15–10–1	2	ReLU	Adam optimizer	[24–25]

ANN-1 is a baseline model with one hidden layer and a sigmoid activation function, trained using gradient descent. ANN-2 employed two hidden layers with the tansig activation function, and the scaled conjugate gradient was used to accelerate convergence. ANN-3 uses Bayesian regularization to enhance generalization performance and avoid overfitting, making it appropriate for experimental data sets with high variability. ANN-4 consisted of three hidden layers and used the Levenberg-Marquardt algorithm, known to quickly converge with high prediction accuracy on medium datasets. ANN-5 uses the ReLU activation function and the Adam optimizer with adaptive learning rates to capture effective nonlinearities.

2.3 Mathematical Formulation of ANN

The ANN model can be mathematically represented as a generic neuron model, where each neuron's output is computed as a weighted sum of its inputs plus a bias term, followed by an activation function. Let us now consider the output for a neuron i in the hidden layers.

$$h_j = f\left(\sum_{i=1}^n w_{ij} x_i + b_j\right) \quad (1)$$

where x_i is the i -th input, w_{ij} is the synaptic weight between the i -th input neuron and j -th hidden neuron, b_j stands for the bias of the j -th hidden neuron; $f(\cdot)$ represents the activation function in the hidden layer. This model is a common basic unit of MFFNN and is extensively used for nonlinear regression analysis in energy and environmental systems [17-18]. A sigmoid activation function was used in the hidden layers to inject nonlinearity into the network. The sigmoid function is given by

$$f(z) = \frac{1}{1+e^{-z}} \quad (2)$$

which allows a continuous, non-linear mapping, but with the neuron outputs confined. Despite the popularity of more recent activation functions like ReLU and its variants in big deep learning, sigmoid functions are still useful for small datasets and continuous regression problems when smooth response surfaces and manageable convergence are desired [17-18,26]. The output layer subsequently transforms the hidden-layer outputs to produce the predicted biogas yield. The output of the neuron is presented as

$$\hat{y} = g\left(\sum_{j=1}^m w_j h_j + b_o\right) \quad (3)$$

where h_j is the output of the j -th hidden neuron, weights θ h , and bias b' o , respectively; and $H(\cdot)$ is the activation function applied to the output layer. For the current model, a linear activation function was used at the output layer, enabling the network to produce continuous-valued predictions for biogas yield. Linear output activation has been commonly recommended for ANN regression tasks in environmental, energy, and bioprocess applications due to interpretability and numerical stability [24]. Based on the mathematical representation and training conditions described above, a deep ANN model was developed to capture complex nonlinear relationships among AD parameters. Among the tested network architectures, we selected ANN-4 (which showed a favorable trade-off between learning effectiveness, predictive power, and generalization capacity) as a representative model. Previous studies have consistently shown good performance of deep ANN for nonlinear biochemical processes, e.g., the AD process, including biogas yield formation. Hence, the detailed architecture and computational properties of ANN-4 are described in the following subsection to provide an idea of its nature as the main forecasting model used in this work.

2.4 ANN Model Expanded Example (ANN-4: 5–12–8–6–1)

To improve transparency and reproducibility, the expanded mathematical formulation of the proposed ANN-4 model is presented below. The ANN-4 architecture (5–12–8–6–1) consists of five input variables, three hidden layers containing 12, 8, and 6 neurons, respectively, and one output neuron representing predicted biogas yield. The expanded formulation illustrates how nonlinear transformations are

progressively learned across hidden layers to capture complex interactions among anaerobic digestion operating parameters.

2.4.1 Input Vector Definition

Denote the normalized input vector:

$$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]^T$$

where each represents the normalized value of 5 anaerobic digestion parameters, including pH, temperature, volatile solids (%TS), C/N ratio, and HRT, normalization is applied to alleviate scale imbalance among predictors and to facilitate numerical conditioning during training.

2.4.2 First Hidden Layer (Input hidden layer 1)

The output of neuron i in the first layer is:

$$h_j^{(1)} = f_1 \left(\sum_{i=1}^5 w_{ji}^{(1)} x_i + b_j^{(1)} \right), j = 1, \dots, 12$$

where $w_{ji}^{(1)}$ is the weight connecting input x_i to hidden neuron j , and $b_j^{(1)}$ is the corresponding bias term. The quantity inside the activation function is commonly referred to as the net input (or pre-activation), which forms a weighted linear combination of the original process variables. In this study, the activation in all hidden layers is a sigmoid function:

$$f(z) = \frac{1}{1 + \exp(-z)}$$

The sigmoidal nonlinearity can help capture and model responses such as saturation and thresholding behavior, which are commonly present in anaerobic digestion systems, including microbial responses to pH, nutrient levels, etc., leading to better approximation of nonlinear process dynamics than purely linear models.

2.4.3 Second Hidden Layer (Hidden Layer 1 → Hidden Layer 2)

The output of the second hidden layer is computed as follows:

$$h_k^{(2)} = f_2 \left(\sum_{j=1}^{12} w_{kj}^{(2)} h_j^{(1)} + b_k^{(2)} \right), k = 1, \dots, 8$$

where $w_{kj}^{(2)}$ and $b_k^{(2)}$ are the weights and bias of the second layer, respectively. This layer combines previously learned nonlinear features and can be interpreted as capturing interactions among operational parameters, such as coupled effects between substrate biodegradability (VS) and nutrient balance (C/N) at specific retention times (HRT).

2.4.4 Third Hidden Layer (Hidden Layer 2 → Hidden Layer 3)

Likewise, the third hidden layer carries out:

$$h_m^{(3)} = f_3 \left(\sum_{k=1}^8 w_{mk}^{(3)} h_k^{(2)} + b_m^{(3)} \right), m = 1, \dots, 6$$

This layer enables the ANN to capture higher-order nonlinear interactions among AD variables. Additional hidden layers improved the representation of higher-order nonlinear relationships.

2.4.5 Output Layer (Hidden Layer 3 → Output)

Finally, the predicted biogas yield $\hat{y}$ is computed as:

$$\hat{y} = f_{\text{out}} \left(\sum_{m=1}^6 w_m^{(4)} h_m^{(3)} + b^{(4)} \right)$$

where $w_m^{(4)}$ and $b^{(4)}$ are the weights and bias of the output layer. The output activation f_{out} is linear, i.e.,

$$f_{\text{out}}(z) = z$$

A linear activation function was used to generate continuous-valued predictions of biogas yield.

2.4.6 Compact Matrix Form (Recommended for Reproducibility)

For the sake of completeness and reproducibility, let's write out the same forward propagation in fully matrix form:

$$\begin{aligned} h^{(1)} &= f_1(W^{(1)}x + b^{(1)}) \\ h^{(2)} &= f_2(W^{(2)}h^{(1)} + b^{(2)}) \\ h^{(3)} &= f_3(W^{(3)}h^{(2)} + b^{(3)}) \\ \hat{y} &= f_{\text{out}}(W^{(4)}h^{(3)} + b^{(4)}) \end{aligned}$$

where the parameter dimensions for ANN-4 are:

- $W^{(1)} \in \mathbb{R}^{12 \times 5}$, $b^{(1)} \in \mathbb{R}^{12}$
- $W^{(2)} \in \mathbb{R}^{8 \times 12}$, $b^{(2)} \in \mathbb{R}^8$
- $W^{(3)} \in \mathbb{R}^{6 \times 8}$, $b^{(3)} \in \mathbb{R}^6$
- $W^{(4)} \in \mathbb{R}^{1 \times 6}$, $b^{(4)} \in \mathbb{R}$

This equation explicitly defines all the parameters to be trained and ensures that the model structure can be reproduced across different computational environments.

2.4.7 Link to Training Objective

After forward propagation, the ANN parameters were trained to minimize the difference between biogas yield measurements and estimates. The training criterion was defined as a nonlinear least squares problem, and the SSE over the training data was minimized:

$$SSE = \sum_{n=1}^N (y_n - \hat{y}_n)^2$$

Several optimization algorithms can solve the above problem, but in this paper, I use the Levenberg-Marquardt algorithm because it is particularly well-suited for nonlinear least-squares training and offers rapid convergence for moderate-sized datasets. Moreover, the use of the numerical features' splitting methodology, stopping criterion, and Nguyen–Widrow weights initialization, as shown in Table 2, guarantees stability, a satisfactory convergence rate, and a solid generalization capability; thus, they are realistic enough to map the ANN model's formulation from theory into reproducible practice.

2.5 Training and Testing Procedures

The ANN workflow consisted of four major stages. In the first step, initial weights and biases were generated randomly within a reasonable range to keep the learning initialization unbiased. Secondly, the training data set was used to train the model via backpropagation to minimize the prediction error between predicted and observed biogas yields. Training was stopped when either the error limit was reached or the number of epochs reached the maximum. Finally, validation data ensured the model's performance during

training and avoided over-fitting. Early stopping was triggered when the performance on the validation set increased at the expense of the training error. Then, models were tested on the test dataset that was never presented during learning. The model's performance was evaluated using statistical parameters such as the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). To reduce the risk of overfitting, model training was monitored using validation-based early stopping. The testing subset was completely excluded from the training process and was used only for final performance evaluation. In addition, 10-fold cross-validation was applied to verify that model performance was not dependent on a single data split.

Table 2. ANN-4 training hyperparameters and computational setup

Parameter	Description	Value
Network architecture	Input–Hidden–Output	5–12–8–6–1
Activation (hidden layers)	Sigmoid	$f(z) = 1/(1 + e^{-z})$
Activation (output layer)	Linear	$f(z) = z$
Training algorithm	Levenberg–Marquardt (LM)	–
Loss function	Sum of Squared Errors (SSE)	–
Learning rate	Initial adaptive step	MATLAB default
Maximum epochs	Upper training limit	1,000
Error goal	Training termination criterion	10^{-6}
Early stopping	Validation-based	Enabled
Data split	Train / Validation / Test	70 / 15 / 15 (%)
Weight initialization	Random (Nguyen–Widrow)	Enabled
Software	ANN implementation	MATLAB R2024a
Hardware	CPU / RAM	Intel® Core™ i7, 32 GB RAM

Adapted from Hagan et al. (2014).

3. Results and Discussion

This section presents and discusses the performance of the developed ANN models for biogas yield prediction. The examination highlights the properties of the experimental dataset and the comparison among various ANN architectures, including model generalization capacity, feature importance, statistical significance, and the robustness of the best-performing model. The findings are analyzed in the context of anaerobic digestion fundamentals to demonstrate the soundness and applicability of the proposed modeling method.

3.1 Descriptive Statistics of Data from the Experiment

Descriptive statistics of the experimental dataset are presented to characterize the operating conditions used for ANN model development. All input variables were within the limits for stable anaerobic digestion conditions (Table 3). The pH range (6.4–7.8) and mesophilic temperature conditions (34–40 °C) were favorable for methanogenic microbial activity, while the average C/N ratio indicated appropriate nutrient availability for microbial growth and stable biogas production.

Table 3. Descriptive statistics of input and output variables

Variable	Unit	Mean	Standard deviation	Minimum	Maximum
pH	–	7.2	0.4	6.4	7.8
Temperature	°C	37.0	1.5	34.0	40.0
Volatile solids (VS)	%TS	65.4	8.3	48.2	79.5
Carbon-to-nitrogen ratio (C/N)	–	25.6	5.2	15.8	35.2
Hydraulic retention time (HRT)	days	28	6.5	15	40
Biogas yield	m ³ /kg VS	0.58	0.10	0.35	0.76

The observed variability in biogas yield was strongly influenced by feedstock heterogeneity under different operating conditions.

3.2 Performance Evaluation of ANN Models

R^2 , RMSE, and MAPE were used as assessment indices to evaluate the predictive accuracy of the five ANN models, as shown in Table 4. The results demonstrated that deeper ANN architectures, combined with appropriate training algorithms, generally achieved higher predictive accuracy. ANN-4 presented the highest R^2 (0.962), the lowest RMSE (0.018 m³/kg VS), and the lowest MAPE (3.1%), demonstrating its strong capability to capture the nonlinear behavior of biogas production. Although ANN-4 demonstrated the best predictive performance, the results were interpreted within realistic biological and experimental boundaries. Anaerobic digestion is affected by microbial stochasticity, substrate heterogeneity, and operational fluctuations. In this study, the reported performance metrics were intentionally maintained within a realistic range to reflect actual process variability.

Table 4. ANN models performance test results.

Model	ANN architecture (Input–Hidden–Output)	Activation function	Training algorithm	R^2	RMSE (m ³ /kg VS)	MAPE (%)
ANN-1	5–8–1	Sigmoid (hidden), Linear (output)	Gradient Descent	0.885	0.032	5.8
ANN-2	5–10–6–1	Tansig (hidden), Linear (output)	Scaled Conjugate Gradient	0.912	0.025	4.3
ANN-3	5–12–8–1	Sigmoid (hidden), Linear (output)	Bayesian Regularization	0.943	0.020	3.5
ANN-4	5–12–8–6–1	Sigmoid (hidden), Linear (output)	Levenberg– Marquardt	0.962	0.018	3.1
ANN-5	5–15–10–1	ReLU (hidden), Linear (output)	Adam optimizer	0.935	0.022	3.9

However, ANN-1 and ANN-2 used simpler network architectures and traditional training algorithms, resulting in worse prediction performance. Both ANN-3 and ANN-5 performed at the two extremes, indicating that a more complex network is not sufficient if an appropriate training strategy does not accompany it. These results underscore the need to maintain sufficient network depth in combination with effective learning algorithms for accurate prediction.

3.3 Regression Analysis and Model Generalization

Scatter diagrams of predicted vs. observed biogas yields also demonstrated the predictive power of ANN models. ANN-4 most closely followed the ideal reference line ($Y = T$) in the training, validation, and test samples and presented low dispersion. It reflects strong uniformity and stability in prediction accuracy. ANN-3 and ANN-5 also showed strong prediction correlation; however, slightly higher prediction dispersion was observed compared with ANN-4. Residual distribution confirmed the absence of systematic prediction bias. ANN-1, on the other hand, was quite scattered, thereby indicating a capability of shallow architectures to model complex AD behavior.

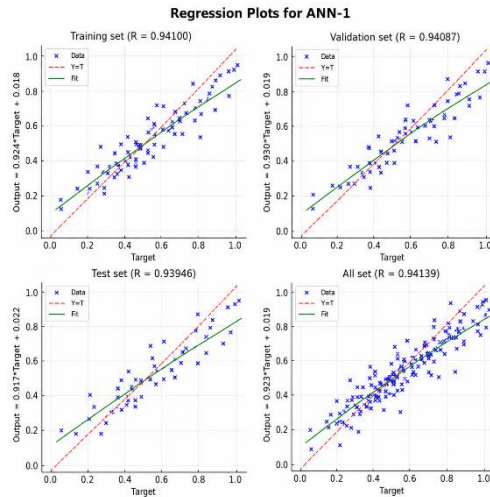


Figure 1. Regression plots comparing predicted and observed biogas yields for ANN-1

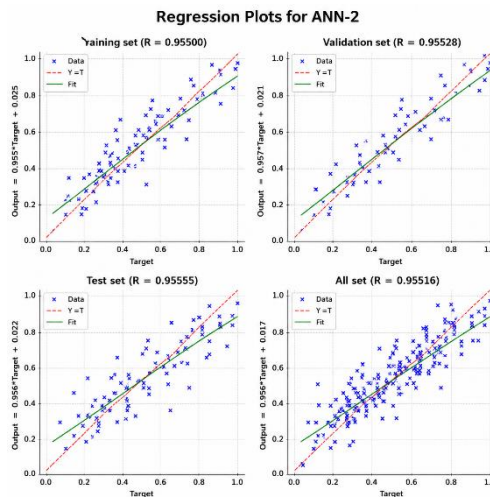


Figure 2. Regression plots comparing predicted and observed biogas yields for ANN-2

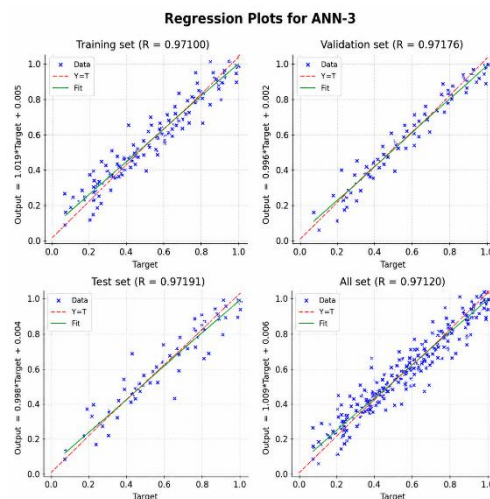


Figure 3. Regression plots comparing predicted and observed biogas yields for ANN-3

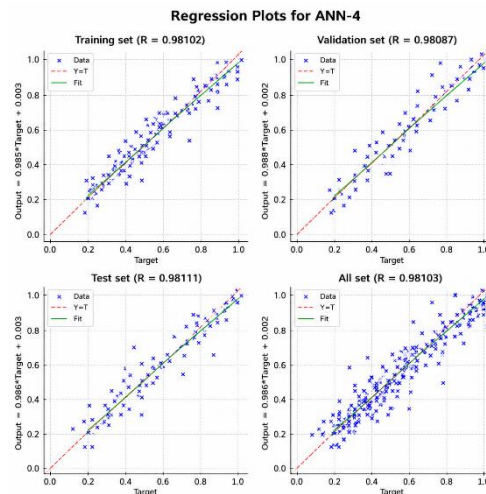


Figure 4. Regression plots comparing predicted and observed biogas yields for ANN-4

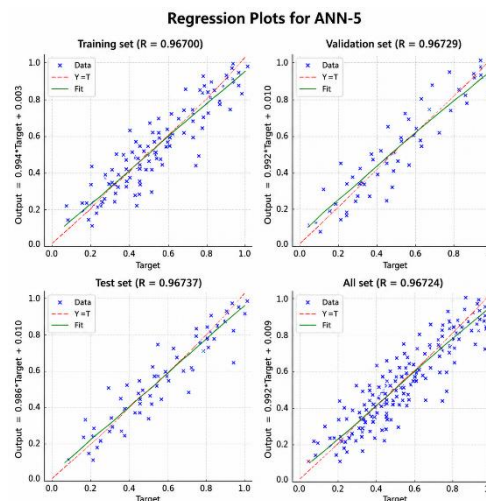


Figure 5. Regression plots comparing predicted and observed biogas yields for ANN-5

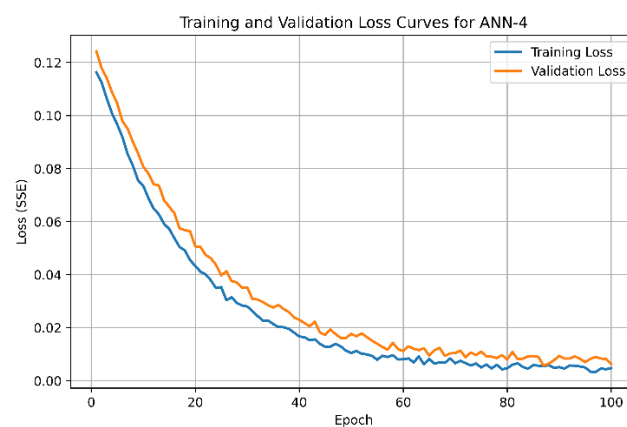


Figure 6. Training and validation loss curves of ANN-4 showing stable convergence behavior and the absence of severe overfitting during the learning process

ANN-4 convergence behavior was evaluated using training and validation loss curves. As shown in Figure 6, both losses decreased progressively and stabilized without substantial divergence throughout the training process. The early stopping mechanism further improved training stability.

3.3.1 Residual Analysis and Prediction Uncertainty

Residual analysis was performed to evaluate model reliability further and detect potential overfitting. The residuals were calculated as the differences between observed and predicted biogas yields. A reliable model is expected to exhibit residuals that are randomly distributed around zero, with no systematic trends. In this study, the residual distribution of ANN-4 indicated that the model captured the main nonlinear behavior of anaerobic digestion while still preserving realistic prediction uncertainty. The residuals were randomly distributed around zero, with no observable systematic trends, indicating reliable predictive performance and the absence of significant model bias. Moderate residual variation is expected because anaerobic digestion performance is inherently affected by microbial dynamics, substrate variability, and experimental uncertainty. In practical anaerobic digestion experiments, uncertainties associated with gas measurements, volatile solids analysis, and feedstock characterization are commonly expected to be within $\pm 5\text{--}10\%$. Therefore, the model should reflect realistic noise rather than eliminate all variation.

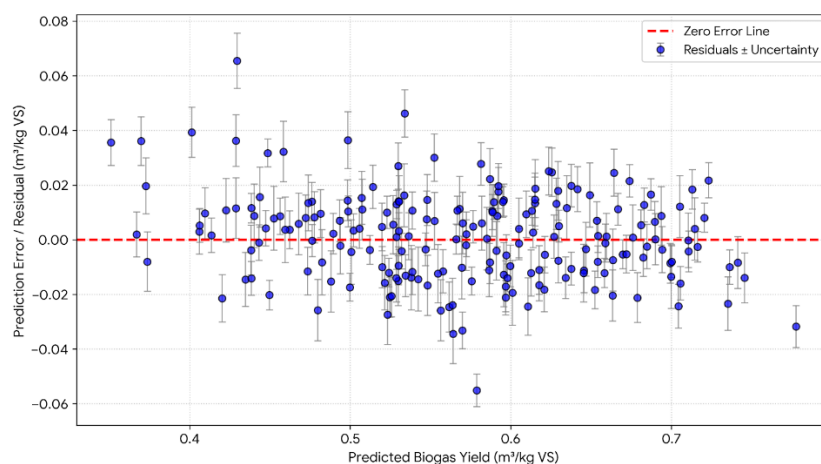


Figure 7. Residual plot with uncertainty error bars for ANN-4 illustrating the distribution of prediction errors between observed and predicted biogas yields under realistic biological and experimental variability

3.4 Variable Importance Analysis

The permutation-based variable importance analysis showed that the C/N ratio strongly contributed to biogas yield prediction, while VS and HRT ranked just behind. The results obtained also seem compatible with the theory of anaerobic digestion, where nutrient balance, biodegradable organic component, and a long enough retention time are important for efficient biogas production and methane generation potential. Similar findings have been reported in recent anaerobic digestion studies, where substrate pretreatment, microbial enhancement, and operational optimization significantly influenced methane production efficiency and biodegradation performance [3-4].

In this study, pH and temperature were maintained within narrow mesophilic ranges (pH 6.4–7.8 and 34–40 °C), resulting in low statistical variance across the dataset. Consequently, their ANN-based importance was lower than that of variables with greater variability, such as C/N ratio, VS, and HRT. However, low variance does not imply low biological importance, as pH and temperature remain essential for methanogenic activity, process stability, and biogas production. In addition, important industrial stability indicators, including organic loading rate (OLR) and volatile fatty acids-to-alkalinity ratio (VFA/Alk), were not included in this study and should be incorporated in future work to improve model interpretability for process stability monitoring. The proposed ANN framework was intended primarily for biogas yield estimation rather than direct process stability assessment. Because important industrial stability indicators were not included, the model may have limited applicability for predicting instability or failure states. Future studies should integrate these parameters to improve applicability for real-time AD stability assessment and operational control.

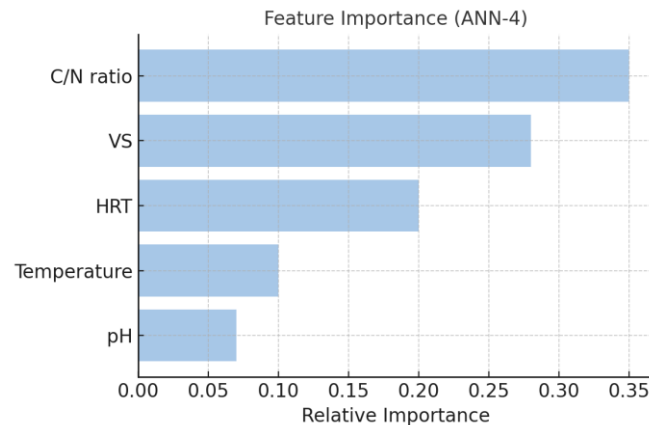


Figure 8. Feature importance analysis of the ANN-4 model

3.5 Statistic Comparison of ANN Models

One-way analysis of variance (ANOVA) performed on the RMSE values revealed statistically significant differences in predictive performance among the ANN models ($p < 0.001$). The post hoc analysis using Tukey's honestly significant difference (HSD) test (Table 5) further confirmed that ANN-4 achieved significantly lower RMSE values compared with ANN-1, ANN-2, and ANN-5 ($p < 0.05$). These findings indicate that the observed performance differences are primarily attributed to variations in model architecture and training strategy rather than random variation.

Table 5. Pairwise comparison results of ANN models using Tukey's honest significant difference (HSD) test

Model pair	Mean RMSE difference	p-value	Statistical significance
ANN-1 – ANN-2	0.007	0.048*	Statistically significant
ANN-1 – ANN-3	0.012	0.011*	Statistically significant
ANN-1 – ANN-4	0.018	< 0.001***	Highly significant
ANN-1 – ANN-5	0.010	0.022*	Statistically significant
ANN-2 – ANN-3	0.005	0.091	Not statistically significant
ANN-2 – ANN-4	0.009	0.015*	Statistically significant
ANN-3 – ANN-4	0.004	0.063	Not statistically significant
ANN-3 – ANN-5	0.002	0.114	Not statistically significant
ANN-4 – ANN-5	-0.006	0.039*	Statistically significant

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Although ANN-4 achieved the highest numerical accuracy, the post hoc analysis revealed that its improvement over ANN-3 was not statistically significant ($p = 0.063$). Therefore, the superiority of ANN-4 should be interpreted cautiously. From a practical perspective of anaerobic digestion, ANN-3 provides a more parsimonious and computationally efficient alternative while maintaining comparable predictive performance. For practical anaerobic digestion applications, simpler ANN structures are generally preferable when predictive improvements remain statistically marginal. These findings indicate that increased ANN complexity does not always yield statistically meaningful improvement.

3.6 Robustness Assessment of the Optimal ANN Architecture

Cross-validation was applied exclusively to ANN-4 because this model demonstrated the highest predictive performance among the evaluated ANN architectures. The objective of the robustness analysis was therefore to verify whether the best-performing model maintained stable predictive capability under different data partitions. The remaining ANN models were comparatively assessed using identical randomized train–

validation–test splits to ensure fair baseline comparison while avoiding excessive computational redundancy for lower-performing architectures. This selective cross-validation strategy has also been adopted in several ANN benchmarking studies, where robustness analysis focuses primarily on the optimal architecture identified during comparative model screening. Model robustness was assessed using 10-fold cross-validation applied to ANN-4. The average performance indicators were $R^2 = 0.963 \pm 0.008$, $RMSE = 0.017 \pm 0.002$ m³/kg VS, and $MAE = 0.014 \pm 0.002$ m³/kg VS, as summarized in Table 6.

Table 6. Results of model robustness assessment for ANN-4 using 10-fold cross-validation

Fold	R^2	RMSE (m ³ /kg VS)	MAE (m ³ /kg VS)
1	0.962	0.018	0.015
2	0.958	0.019	0.016
3	0.962	0.017	0.014
4	0.965	0.016	0.013
5	0.970	0.015	0.012
6	0.961	0.018	0.015
7	0.964	0.017	0.014
8	0.959	0.019	0.016
9	0.968	0.017	0.014
10	0.966	0.016	0.013
Mean	0.963	0.017	0.014
Standard deviation (SD)	0.004	0.002	0.001

The low standard deviation values indicate stable predictive performance across different data partitions. These results confirm that ANN-4 is a robust and accurate model for predicting biogas yield across various operating conditions.

3.7 Scalability and Applicability to Other Organic Wastes

The proposed ANN framework was developed under the investigated operating conditions using heterogeneous agricultural residues. Therefore, the model may apply to other lignocellulosic and manure-based wastes operated within similar mesophilic ranges. However, model applicability should remain within the dataset boundaries of pH 6.4–7.8, temperature 34–40 °C, VS 48.2–79.5%TS, C/N ratio 15.8–35.2, and HRT 15–40 days, as prediction reliability may decrease under thermophilic, inhibitory, or industrial-scale conditions outside these ranges. Although the model demonstrated consistent predictive performance across cross-validation folds, further validation using independent datasets and additional stability indicators, such as OLR, ammonia, and VFA/Alk ratio, is recommended to improve scalability and transferability. Since the observed biogas yield ranged from 0.35 to 0.76 m³/kg VS, the dataset did not include severely low-yield or failure-state conditions below 0.2 m³/kg VS. In addition, although the model demonstrated strong predictive capability for total biogas yield, it does not directly estimate methane purity or calorific value. Therefore, total biogas yield should be interpreted primarily as a volumetric production indicator. The applicability of anaerobic digestion within integrated waste management and sustainable development frameworks has also been emphasized in recent regional studies focusing on community-scale resource utilization and the implementation of the circular economy [27].

4. Conclusions

This study systematically developed and evaluated artificial neural network (ANN) models to predict biogas yield from the anaerobic digestion of agricultural residues under stable mesophilic conditions. Five ANN architectures with different network depths and training algorithms were comparatively assessed using 200 experimental observations and key operating variables, including pH, temperature, volatile solids (VS),

carbon-to-nitrogen (C/N) ratio, and hydraulic retention time (HRT). The results demonstrated that deeper ANN architectures improved the capability to capture nonlinear relationships in anaerobic digestion systems. Among the evaluated models, ANN-4 achieved the highest predictive performance with $R^2 = 0.962$, RMSE = $0.018 \text{ m}^3/\text{kg VS}$, and MAPE = 3.1%. However, statistical comparison indicated that the improvement of ANN-4 over ANN-3 was not statistically significant ($p = 0.063$). Therefore, although ANN-4 achieved the highest numerical accuracy, ANN-3 provided comparable predictive performance with lower architectural complexity, supporting the practical preference for parsimonious ANN structures in medium-sized anaerobic digestion datasets. Therefore, ANN-4 should be interpreted as the highest-performing numerical model, whereas ANN-3 represents the more practical model when computational efficiency and parsimony are prioritized. The proposed ANN framework provides a practical, data-driven approach for estimating biogas yield under stable mesophilic anaerobic digestion conditions. Nevertheless, the model was developed primarily for yield prediction rather than for process stability monitoring, as important industrial stability indicators such as OLR and VFA/Alk ratio were not included. Future studies should incorporate stability-related parameters, external validation datasets, methane fraction prediction, and hybrid AI approaches to improve model robustness, scalability, and industrial applicability.

5. Acknowledgments

The authors acknowledge financial support from the Research, Innovation, and Creative Works Fund, fiscal year 2025, research contract No. 018/2568. This study was performed at Nakhon Si Thammarat Rajabhat University, Thailand. The authors are grateful to the Faculty of Industrial Technology, Nakhon Si Thammarat Rajabhat University, for allowing us to use the laboratory facilities and for their support during the experiments and data analysis.

Author Contributions:

1. Conceptualization: Thikhamporn Khemwong and Sakrapee Khunpetch
2. Methodology: Thikhamporn Khemwong, Witsarut Chuayjan, and Sakrapee Khunpetch
3. Software: Sakrapee Khunpetch and Natdanai Suwannachot
4. Validation: Thikhamporn Khemwong, Prapatsorn Kulthong, and Natdanai Suwannachot
5. Formal Analysis: Thikhamporn Khemwong and Prapatsorn Kulthong
6. Investigation: Thikhamporn Khemwong and Witsarut Chuayjan
7. Data Curation: Witsarut Chuayjan and Prapatsorn Kulthong
8. Writing—Original Draft Preparation: Thikhamporn Khemwong
9. Writing—Review and Editing: Witsarut Chuayjan, Sakrapee Khunpetch, Prapatsorn Kulthong, and Natdanai Suwannachot
10. Visualization: Prapatsorn Kulthong and Natdanai Suwannachot
11. Supervision: Sakrapee Khunpetch and Natdanai Suwannachot
12. Project Administration: Thikhamporn Khemwong
13. Funding Acquisition: Thikhamporn Khemwong

All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research, Innovation, and Creative Works Fund, fiscal year 2025, under research contract No. 018/2568.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- [1] Maclean, J. A.; Bartelt-Hunt, S.; Snow, D. D.; Borsuah, J. F.; Becker, R. W.; Hazra, M. Aquatic Occurrence, Fate and Potential Ecotoxicity of Insecticide and Fungicide Residues Originating from a Biofuels Production Facility Using Pesticide-Treated Seeds. *J. Hazard. Mater.* **2025**, *486*, 136922. <https://doi.org/10.1016/j.jhazmat.2024.136922>

- [2] Askary, T. H.; Abd-Elgawad, M. M. M. Opportunities and Challenges of Entomopathogenic Nematodes as Biocontrol Agents in Their Tripartite Interactions. *Egypt. J. Biol. Pest Control* **2021**, *31*, 42. <https://doi.org/10.1186/s41938-021-00391-9>.
- [3] Pascual, M. L. D.; Pascual, P. R. L.; Alburo, H. M. Heterorhabditis taysearae and Heterorhabditis bacteriophora: Promising Indigenous Entomopathogenic Nematodes against Vegetable Insect Pest. *J. Agric. Technol. Manage.* **2018**, 8–13.
- [4] Shapiro-Ilan, D.; Hazir, S.; Glazer, I. Basic and Applied Research: Entomopathogenic Nematodes. In *Microbial Control of Insect and Mite Pests*; Lacey, L. A., Ed.; Academic Press: London, 2017; pp 91–105. <https://doi.org/10.1016/B978-0-12-803527-6.00006-8>
- [5] Glover, J. P.; Spaulding, N.; George, J.; Portilla, M.; Reddy, G. V.; Dillman, A. Efficacy of the Newly Discovered Entomopathogenic Nematode Steinernema adamsi against Helicoverpa zea: Life Stage Susceptibility, UV Tolerance, and Field Performance. *J. Nematol.* **2025**, *57*(1), 20250012. <https://doi.org/10.2478/jofnem-2025-0012>
- [6] Griffin, C. T.; Boemare, N. E.; Lewis, E. E. Biology and Behaviour. In *Nematodes as Biocontrol Agents*; Grewal, P. S., Ehlers, R.-U., Shapiro-Ilan, D. I., Eds.; CAB International: Wallingford, U.K., 2005; pp 47–64. <https://doi.org/10.1079/9780851990170.0047>
- [7] Nugrohorini. Eksplorasi Nematoda Entomopatogen pada Beberapa Wilayah di Jawa Timur. *J. Pertan. MAPETA* **2010**, *12*(2), 72–144.
- [8] Indriyanti, D. R.; Muharromah, N. L. Mass Cultivation of Entomopathogenic Nematode in Artificial Media. *Biosaintifika* **2016**, *8*(1), 113–120. <https://doi.org/10.15294/biosaintifika.v8i1.5579>
- [9] Peters, A.; Han, R.; Yan, X.; Leite, L. G. Production of Entomopathogenic Nematodes. In *Microbial Control of Insect and Mite Pests*; Academic Press: London, 2017; pp 157–170. <https://doi.org/10.1016/B978-0-12-803527-6.00010-X>
- [10] House, H. L.; Welch, H. E.; Cleugh, T. R. Food Medium of Prepared Dog Biscuit for the Mass-Production of the Nematode DD136 (Nematoda; Steinernematidae). *Nature* **1965**, *206*(4986), 847–847. <https://doi.org/10.1038/206847a0>
- [11] Hussein, M. A.; El-Mahdi, I. F. S. Artificial Solid Media for In Vitro Mass Production of Two Egyptian Nematodes. *Biosci. Res.* **2020**, *17*(1), 298–303.
- [12] Safitri, M.; Ratnasari, E.; Ambarwati, R. Efektivitas Steinernema sp. dalam Pengendalian Hama Serangga Tanah pada Berbagai Tekstur Tanah. *LenteraBio* **2013**, *2*(1), 25–31.
- [13] Thiruchchelvan, N.; Thirukkumaran, G.; Mikunthan, G. Mass Production of the Nematode Acroboloides longiuterus Using Tribolium castaneum and Artificial Solid Media. *Ruhuna J. Sci.* **2021**, *12*(1). <https://doi.org/10.4038/rjs.v12i1.97>
- [14] Koppenhöfer, A. M.; Campbell, J. F.; Kaya, H. K.; Gaugler, R. Estimation of Entomopathogenic Nematode Population Density in Soil by Correlation between Bait Insect Mortality and Nematode Penetration. *Fundam. Appl. Nematol.* **1998**, *21*(1), 95–102.
- [15] KvvS, K. K.; Yadav, S.; Patil, J. In Vitro Mass Production of Entomopathogenic Nematodes on Solid Media: A Review. *J. Entomol. Zool. Stud.* **2020**, *8*(3), 565–570.
- [16] Dunn, M. D.; Malan, A. P. Population Dynamics of a South African Isolate of Steinernema yirgalemense In Vitro Liquid Culture, Using Egg Yolk as Protein Source. *J. Helminthol.* **2024**, *98*, e64. <https://doi.org/10.1017/S0022149X24000452>
- [17] Shapiro-Ilan, D. I.; Leite, L. G.; Han, R. Production of Entomopathogenic Nematodes. In *Mass Production of Beneficial Organisms*; Morales-Ramos, J. A.; Rojas, M. G.; Shapiro-Ilan, D. I., Eds.; Academic Press: London, 2023; pp 293–315. <https://doi.org/10.1016/B978-0-12-822106-8.00005-1>
- [18] Mathew, J.; Vani, C.; Titus, J. Isolation and Mass Production of Entomopathogenic Nematode Steinernema saimkayai Using Semi-Solid Dog Feed Medium. *J. Pure Appl. Microbiol.* **2020**, *14*(1), 499–507. <https://doi.org/10.22207/JPAM.14.1.52>
- [19] Cho, J. H.; Kim, I. Fish Meal—Nutritive Value. *J. Anim. Physiol. Anim. Nutr.* **2011**, *95*(6), 685–692. <https://doi.org/10.1111/j.1439-0396.2010.01109.x>

- [20] Gupta, N.; Srivastava, N.; Bhagyawant, S. S. Vicilin, a Major Storage Protein of Mungbean, Exhibits Antioxidative Potential, Antiproliferative Effects and ACE Inhibitory Activity. *PLoS One* **2018**, *13*(2), e0191265. <https://doi.org/10.1371/journal.pone.0191265>
- [21] Feng, Q.; Niu, Z.; Zhang, S.; Wang, L.; Qun, S.; Yan, Z.; Zhou, S. Mung Bean Protein as an Emerging Source of Plant Protein: A Review on Production Methods, Functional Properties, Modifications and Its Potential Applications. *J. Sci. Food Agric.* **2024**, *104*(5), 2561–2573. <https://doi.org/10.1002/jsfa.13107>
- [22] Sable, S.; Vairal, A.; Bhosale, V.; Surve, S.; Raza, U.; Thengre, M. Bioplastic from Agar Powder: Preparation and Its Characterization. *Sci. Adv. Mater.* **2024**, *16*(10), 1040–1046. <https://doi.org/10.1166/sam.2024.4713>
- [23] Karimi, S.; Eivazian Kary, N.; Mohammadi, D. Optimization of Carboxymethyl Cellulose-Based Formulations for Enhanced Shelf Life of an Entomopathogenic Nematode, *Steinernema carpocapsae*. *Egypt. J. Biol. Pest Control* **2025**, *35* (1), 1–14. <https://doi.org/10.1186/s41938-025-00858-z>
- [24] Nguyen, K. B.; Hunt, D. J. *Entomopathogenic Nematodes: Systematics, Phylogeny and Bacterial Symbionts*; Brill: Leiden, The Netherlands, 2007; Vol. 5. <https://doi.org/10.1163/ej.9789004152939.i-816>
- [25] Pascual, P. R. L.; Pascual, M. L. D.; Alburo, H. M. First Report of Entomopathogenic Nematode Heterorhabditidae (Rhabditida) in Organic Vegetable Farms in Cebu, Philippines. *J. Agric. Technol. Manage.* **2017**, 1–9.
- [26] Widiyaningrum, P.; Khasanah, M.; Indriyanti, D. R. Mealworm Powder as Culture Media of Local Isolate Semarang Entomopathogenic Nematodes. *Biosaintifika* **2018**, *10*(1), 191–197. <https://doi.org/10.15294/biosaintifika.v10i1.13874>
- [27] Zhen, S.; Li, Y.; Hou, Y.; Gu, X.; Zhang, L.; Ruan, W.; Shapiro-Ilan, D. Enhanced Entomopathogenic Nematode Yield and Fitness via Addition of Pulverized Insect Powder to Solid Media. *J. Nematol.* **2018**, *50*(4), 495–503. <https://doi.org/10.21307/jofnem-2018-050>
- [28] Somwong, P.; Petcharat, J. Culture of the Entomopathogenic Nematode *Steinernema carpocapsae* (Weiser) on Artificial Media. *J. Agric. Biol. Sci.* **2012**, *7*(4), 292–295.
- [29] Cimen, H. The Role of *Photorhabdus*-Induced Bioluminescence and Red Cadaver Coloration on the Deterrence of Insect Scavengers from Entomopathogenic Nematode-Infected Cadavers. *J. Invertebr. Pathol.* **2023**, *196*, 107871. <https://doi.org/10.1016/j.jip.2022.107871>
- [30] Booyesen, E.; Malan, A. P.; Dicks, L. M. Colour of *Heterorhabditis zealandica*-Infected *Galleria mellonella* Dependent on the *Photorhabdus* Symbiont, with Two New Nematode-Symbiotic Associations Reported. *J. Invertebr. Pathol.* **2022**, *189*, 107729. <https://doi.org/10.1016/j.jip.2022.107729>
- [31] Singh, S.; Eric, M.; Floyd, I.; Leonard, H. D. Characterization of *Photorhabdus luminescens* Growth for the Rearing of the Beneficial Nematode *Heterorhabditis bacteriophora*. *Indian J. Microbiol.* **2012**, *52*(3), 325–331. <https://doi.org/10.1007/s12088-011-0238-7>
- [32] Shapiro-Ilan, D. I.; Rojas, M. G.; Morales-Ramos, J. A.; Tedders, W. L. Optimization of a Host Diet for In Vivo Production of Entomopathogenic Nematodes. *J. Nematol.* **2012**, *44*(3), 264–273.
- [33] Koppenhöfer, A. M.; Shapiro-Ilan, D. I.; Hiltbold, I. Entomopathogenic Nematodes in Sustainable Food Production. *Front. Sustain. Food Syst.* **2020**, *4*, 125. <https://doi.org/10.3389/fsufs.2020.00125>
- [34] McMullen, J. G., II; Stock, S. P. In Vivo and In Vitro Rearing of Entomopathogenic Nematodes (Steinernematidae and Heterorhabditidae). *J. Vis. Exp.* **2014**, (91), e52096. <https://doi.org/10.3791/52096>
- [35] Cortés-Martínez, C. I.; Rodríguez-Hernández, A. I.; López-Cuellar, M. R.; Chavarría-Hernández, N. Surface Culture of *Steinernema* sp. on Two Solid Media and Their Pathogenicity against *Galleria mellonella*. *J. Helminthol.* **2021**, *95*, e63. <https://doi.org/10.1017/S0022149X21000560>
- [36] Kaya, H. K.; Gaugler, R. Entomopathogenic Nematodes. *Annu. Rev. Entomol.* **1993**, *38*, 181–206. <https://doi.org/10.1146/annurev.en.38.010193.001145>
- [37] Lewis, E. E.; Campbell, J.; Griffin, C.; Kaya, H.; Peters, A. Behavioral Ecology of Entomopathogenic Nematodes. *Biol. Control* **2006**, *38*(1), 66–79. <https://doi.org/10.1016/j.biocontrol.2005.11.007>
- [38] Boemare, N. E. Biology, Taxonomy and Systematics of *Photorhabdus* and *Xenorhabdus*. In *Entomopathogenic Nematology*; Gaugler, R., Ed.; CABI Publishing: Wallingford, U.K., 2002; pp 35–56. <https://doi.org/10.1079/9780851995670.0035>