Vol. 5(1), pp. 1-8, June 2026
eISSN: 2976-4009

ASEAN Journal of Process Control

Research Article

Modeling And Optimization of Methane Yield and H2S Reduction from Palm Oil Mill Effluent Using Artificial Neural Network

Nor Fatin Zulaikha Binti Mohamad Mokhtar[1]*, Nurhazwani Yusoff Azudin[1]

1 Faculty of Engineering, Universiti Malaysia Sabah, 88400, Jalan UMS, Kota Kinabalu, Sabah, Malaysia
*Corresponding Author: hazwaniyusoff@ums.edu.my

Abstract: The treatment of palm oil mill effluent (POME) through anaerobic digestion proposes a sustainable corresponding for renewable energy production in the form of methane-rich biogas. However, the performance of this process is highly dependent on several operational parameters including pH, temperature, and recirculation ratio which under industrial conditions can fluctuate. Conventional modelling approaches are limited in capturing complex and nonlinear interactions among these variables. In this study, an Artificial Neural Network (ANN) model was developed and optimized to predict methane yield and hydrogen sulphide (H_2S) concentration based on real industrial data collected from a biogas facility in Pahang (Chen et al.,2023) gathering over two-year period. The ANN model was constructed using MATLAB R2022a and trained using the Bayesian Regularization algorithm. Optimum architecture was identified as a feedforward neural network structure with one hidden layer containing 19 neurons and the log-sigmoid activation function. Excellent predictive performance of ANN model shown with a high regression coefficient (R = 0.99625) and a low mean absolute percentage error (MAPE = 3.53 %) in predicting methane yield and H_2S concentration. Sensitivity analysis revealed that pH was the most influential factor affecting both outputs. The reduction of H_2S was done through the determination of the most optimum operating condition that would result in low levels of predicted H_2S concentration and high level of methane yield using the trained ANN model. Based on the trained model, the optimum operating conditions for anaerobic digestion were determined to be a pH of 7.0934, a temperature of 38.86 °C and a recirculation ratio of 1.7817. The outcomes of this research highlight the reliability and practical applicability of ANN models for industrial-scale process prediction and optimization. This model has been developed using real plant data, while its predictions have been compared with actual values from the same dataset as its accuracy evaluation. It further highlights the great potential of models to support decision-making accurately and further improve industrial POME treatment of biogas production system.

Keywords: Palm Oil Mill Effluent (POME), Methane Yield, Artificial Neural Network (ANN), Anaerobic Digestion (AD), Biogas Optimization.

1. Introduction

Palm oil mill effluent (POME) is a liquid waste produced during manufacturing process of crude palm oil (CPO). Yong et al., [1] states that only 21% of Fresh Fruit Brunches (FFB) are converted into CPO, while the remaining 79% considered as waste. POME is formed up of three main sources: sterilization process, digestion process and clarification process [2]. As a major wastewater stream from palm oil mills, POME poses considerable environmental risk to water quality if released directly into water bodies without an early treatment system [3]. This effluent is rich in organic matter which is reflected by elevated COD and BOD values. In recent years, anaerobic digestion (AD) process has been recognized as an effective biological treatment method for treating POME. Converting POME into methane-producing biogas through AD process has gained increasing attention for its dual benefits of pollutant reduction and renewable energy generation. However, Hoon et al. [2] documented that the biogas product from the AD process develops 60-70% methane (CH_4), 30-40% carbon dioxide (CO_2), and trace hydrogen sulphide (H2S). The presence of H2S poses a major challenge among these components due to its corrosive and highly hazardous nature. It might constrain the efficiency and safety of biogas treatment systems. Therefore, boosting methane yield production while lowering H2S concentration are the major points for optimizing biogas quality.

The process of AD occurs in four stages: Hydrolysis, Acidogenesis, Acetogenesis, Methanogenesis [4]. Maintaining optimum condition during AD process is essential to ensure that the microbial community remains balanced to undergo each stage effectively. Even small fluctuations can disturb microbial activity which cause in reducing methane yield and increasing undesirable by-products of H_2S. Chen et al. [5] highlights that pH, temperature, and recirculation ratio are the most critical factors that affect AD activities and methane yield production. This emphasizes the need for accurate modelling tools that are capable in reflecting real process behaviors of microbes in POME treatment.

Various approaches, encompassing both traditional and modern technologies, are employed in the development of models aimed at enhancing methane production within the context of AD. Conventionally, optimization methods depending on empirical experimentation and classical statistical tools, such as Gaussian and Surge models [6]. However, these methods poorly represent the complex relationships among operational variables since they often assume idealized reactor behavior, which fails to capture real-world behavior of microbes in industrial bioreactors. The complexity of the systems makes mathematical modelling insufficient to capture the effect of specific independent variables on the response. Thus, recent studies have applied advanced mathematical methods such as Response Surface Methodology (RSM), and Machine Learning (ML) algorithm such as Artificial Neural Network (ANN) to optimize the methane yield from POME.

Despite the development of RSM model offers a simple statistical framework for determining optimal conditions, its reliance on predefined quadratic models limits its ability to capture the complex nonlinear among the AD processes. ANN approach on the other hand, showed high performed in learning complex nonlinear relationship directly from the data. Supporting this, previous studies such as Chen et al. [5] demonstrated that ANN models showed impressive predictive accuracy for methane yield, with R^2 values of 0.9799 in testing phase and 0.9891 in validation phase. However, the dataset used in their study was not entirely based on actual plant operations; instead, it was expanded using a pre-experimental design (Box-Behnken Design) to generate 80 datasets from only 24 real operational datasets. Thus, the data may not fully reflect real industrial conditions.

Therefore, this study addresses this gap by modelling and optimizing ANN using raw operational data rather than simulated datasets to provide a more realistic evaluation of model performance, generalization capability, and practical applicability. Furthermore, the study seeks to identify the optimum AD conditions in a palm oil mill based on the optimized ANN model.

2. Materials and Methods

2.1. Data Source and collection

Data were obtained from a full-scale biogas plant in Pahang, Malaysia, utilizing POME as substrate. Operating parameters such as pH (7.0–7.2), temperature (35.8–42.5°C), and recirculation ratio (0.76–3.6) were varied, with methane yield and H_2S concentration measured. The industry data used was extracted from the Supervisory Control and Data Acquisition (SCADA) control system of the biogas plant over two years with monthly average values.

2.2. Development of artificial neural network (ANN) model

2.2.1 Data Pre-processing

The dataset was then normalized using the 'mapminmax' function (−1 to 1) in MATLAB 2022a to enhance model training and numerical stability. It was randomized and split into 70% training, 15% validation, and 15% for testing to avoid any bias in estimating predictive accuracy.

2.2.2 Define ANN model structure

The initial ANN structure was first developed as baseline model to evaluate the relationship between input and output parameters. The ANN structure was set as a feedforward network consisting of 3–10– 2 neurons (input–hidden–output). The network was trained using the 'trainlm' algorithm, with 'tansig' activation in the hidden layer and 'purelin' in the output layer. This configuration represents the initial model prior to optimization.

2.2.3 Evaluate Network Performance

The performance of ANN model was evaluated using Mean Squared Error (MSE), Coefficient of Correlation (R), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) shown in Equations (1) - (4). These metrics were used as indicators to evaluate the accuracy and reliability of the ANN model.

$$MSE = \frac{\sum(E_i - P_i)^2}{N} \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|E_i - P_i| \tag{2}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|E_i - P_i|}{|E_i|} \tag{3}$$

$$R = \frac{\sum_{i=1}^{N}(P_i - \bar{P}_i)(E_i - \bar{E}_i)}{\sqrt{\sum_{i=1}^{N}(E_i - \bar{E}_i)^2(P_i - \bar{P}_i)^2}} \tag{4}$$

where N is number of observations, Ei is the experiment (actual) value, and Pi Predicted value, $E\ i$ and $P\ i$ are mean of actual data and mean of observed data, respectively. MSE is a statistic for measuring the efficiency of ANN model, since it indicates how much its predicted values differ from the actual data. The smaller the MSE value approaching 0, the better the model fits the data and the more accurate the model. Furthermore, the R value provides an indication of the degree of fit of the data to the line of regression and takes the value of between 0 and 1.

2.3 Optimization of artificial neural network (ANN) model

2.3.1 Selection of Training Algorithm and Hidden Neuron Number

The ANN were trained using 5 different training algorithms in MATLAB 2022a software as shown in Table 1. Each algorithm was tested across a range of hidden neurons from 1 to 20 neurons with 50 runs to determine which combination provided the highest accuracy based on MSE, MAPE, and R^2 values.

Table 1. Types of training algorithms.

Training Algorithm	Description
'trainlm'	Levenberg-Marquardt
'trainbr'	Bayesian Regularization
'trainbfg'	BFGS Quasi-Newton
'trainrp'	Resilient Backpropagation
'trainscg'	Scaled Conjugate Gradient

2.3.2 Selection of Training Function at Hidden and Output Layer

After determining the best training algorithm and neuron number, the ANN models were further tested using different transfer function combinations for the hidden and output layers. Three common functions (purelin', 'tansig', and 'logsig') were evaluated. A total of nine combinations were tested to identify which pairing produced the most accurate and stable predictions for both methane and H_2S outputs.

The trained ANN model was then employed as an optimization tool to identify operating conditions that maximize the yield of methane and at the same time minimizing the H_2S concentration, thus demonstrating an indirect H_2S reduction method through process optimization.

2.4. Sensitivity Analysis

To evaluate the influence of each input parameter on the ANN outputs, the Garson connection weight method was applied. This method analyses the internal connection weights of the trained network to estimate the relative importance of each input variable which are temperature, pH, and recirculation ratio on methane yield and H_2S concentration.

3. Results and Discussion

3.1. Analysis of the developed ANN model

To find the best ANN model, different training algorithms, numbers of hidden neurons, and transfer function combinations were tested. Box plots were used to illustrate each algorithm performed. As shown in Fig. 1, the Levenberg-Marquardt (LM), Bayesian Regularization (BR), BFGS Quasi-Newton (BFG), Resilient Backpropagation (RP) and Scaled Conjugate Gradient (SCG) showed different accuracy levels when hidden neurons were changed from 1 to 20 for each output. Across all experiments, the Bayesian Regularization (BR) algorithm consistently produced small variation and high median prediction accuracy indicating superior learning efficiency and stability. Hence, BR was chosen for further evaluation of ANN model.

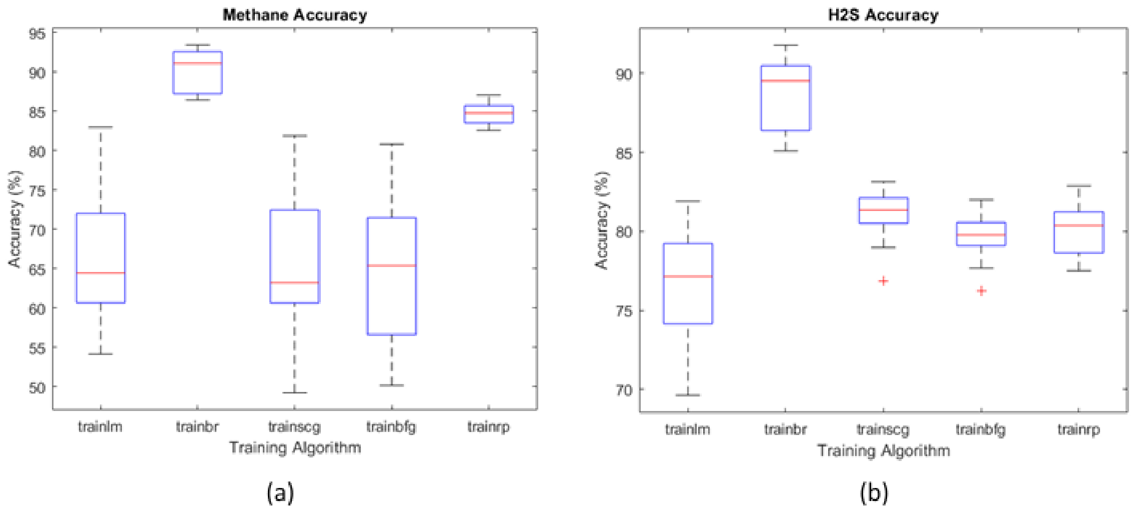

Figure 1: Boxplots showing (a) methane prediction accuracy (%) and (b) H_2S prediction accuracy (%) for various training algorithms.

To identify the optimal number of hidden neurons for the ANN model, trainbr was evaluated throughout the range of neurons from 1 to 20 with each model trained 50 runs to ensure consistency. Model performance was measured using four performance metrics as plotted and illustrated in Fig. 2 Among all configurations tested, the ANN model with 19 hidden neurons performed the best according to the evaluation metrics.

To determine the most suitable transfer function for the ANN model, several combinations of activation functions for the hidden and output layers were tested. Each configuration was trained and illustrated in boxplot as shown in Fig. 3. Based on the results, the logsig–purelin combination achieved the highest overall accuracy and was selected as the optimal transfer function for the final ANN model.

This concluded model 3–19–2 was the best topology where 3 represents the input variables, 19 as hidden neurons and 2 for the response variables in forecasting methane yield and H2S concentration in raw biogas. The optimized ANN model demonstrated high predictive accuracy, achieving an overall correlation coefficient (R) of 0.98341 across all data. As shown in Fig. 4, the model achieved an R value of 0.99489 for the training dataset indicating an almost perfect fit, while 0.93019 achieved an R value of 0.93019 for the test dataset, reflecting strong generalization performance on unseen data. The Mean Squared Error (MSE) stood at 2854.99 while the Mean Absolute Percentage Error (MAPE) was 3.53%, point out that the prediction of the model turned from the actual values was only 3.53%. These results show that the ANN was able to accurately capture the nonlinear relationships between the input parameters and the target outputs.

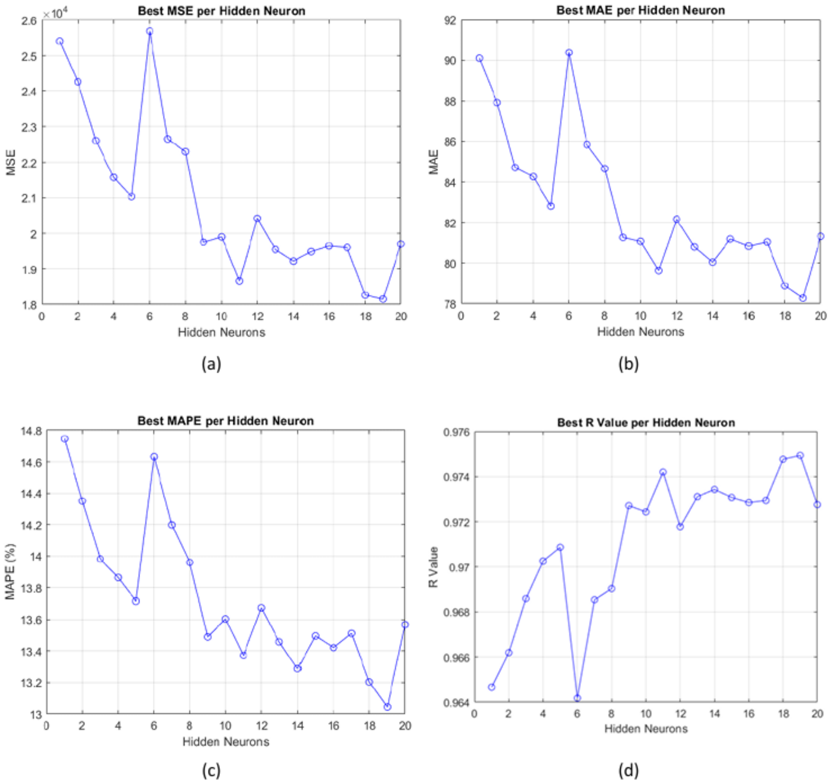

Figure 2: Performance of the ANN model across different numbers of hidden neurons based on (a) MSE, (b) MAE, (c) MAPE, and (d) R-value.

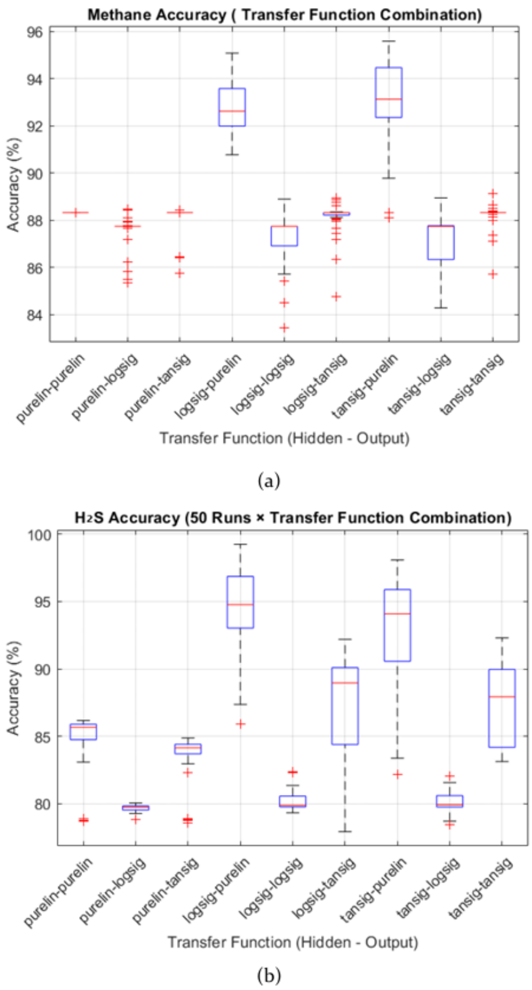

Figure 3: Boxplots showing (a) methane prediction accuracy (%) and (b) H_2S prediction accuracy (%) across different transfer function combinations.

3.2. Sensitivity Analysis of the ANN model

To further evaluate the contribution of each input variable to the model outputs, a sensitivity analysis was performed using the Garson connection weight method. Based on these simulations, it can be seen from Fig. 5 that pH had a sensitivity value of 54.5% on methane production and 50% on H2S concentration, followed by recirculation ratio with 34.8% on methane production and 40.2%, followed by temperature with 11.2% and 9.8%, respectively, on H2S concentration. The high sensitivity obtained on pH regards its importance on microbial activities and ensures that microbial growth and functions are not significantly hampered due to pH imbalances and thus ensures AD processes are optimized. Also, it confirms that optimal biogas production requires optimal control of the pH.

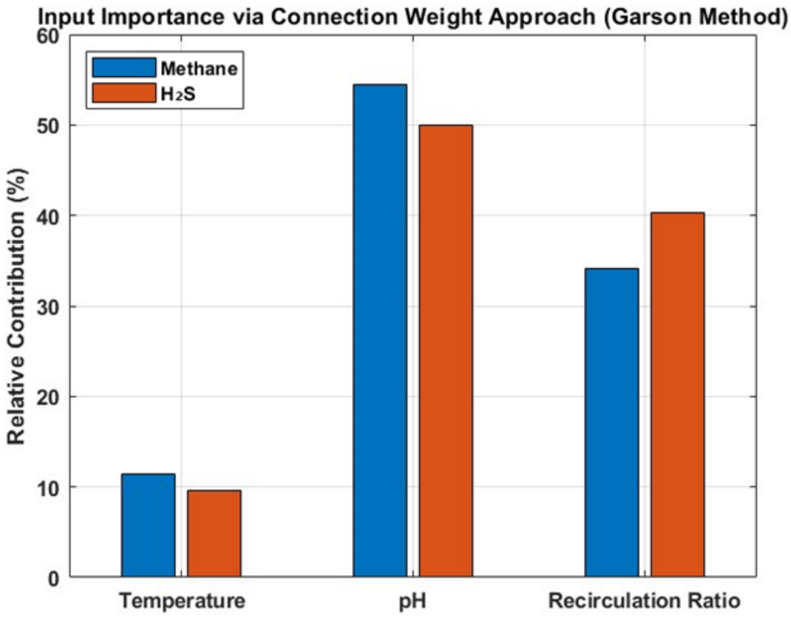

Figure 5: Relative importance of each input factor on methane yield.

3.3. Optimum Operating Conditions of AD process

Based on the optimization results obtained using the trained ANN model, the optimum operating conditions for maximizing methane production while minimizing H_2S concentration were summarized in Table 2. The optimum pH value identified was 7.09, implying it is slightly alkaline and thus an appropriate pH range for AD. The optimum value for temperature at 38.86°C represents mesophilic temperature. Moreover, Retention Ratio (RR) with an optimum value of 1.78 signifies the duration required for POME retention in the reactor for optimal methane production. At these optimal conditions within the anaerobic digestion process, the ANN model does not directly reduce H_2S concentration. Instead, it learns trends in the operational data and highlights operating conditions that favor methanogenesis. These conditions maximize methane and are consistently associated with lower H_2S concentration, leading to an indirect reduction in predicted H_2S level.

Table 2. Optimum operating conditions predicted by the ANN model.

Parameters	Optimum Conditions
pH	7.09
Temperature	38.86
RR	1.78

A comparative study was conducted to evaluate the performance of the developed ANN model against existing research. The comparison focuses on three main areas which are the ANN structure used, how accurate the model is in predicting outputs, and the optimum anaerobic digestion (AD) conditions for maximizing methane yields and reducing H2S concentration. Table 3 provides a comparative summary of ANN model structures, predictive performance and optimum AD conditions.

Table 3. Summary of Performances ANN model.

Structure	Performances (R2)	Optimum Condition			References
		pH	T	RR	
3-14-2	0.9762	7.36	42◦C	2.5	Chen et al., 2023
3-15-15-2	0.9797	7.03	38.9◦C	1.89	Chong et al., 2023
3-19-2	0.9920	7.09	38.9 ◦C	1.78	Current work

In this study, a 3-19-2 ANN model with Bayesian regularization was developed to predict both methane yield and H_2S concentration. The model achieved a high R^2 of 0.996, with optimum conditions of pH 7.09, temperature 38.9°C, and RR 1.78, which fall within the range reported by previous studies such as Chen et al. [5] and Chong et al. [7]. Their studies also used pH, temperature, and recirculation ratio as inputs but produced slightly different optimum conditions due to differences in ANN structure and output focus. Compared to these studies, the model developed here showed higher prediction accuracy and provided realistic operating conditions suitable for industrial AD processes. This comparison demonstrates that variations in ANN structure, training algorithms, and target outputs can influence model performance, and it supports the reliability and relevance of the ANN model developed in the current work.

4. Conclusions

This study successfully developed and optimized an artificial neural network (ANN) model to predict methane yield and H_2S concentration using real industrial data from a palm oil mill anaerobic digestion (AD) system. Unlike previous studies that relied on Response Surface Methodology (RSM) and simulated datasets, this work used actual plant data without any pre-experimental design. Despite the limitations of working with a small dataset, the ANN model achieved strong training performance, with an R-value of 0.99625 and a low MAPE of 3.53%, showing that the model was able to learn complex nonlinear patterns and generate reliable predictions from real operational conditions. However, when estimating new input values, the model's accuracy decreased with signs of overfitting. This seems to show that the dataset size was not representative for the model to generalize well outside the range of its training. To enhance the robustness of the model, future studies ought to be directed toward the acquisition of a bigger and more diverse dataset that covers wider ranges of temperature, pH, and recirculation ratio. Growing the dataset will assist the model in better representation of the variability in actual AD operations.

Future efforts should also investigate the incorporation of more process parameters like Organic Loading Rate (OLR), or Volatile Fatty Acid (VFA) concentrations to have a full representation of the biogas production process. Moreover, advanced techniques of machine learning such as crossvalidation could be applied to reduce the problem of overfitting and enhance generalization. Integrating these enhancements will further develop the ANN model into a more reliable and industrially practical tool capable of predicting methane yield and assisting in the optimization of anaerobic digestion processes.

Declaration of Competing Interest: The authors declare no conflict of interest.

Acknowledgments: The authors would like to acknowledge Ministry of Higher Education Malaysia for providing research funding through FRGS-EC Grant (FRGS-EC/1/2024/TK05/UMS/02/2) and Universiti Malaysia Sabah for support.

References

1. Yong, Gloria Tung Xin, Yi Jing Chan, Phei Li Lau, Baranitharan Ethiraj, Ayman A. Ghfar, Abdallah A. A. Mohammed, Muhammad Kashif Shahid, and Jun Wei Lim. 2023. "Optimization of the Performances of Palm Oil Mill Effluent (POME)-Based Biogas Plants Using Comparative Analysis and Response Surface Methodology." Processes 11(6). doi:10.3390/pr11061603.
2. Hoon, Yee Theng Jessy, Yi Jing Chan, Yoke Kin Wan, Yong Kheng Goh, and Sara Kazemi Yazdi. 2024. "Industrial-Scale Anaerobic Co-Digestion (ACoD) of Palm Oil Mill Effluent (POME) and Decanter Cake (DC) for Maximizing Methane Yield: An Integrated Machine Learning and Simulation-Based Economic Analysis Approach." Energy 289. doi:10.1016/j.energy.2023.129939.
3. Panagopoulos, Argyris, and Katherine Joanne Haralambous. 2020. "Minimal Liquid Discharge (MLD) and Zero Liquid Discharge (ZLD) Strategies for Wastewater Management and Resource Recovery – Analysis,

Challenges and Prospects." Journal of Environmental Chemical Engineering 8(5):104418. doi:10.1016/J.JECE.2020.104418.

4. Meegoda, Jay N., Brian Li, Kush Patel, and Lily B. Wang. 2018. "A Review of the Processes, Parameters, and Optimization of Anaerobic Digestion." International Journal of Environmental Research and Public Health 15(10).
5. Chen, Wei Yao, Yi Jing Chan, 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)." Water (Switzerland) 14(9). doi:10.3390/w14091410.
6. Abubakar, Umar Alfa, Gul Sanga Lemar, Al Amin Danladi Bello, Aliyu Ishaq, Aliyu Adamu Dandajeh, Zainab Toyin Jagun, and Mohamad Rajab Houmsi. 2024. "Evaluation of Traditional and Machine Learning Approaches for Modeling Volatile Fatty Acid Concentrations in Anaerobic Digestion of Sludge: Potential and Challenges." Environmental Science and Pollution Research. doi:10.1007/s11356-024-33281-2.
7. Chong, Daniel Jia Sheng, Yi Jing Chan, Senthil Kumar Arumugasamy, Sara Kazemi Yazdi, and Jun Wei Lim. 2023a. "Optimisation and Performance Evaluation of Response Surface Methodology (RSM), Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) in the Prediction of Biogas Production from Palm Oil Mill Effluent (POME)." Energy 266. doi:10.1016/j.energy.2022.126449.