


## Article

# Derivation of Optimal Operation Factors of Anaerobic Digesters through Artificial Neural Network Technology

Yumeng Bao <sup>1</sup>, Ravindranadh Koutavarapu <sup>2,\*</sup>  and Tae-Gwan Lee <sup>1,\*</sup>

<sup>1</sup> Department of Environmental Science, Keimyung University, Daegu 42601, Republic of Korea; yumengbao994@gmail.com

<sup>2</sup> Department of Robotics Engineering, College of Mechanical and IT Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

\* Correspondence: ravindra\_physicist@ynu.ac.kr (R.K.); wateree@kmu.ac.kr (T.-G.L.)

**Abstract:** The anaerobic digestion of sewage sludge in South Korean wastewater treatment plants is affected by seasonal factors and other influences, resulting in lower digestion efficiency and gas production, which cannot reach optimal yields. The aim of this study was to improve the digestion efficiency and gas production of sludge anaerobic digestion in a wastewater treatment plant (WWTP) by using data mining techniques to adjust operational parameters. Through experimental data obtained from the WWTP in Daegu City, South Korea, an artificial neural network (ANN) technology was used to adjust the range of the organic loading rate (OLR) and hydraulic retention rate (HRT) to improve the efficiency and methane gas production from anaerobic sludge digestion. Data sources were normalized, and data analysis including Pearson correlation analysis, multiple regression analysis and an artificial neural network for optimal results. The results of the study showed a predicted 0.5% increase in digestion efficiency and a 1.3% increase in gas production at organic loads of 1.26–1.46 kg/m<sup>3</sup> day and an HRT of 26–30 days. This shows that the ANN model that we established is feasible and can be used to improve the efficiency and gas production of sludge anaerobic digestion.

**Keywords:** data mining technology; digestion efficiency; gas production; organic matter load; HRT



**Citation:** Bao, Y.; Koutavarapu, R.; Lee, T.-G. Derivation of Optimal Operation Factors of Anaerobic Digesters through Artificial Neural Network Technology. *Systems* **2023**, *11*, 375. <https://doi.org/10.3390/systems11070375>

Academic Editors: Wendong Yang, Jinpei Liu and Jianzhou Wang

Received: 3 May 2023

Revised: 14 July 2023

Accepted: 20 July 2023

Published: 22 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sludge is produced as a byproduct of treating wastewater; it is made up of solid material that has been removed from the water. As sludge is composed of waste originating from various sources, it contains not only common microorganisms and floating and dissolved solids, but also carbon compounds, which are usually derived from the waste of living organisms, and inorganic substances, such as heavy metals and toxic compounds [1]. If these substances are released into the environment without proper treatment, they can contaminate water bodies and cause eutrophication, or if they penetrate into the soil, they can lead to plant death. Additionally, the release of sulfide from the sludge into the air can cause air pollution [2]. Therefore, the sludge is then further processed to (i) reduce the toxicity, (ii) recover any valuable nutrients, (iii) produce biogas and (iv) reduce greenhouse gas production. Sludge from wastewater treatment plants can be put through anaerobic digestion process, in which the organic matter is broken down in the absence of oxygen. Methane and carbon dioxide are the main components of the biogas produced in this way [3,4]. However, digestion efficiency and gas production are affected by many factors, such as sludge characteristics and external factors, such as operation, temperature, and pH inside the digester [5]. The demand for energy in modern society is constantly increasing, while traditional fossil energy is still the main source [6]. With the growth of the global population and economy, the enhancement of people's awareness of environmental protection and the adjustment of energy policies, the diversification trend of human energy demands is also increasing. Thus, the demand for renewable energy is also showing an

increasing trend. Gas is a renewable gaseous fuel derived from biomass that can be used as an alternative to traditional fossil fuels. For that reason, energy harvesting from waste sludge through anaerobic digestion could help in energy recovery and hence improving sludge anaerobic digestion efficiency, and gas production is an important issue in current WWTP research and practice [7,8]. All in all, through the anaerobic digestion of sludge, the amount of waste disposed and the negative impact of landfills on the environment are reduced, and the gas generated during the digestion process is used as energy, reducing greenhouse gas emissions and improving the environment and soil quality [9].

Anaerobic digestion takes place in the absence of oxygen. Anaerobic microorganisms, such as bacteria, produce biogas by decomposing organic matter. This process is divided into four stages: hydrolysis, acid production, acetic acid production, hydrogen production, and methane production [10,11]. Generally, the traditional WWTP uses a two-stage digestion process [12,13]. Before anaerobic digestion, pretreatment is required, such as thermal pretreatment, ultrasonic pretreatment, and chemical pretreatment. The raw sludge is sent to the thickening tank for treatment and then continuously or batchwise put into the first-level biogas digester for stirring [14,15]. Stirring can make the sludge fully contact the microorganisms at the bottom and keep the mixing state in the biogas digester stable [16,17]. The generated methane is then collected into the methane production unit, and the remaining sludge is sent to the secondary digester for concentration and supernatant separation. In general, mesophilic anaerobic digestion is used during the operation of WWTP biogas digesters, and the temperature is maintained at 30~37 °C [18]. As a parameter to judge whether the digestion process is running normally, pH is generally maintained at 6.8~7.2 [19]. However, due to the complexity of anaerobic digestion, the overall digestion efficiency and gas production are in general low. Therefore, in order to improve its efficiency and methane production, data mining techniques are used to analyze relevant data in the anaerobic digestion process, identify and analyze relevant variables and parameters in the anaerobic digestion process, and then propose optimized parameters to improve efficiency and gas production.

In addition, the methane produced from anaerobic digestion can be used to provide energy for equipment operation within the sewage treatment plant, thus reducing operational costs. Moreover, methane is a clean and renewable energy source that can be used as fuel in various fields, such as transportation and residential gas supply. With this in mind, the development of anaerobic digestion technology can not only solve the problem of organic waste disposal, but also contribute to the development and utilization of new energy sources.

The performance of anaerobic digestion systems depends on various factors, including growth factors, operational parameters, system type, and digester type [20]. The digester can control growth factors, such as temperature, pH, and organic acids, while operational parameters, such as the HRT (hydraulic retention time) and OLR (organic loading rate) can directly affect the system's stability and treatment efficiency and affect the metabolism of growth factors. Due to the complexity of anaerobic digestion, the overall digestion efficiency and gas production are, in general, low. Thus, effectively controlling the OLR and HRT is one of the methods to improve the efficiency and gas production of the anaerobic digestion [21].

In order to improve anaerobic digestion efficiency and methane production, data mining techniques are used to analyze relevant data in the anaerobic digestion process. Due to the large amount of data, data mining techniques can be used to automate the analysis and discover trends and patterns in the data, establish corresponding data models, analyze parameter data under different operations, and find the optimal operating parameters, effectively achieving the goal of optimizing production. In addition, it is possible to predict the efficiency and gas production of anaerobic digestion, to adjust parameters and treatment processes in a timely manner. Identifying and analyzing relevant variables and parameters in the anaerobic digestion process can propose optimized parameters to improve efficiency and gas production.

Elena Rossi et al. [22] aimed to develop an experiment using a multiple regression model to predict biogas production from the dry anaerobic digestion of organic waste. In this paper, the author first evaluated the correlation results between variables by analyzing the Pearson correlation analysis and then used MLR and selected the temperature, pH, TS content, mixture, carbon–nitrogen ratio, OLR and HRT, nutritional availability, and toxic compounds as input variables, with biogas as the output model. Experiments demonstrate that MLR can be used for preliminary evaluation and potential energy prediction. Fuqing Xu et al. [23] aimed to develop a method to predict the gas production of lignocellulosic biomass in low-temperature solid-state anaerobic digestion (SS-AD). The authors used MLR to predict gas production, by examining different types of biomasses and chemical compositions and operating parameters to predict gas production and compared the accuracy of prediction models with traditional prediction models. Through the experimental data, the establishment of a multiple regression model can accurately predict the gas production of different biomasses in SS-AD. The results show that this forecasting method can provide a valuable reference for forecasting.

NA Perendeci et al. [24] analyzed the correlation between the biochemical composition of microalgal and cyanobacterial biomass and gas production using Pearson correlation analysis. The gas production and biochemical composition of microalgal and cyanobacterial samples from different sources were assessed through Pearson correlation analysis. The results showed that the gas production and gas production were positively correlated with the total protein, total sugar, total organic carbon, and total ash content of the biomass and negatively correlated with the crude fat and cellulose content of the biomass. This result can be used to analyze biochemical components to predict gas production, which provides a reference value for predicting and optimizing gas production. Yongwoon Park et al. [25] used Pearson correlation analysis to study the effect of different biomass mixtures on steam production during anaerobic digestion and calculate the Pearson correlation coefficient between different biomass mixing ratios and gas production. The results showed that the mixing ratio of different biomasses had a significant effect on the gas production. The Pearson correlation analysis can provide value for predicting and optimizing the gas production of anaerobic digestion.

Artificial Neural Networks (ANNs) are mathematical models that simulate the artificial neural network of the human brain and are used to deal with nonlinear problems and perform pattern recognition [26,27]. The use of ANNs in the field of sludge anaerobic digestion has been rarely explored and applied in real field applications [28–30]. Previously, Liliana Mafalda studied the use of ANN methods to predict gas production and chemical oxygen demand (Chemical Oxygen Demand COD) removal rates during anaerobic digestion [31]. The model uses six parameters. The author standardized the initial data and used ANN and BP (Backpropagation Neural Network—BP) to train and verify the model. This experiment proved that the ANN method can predict the gas production and COD removal rate well in the anaerobic digestion process. ANN can be used as an effective tool to predict key parameters in the anaerobic digestion process. However, the authors did not analyze the error of the prediction results, nor did they verify the stability and reliability of the model, all of which require further research. Yuchen Wu et al. established a backpropagation neural network and used a genetic algorithm to establish a model to optimize anaerobic digestion. The model can be used to predict gas production at different temperatures. The results show that the model has a good fitting degree with the actual data and has important practical application value. However, the author did not fully discuss the scope and limitations of the model, which will affect the accuracy of the model in practical applications [32]. P. Sakiewicz et al. proposed an innovative artificial neural network method for simulating the operation of a biogas–wastewater treatment system. The method predicts the relationship between the output of the system and the parameters of the system's operation. The results show that the model can predict the impact of operating parameters on system performance and provide a reference for system optimization. However, the author did not conduct a comparative analysis between other modeling

methods and the ANN and did not fully discuss the scope of the application and stability of the model [33]. Through numerous experiments, it can be proven that an ANN is useful for the optimization of anaerobic digestion parameters. It can model complex nonlinear relationships and conduct the comprehensive analysis and optimization of multiple factors, thereby effectively improving anaerobic digestion efficiency and gas production.

The ANN is applied in many fields, such as digital twin, mathematics field, medical field, etc. T.I. Zohdi [34] used self-adaptive digital twin technology to simulate and predict the spread of flames in complex environments and used digital twins to model and predict fire spread trends. Another article of the authors [35] optimized the digital twin framework applied to aerial firefighting and pilot safety; by using machine learning algorithms (MLA) to determine the optimal dynamics of the aircraft, the high flame-retardant release efficiency is maximized. Jinyong Ying et al. [36] proposed a new deep learning structure of a multi-scale fusion network, which can accurately calculate the solution while maintaining the physical properties and continuity of the solution and can solve the ellipse interface problem well. Yinghao Chen et al. [37], aiming at the many similarities and overlaps between Crohn's disease (CD) and intestinal tuberculosis (ITB) clinically, used the fusion neural network (FCNN) method to distinguish the two CDs and ITB and trained the patient's brain through FCNN. The relevant inspection indicators were used to establish a model, and the accuracy of the model is higher than that of MLR and other models, showing higher performance and accuracy.

As a new energy source, the output of methane in WWTPs is not high. This study aimed to investigate and identify the underlying factors contributing to the low digestion efficiency and gas production in a wastewater treatment plant (WWTP) by collecting and processing mining data, finding the best operating parameters, and optimizing and predicting anaerobic digestion efficiency and gas production. By using data mining, Pearson correlation, multiple linear regression, and artificial neural networks are analyzed for sludge anaerobic digestion in order to improve its efficiency and application.

## 2. Materials and Methods

### 2.1. Data Mining

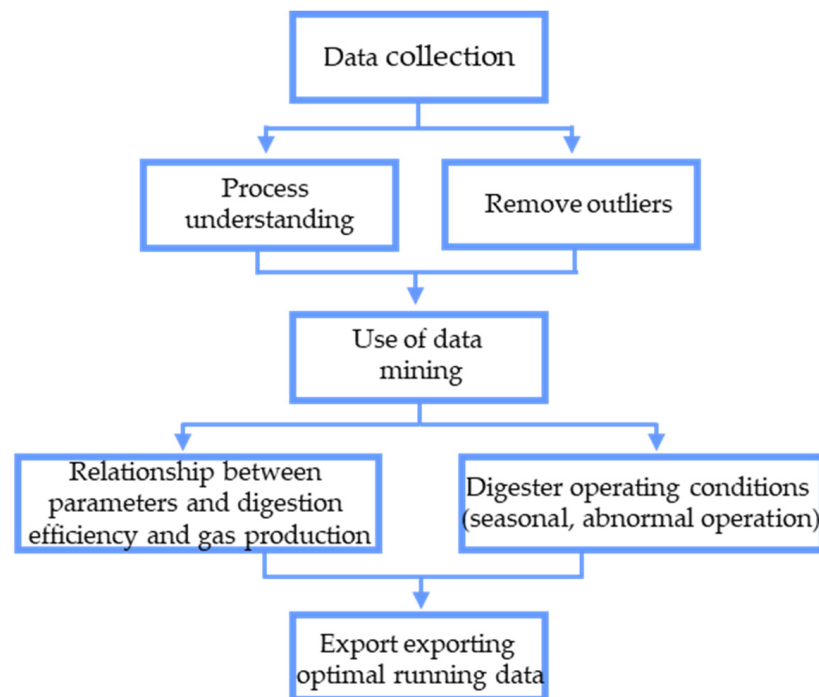
The data mining was collected from the Korean Wastewater Treatment Plant in Daegu. The wastewater treatment facility is combined with an anaerobic digestion process where the effluent is sequentially filtered and purified through sand chambers and settling tanks, and the remaining material is sent to thickening tanks and then biologically treated, scilicet, fermented in digesters where the resulting biogas is collected and used in boilers or other applications and the remaining dewatered sludge is burned as fuel. The WWTP provided operational data for almost seven years, which recorded various parameters during the operation of the WWTP, including sludge thickening, digesters, storage tanks, dewatering, and gas collection. The analysis in this study focused on digester digestion variables that are closely related to the anaerobic digestion process of the sludge. The data used in the study were collected automatically by the wastewater treatment plant on a daily basis and recorded for seven years (January 2014 to December 2021).

### 2.2. Statistical Pre-Analysis

The initial stage of statistical analysis involves identifying the relevant variables of interest, including both independent and dependent variables. After identifying the variables, exploratory data analysis (EDA) is conducted to better understand the data and uncover any patterns or trends. This is accomplished by utilizing various visualization techniques, such as scatterplots or histograms, to examine the distribution of the data and identify relationships between variables.

Seasonal data analysis of this data provides insight into the performance of the digester digestion process and explores how to optimize anaerobic digestion parameters to improve treatment efficiency and gas production. Next, data mining was performed by normalizing and removing outliers to determine the optimal operating factors that would improve

the efficiency of the digestion process. The relationship between operational factors and digestion efficiency was tested, and the conditions for the optimal operation of digestion were then determined. Finally, based on the preliminary results, the optimal operating factors that can increase digestibility were successfully derived (Figure 1). The main factors in the operation of the biogas digester include sludge inflow, temperature, pH, TS, VS, VS%, organic acid, alkalinity, OLR, and HRT. In addition, by using four scores, the season data are divided into three categories, the upper quartile (Q3, Q4) is the Good Case, the lower quartile (Q2, Q1) is the Bad Case, and the middle data are the Normal Case.



**Figure 1.** Framework for running experiments.

### 2.3. Pearson Analysis

In the research, firstly, Pearson correlation analysis was used to explore and verify the correlation between digestion efficiency and methane production and other variables. In order to ensure the quality and accuracy of the data, processed and standardized data were used. Pearson correlation analysis is a descriptive statistical analysis that describes the linear relationship between variables in a dataset and is often used to measure the correlation between two variables. The value of this correlation coefficient ‘ $r$ ’ (Formula (1)) is between 1 and  $-1$ , and the closer it is to 1 or  $-1$ , the stronger the correlation between the two variables, that is, a positive or negative correlation. When  $r$  is between 1 and  $-1$ , it can indicate that there is some degree of correlation between the two variables, for example,  $r = 0.3$ , which means that there is a weak correlation between the two variables. An  $r$  value closer to 1 and  $-1$  means that the stronger the correlation between the two variables, the more there is a complete positive or negative correlation between the two variables, meaning that when one variable increases, the other variable decreases. However, Pearson correlation analysis can only describe whether there is a linear relationship between two variables, not simply a causal relationship between two variables [38]. When the Pearson correlation coefficient is close to 0, it indicates that the degree of correlation between the two variables is very low, that is, there is an irrelevant or nonlinear relationship. The results of this statistic can play an important role in data analysis and forecasting because it can help determine the strength and direction between variables [39]. In addition, when performing Pearson correlation analysis, in addition to calculating the correlation coefficient  $r$ , it is also necessary to verify the significance level,  $p$ -value, of the correlation coefficient; the  $p$ -value



is an indicator used to judge whether the Pearson correlation coefficient of the sample data is statistically significant, and the results with a  $p$ -value less than 0.05 are generally considered to be statistically significant [40].

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (1)$$

$x, y$ : Variables.

$x^2, y^2$ : The square of the variables.

$r$ : Pearson correlation coefficient.

$n$ : Variable coefficient.

#### 2.4. Multiple Linear Regression Analysis (MLR)

Multiple linear regression analysis (MLR) is a traditional statistical method that establishes the relationship between variables based on a mathematical model. It can establish a linear relationship between the response variable and multiple predictor variables and use this relationship to make predictions. One of the advantages of MLR is the rapid establishment of models and predicting the outcome [41]. MLR was performed on parameters that have a correlation between digestion efficiency and gas production. MLR was used to model a linear relationship between two or more independent variables  $x$  and a dependent variable  $y$ . The general form of the multiple regression model is shown as Formula (2) [38]

$$y = a + b_1x_1 + b_2x_2 + \dots + b_px_p + \epsilon \quad (2)$$

$a$ : Y-intercept.

$y$ : dependent variable.

$x_1, x_2 \dots x_p$ : independent variable.

$b_1, b_p$ : parameter.

$\epsilon$ : the error term.

When building a regression model, it is assumed that the relationship between the model and the actual value is approximately reasonable; that is, a linear equation is satisfied between the predictor variable ' $y$ ' and the dependent variable ' $x$ '. The least squares method, Formula (3), is used to estimate the regression coefficients, the purpose of which is to find the minimum sum of squares of the error between the model predicted value and the actual value. The error between the predicted value of the regression model and the actual value is minimized by minimizing the error sum of squares. The process of minimizing the sum of squared errors is used to find the optimal regression coefficient estimates, so that the prediction values of the regression model are more accurate, and the model ability is improved.

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (3)$$

$a$ : Y-intercept.

$\hat{y}$ : predicted Value.

$x_1, x_2 \dots x_p$ : independent Variables.

$b_1, b_p$ : parameters.

$R^2$  is the coefficient of determination, which is a measure of the fitting degree of the regression model and is used to indicate the proportion of the predicted value that can be explained by the model. The value ranges from 0 to 1, the difference between the predicted value and the actual value is small, and the closer  $R^2$  is to 1, on the contrary, there is no relationship between the predicted value and the actual value,  $R^2 = 0$ . Formula (4) and related Formulas (5)–(7) are as follows:

$$R^2 = \frac{SSR}{SST} \quad (4)$$

$R^2$ : coefficient of Determination.

*SSR*: regression sum of squares.  
*SST*: total sum of squares.

$$SST = SSR + SSE \quad (5)$$

*SST*: total sum of squares.  
*SSR*: regression sum of squares.  
*SSE*: error sum of squares.

$$SSR = \sum (\hat{y}_i - \bar{y})^2 \quad (6)$$

*SSR*: regression sum of squares.  
 $\hat{y}_i$ : predictive value.  
 $\bar{y}$ : the mean of the sample.

$$SSE = \sum (y_i - \hat{y}_i)^2 \quad (7)$$

*SSE*: error sum of squares.  
 $\hat{y}_i$ : predictive value.  
 $y_i$ : the actual value of the  $i$  observation.

However, the performance of the regression model is not only evaluated based on the coefficient of determination  $R^2$ , but also needs to use other indicators, such as the percentage error (*PE*), root mean square error (*RMSE*), mean absolute error (*MAE*), etc., to evaluate the accuracy of the model. *PE* is the percentage used to evaluate the difference between the predicted result and the real result; *RMSE* is an indicator to measure the difference between the predicted value and the real value; *MAE* is an indicator to measure the difference between the predicted value and the real value. *PE*, *RMSE*, and *MAE* are all indicators used to evaluate the accuracy of the model. The smaller the error value, the higher the accuracy of the model, and vice versa, the lower the accuracy of the model. However, *PE* is used to compare the prediction accuracy of different models, and *RMSE* and *MAE* are used to evaluate the prediction accuracy of a single model. Relevant formulas such as (8)–(10) are as follows.

$$PE = \frac{[(\hat{y}_i - y_i)]}{y_i} \times 100 \quad (8)$$

*PE*: percentage error.  
 $\hat{y}_i$ : predictive value.  
 $y_i$ : the actual value of the  $i$  observation.

$$RMSE = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{n}} \quad (9)$$

*RMSE*: root mean square error.  
 $\hat{y}_i$ : predictive value.  
 $y_i$ : the actual value of the  $i$  observation.  
 $n$ : total samples.

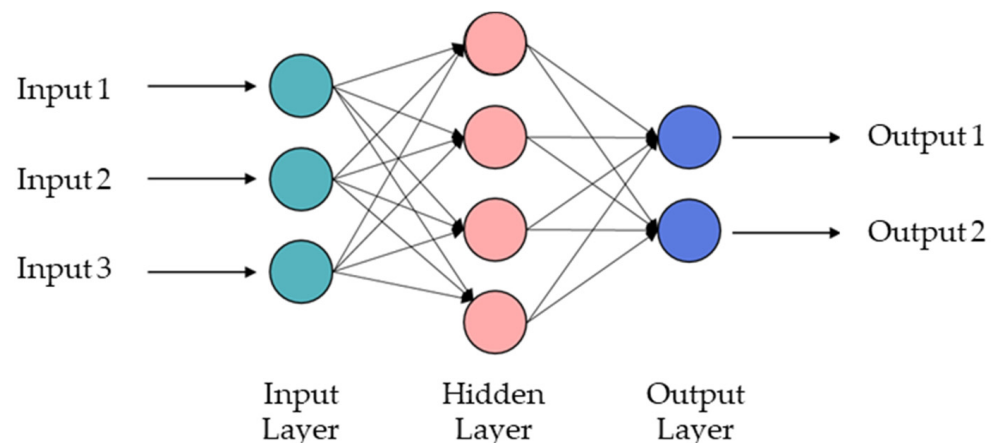
$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n} \quad (10)$$

*MAE*: mean absolute error.  
 $\hat{y}_i$ : predictive value.  
 $y_i$ : the actual value of the  $i$  observation.  
 $n$ : total samples.

### 2.5. Artificial Neural Network (ANN)

An artificial neural network (ANN) is a technique that mimics the workings of neurons in the human brain by learning patterns between inputs and outputs to predict outcomes [42]. Its power lies in its ability to capture complex nonlinear relationships between input and output variables, and it can be used to process large, complex datasets. It can learn from data, adjust weights and biases, and predict them to generate new data.

In this study, the artificial neural network multilayer perceptron (Multilayer Perceptron-MLP) was used to avoid the performance problems caused by a large number of neurons and too few neurons. MLP is a feed-forward neural network consisting of an input layer (independent variable), a hidden layer, and an output layer (dependent variable) (Figure 2). The input layer receives data, the hidden layer processes the data, and the output layer produces the result [33]. The connection between the input layer and the hidden layer is the weight of the input data, and the link between the hidden layer and the output layer is the weight of the hidden layer nodes. Each node in an MLP has an activation function that transforms the node's weighted input into an output signal. The connections between these nodes are achieved through weights, which determine how much each node contributes to the output. In MLP, the output of each node is connected to all nodes in the previous layer, and each connection has a weight, and these weights are used to calculate the output of each node. Therefore, the hidden layer nodes and weights affect the output layer results, and the adjustment of weights is a key training process to ensure that the network produces accurate output results. Both the hidden layer and the output layer have a bias (bias) node, of which output is a constant, which is used to adjust the offset of the entire network.



**Figure 2.** Principles of artificial neural network.

The purpose of this study was to investigate the improvement in anaerobic digestion efficiency and gas production during the digestion process. Therefore, the digestion efficiency and gas production were selected as output variables, and many factors related to anaerobic digestion, such as sludge inflow, temperature, pH value, TS, VS, VS%, organic acid, alkalinity, the HRT, and OLR were used as input variables, and the ANN model was established for prediction and analysis, so as to reflect the digestion efficiency and gas production of the anaerobic digestion process.

## 3. Results and Discussion

### 3.1. Effect of Seasonality on Gas Production

Generally, gas production is affected by seasonal factors, and the higher the temperature, the higher the gas production. This is because the higher temperature promotes the metabolism and growth of microorganisms in the digester, increasing the degradation rate of organic waste. Relevant studies have shown that the digesters were monitored in autumn and winter, and the results showed that the gas production in autumn was



significantly higher than that in winter, and the gas production when the temperature was low was lower than that when the temperature was high [43].

Dong Jin Lee et al. [44]. investigated the seasonal effects of the organic loading rate and acid phase on methane production during anaerobic digestion at a food wastewater treatment plant in southern Korea. The results showed that seasonal changes have a significant impact on methane production during anaerobic digestion. Due to the high temperature and rain in the summer, the acid production in the digester of the food processing plant is high, and the concentration of VFAs increases sharply, which inhibits methane production and leads to a decrease in gas production in the summer. In November, methane production rose, and gas production increased due to cooler temperatures and less rainfall.

### 3.2. The Relationship between Gas Production and Season

Based on the existing gas production data, this study uses the quartile approach to divide the data set into three categories [45]: Good Case, Bad Case, and Normal Case. Among them, Good Case and Bad Case represent the upper quartile and lower quartile respectively, while Normal Case refers to the data between these two quartiles. This classification helps to analyze gas production data and identify trends. In particular, quartile values are crucial for distinguishing Good Cases from Bad Cases and represent boundaries that divide gas production data into different categories. This classification method can effectively help researchers better understand gas production data, extract useful information, and provide a reference for follow-up research.

The upper and lower quartiles divide the dataset into four equal parts. In this method, the upper quartile of digestion efficiency and gas production is defined as the Good Case, and the lower quartile is defined as the Bad Case. Specifically, Good Case values for digestive efficiency range from 45.1% to 83%, Bad Case values range from 0% to 29%, and values between 29% and 45.1% are considered the Normal Case. Good Case values for gas production range from 14,000 to 21,000 m<sup>3</sup>/day, Bad Case values range from 1400 to less than 10,000 m<sup>3</sup>/day, and values between 10,000 and 15,000 m<sup>3</sup>/day are the Normal Case. These interquartile ranges can be used to assess the performance of the digestion efficiency and gas yield production process. When the digestion efficiency is higher than 45.1% or the gas production is higher than 14,000 m<sup>3</sup>/day, this means the digestion performance is good; on the contrary, when the digestion efficiency is lower than 29% or the gas production is lower than 10,000 m<sup>3</sup>/day, this means the digestion performance is poor. Therefore, this statistical method can be used to monitor and evaluate the performance of the production process and provide a way to judge the quality of the production process based on statistical principles. Therefore, the low production in the summer and autumn is classified as a “Bad Case”, and the high production in the spring and winter is classified as a “Good Case”.

Gas production and digestion efficiency are affected by seasonal factors. Factors, such as external temperature and rainfall, will affect the digester, thereby affecting gas production and digestion efficiency [44,46]. According to the results in Figures 3 and 4, the Good Case ratio is relatively large in the winter and spring; the percentages of digestion efficiency are 10% and 49%, and the percentages of gas production are 47% and 80%, respectively. In contrast, the digestion efficiency and gas production in the summer and autumn are not good, accounting for 57% and 22% of the total percentage respectively, which is different from the general situation that the digestion efficiency and gas production are higher in the summer and autumn. According to the research of E Sánchez et al. [47], tropical climate, especially in the rainy season, due to the high temperature and high humidity environment, will have an adverse effect on anaerobic digestion, because the temperature of the digester is too high, with too much water, so the substrate in the digester is diluted, and the humid and high-temperature environment reduces the activity of microorganisms, which in turn affects the stability and digestion efficiency of the internal environment of the anaerobic digester and further affects the gas production. Further explained, this seasonal trend in gas production and digestion efficiency could be attributed to the influence of the

rainy season in South Korea. During the rainy season, the concentration of organic matter and substrates in the wastewater flowing into the wastewater treatment plant (WWTP) decreases. This reduction ultimately affects the efficiency of anaerobic digestion, the main process that produces gas in wastewater treatment plants. Therefore, gas production is lower in the summer and autumn. Conversely, in the spring and winter, fewer raindrops and increased concentrations of organic matter and substrates boosted the efficiency of anaerobic digestion, leading to increased gas production, and thus, this fell into the “Good Case” category.

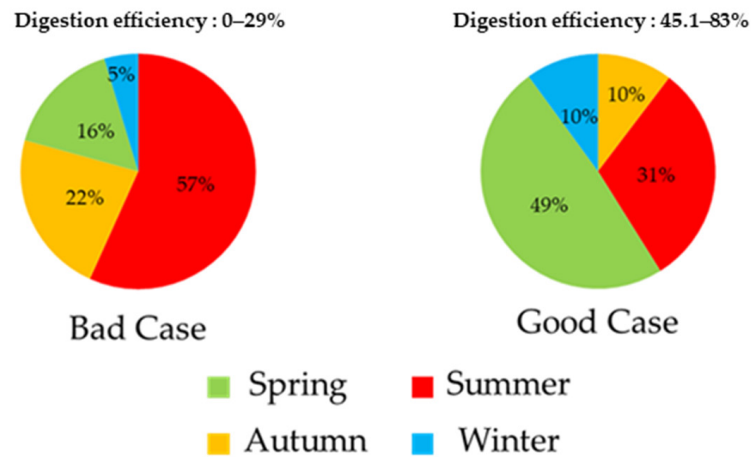


Figure 3. Digestion efficiency under a seasonal influence.

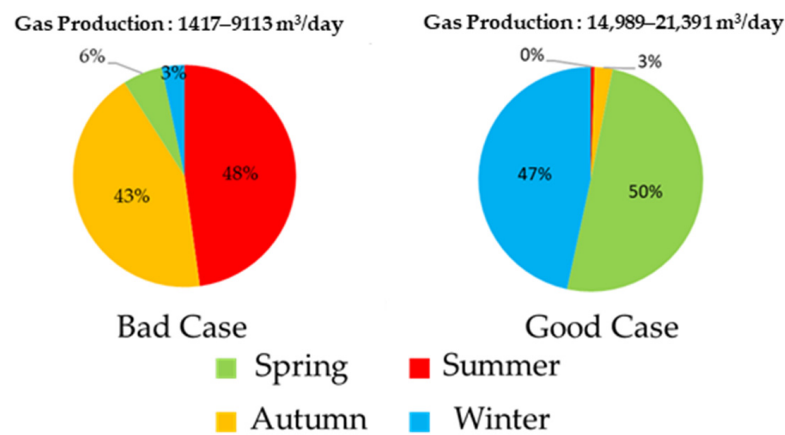


Figure 4. Gas production under a seasonal influence.

Tables 1 and 2 provide information on the various parameters associated with the digestion system. This table shows the various parameters within the digester, including temperature, pH, and others, along with the maximum, minimum, and average values for each parameter in both Good and Bad Cases under different conditions, and these data assist in the assessment of the performance and reliability of the system. A Good Case is a situation where the system is performing as expected, and a Bad Case is a situation where the system is not performing as expected. No significant difference between the main parameters was observed in the values between the Good Case and the Bad Case, indicating that the system operates in a relatively stable state at this stage. This situation may be due to low gas production due to seasonal factors. South Korea’s summer and autumn are the rainy seasons. When rainwater flows into the sewage treatment plant, the temperature and pH in the digester change. Typically, in order to maintain a good digestion environment, the temperature fluctuation should not exceed 1 °C, and if it exceeds 2–3 °C, it will have a great impact on anaerobic digestion [48]. In addition, the organic matter

load and HRT, marked in red in the table, are parameters that can only be adjusted in the actual operation of this study. This means that for other parameters, accurate data may not be able to be obtained through a laboratory operation. Therefore, during the actual operation, special attention needs to be paid to these operable parameters in order to be able to optimize the performance and efficiency of the system. At the same time, for those parameters that cannot be changed, we need to treat them as constants in the operation of the system in order to better understand the performance and behavior of the digester. Based on this, statistical analysis is necessary, which can help us understand in more detail the impact of existing parameters on the production gas yield of WWTPs in the west. By performing statistical analysis of the data, we can identify and determine which parameters have the most significant impact on gas production and how to optimize these parameters to improve the efficiency and quality of gas production. This statistical analysis can help us better understand and control the variables in the production process so that we can more reliably produce high-quality gas products. On this basis, we can formulate more effective production strategies and programs to improve the production capacity and benefits of a western WWTP.

**Table 1.** Parameter values under abnormal gas production conditions.

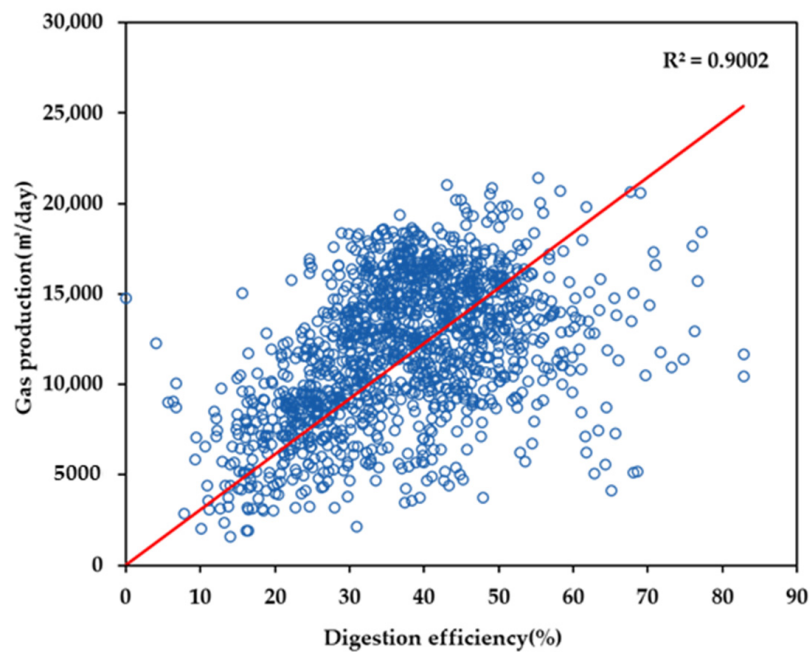
Parameter				Good Case			Bad Case		
				Average	Minimum	Maximum	Average	Minimum	Maximum
Inflow Sludge	TS	mg/L	32,978	23,280	49,115	27,801	15,640	45,470	
	VS	mg/L	25,336	17,565	37,085	19,018	10,860	31,015	
	VS (%)	%	77	68	84	68	60	80	
Digester	Inflow	m <sup>3</sup> /day	1348	515	2104	1236	540	1976	
	Temperature	°C	34.5	27.3	40.2	37.4	26.1	45.7	
	pH		7.3	7.0	7.7	7.4	7.0	7.9	
	TS	mg/L	22,282	11,685	45,300	23,231	9505	43,315	
	VS	%	66	40	75	58.0	36.4	68.4	
	Acid	mg/L	219	43	799	62	25	930	
	Alkalinity	mg/L	2765	1068	4040	1989	1025	3130	
	Digestion Efficiency	%	42.2	15.6	77.2	34.5	10.7	68.6	
	OLR	kg/m <sup>3</sup> day	1.34	0.39	2.24	0.98	0.49	1.77	
	HRT	day	20.2	12.6	48.0	19.5	11.0	50.9	
Secondary digester	Sludge Volume	m <sup>3</sup>	1145	27	1965	1169	158	1914	
	TS	mg/L	21,663	16,425	34,500	24,021	3985	38,955	
Digestive gas	VS	%	65.3	55.5	71.9	59.1	48.9	77.9	
	Solid Content	kg/day	24,936	9870	43,975	27,668	5679	63,341	
	Gas production	m <sup>3</sup> /day	16,776	14,991	21,391	7376	2151	9699	
	CH <sub>4</sub>	%	63	44	65	63.6	61.0	67.7	
	CO <sub>2</sub>	%	37	32	37.5	36.4	32.3	39.0	

Digestion efficiency generally refers to the rate at which organic matter is degraded by anaerobic digestion microorganisms into organic products, and gas is produced in the process. In the example in Figure 5, the relationship between the digestion efficiency and gas production was examined. An  $R^2$  value of 0.9002 indicated a very strong positive correlation between sludge anaerobic digestibility and gas production, implying a direct relationship between these two variables. It has been shown that a higher sludge digestibility leads to an increase in gas production during anaerobic digestion, and conversely, a lower digestibility leads to a decrease in gas production [49]. The strength of this relationship was significant, and 90.02% of the variation in gas production could be explained by the variation in sludge

anaerobic digestibility. In practice, this information is useful to the anaerobic digestion operator and can be used to optimize the process.

**Table 2.** Parameter values under abnormal operating conditions of digestion efficiency.

Parameter			Good Case			Bad Case		
			Average	Minimum	Maximum	Average	Minimum	Maximum
Inflow Sludge	TS	mg/L	30,335	16,365	50,335	27,994	17,565	52,945
	VS	mg/L	22,985	11,080	41,345	19,150	11,390	32,220
	VS (%)	%	75	62	84	68	57	79
Digester	Inflow	m <sup>3</sup> /day	1277	533	2133	1560	820	2536
	Temperature	°C	36.2	26.4	46.5	36.5	26.1	42.5
	pH		7.4	6.8	7.9	7.2	7.0	7.8
	TS	mg/L	23,255	12,538	48,260	22,461	10,375	34,695
	VS	%	59	32	72	62	26	75
	Acid	mg/L	119	23	799	158	30	930
	Alkalinity	mg/L	2486	1068	4040	1773	870	3270
	Digestion Efficiency	%	52.3	45.1	82.8	22.9	4.1	29.9
	OLR	kg/m <sup>3</sup> day	1.20	0.39	2.26	1.12	0.54	1.94
	HRT	day	21.0	11.0	38.7	17.4	10.8	33.5
Secondary digester	Sludge Volume	m <sup>3</sup>	1162	439	8953	1156	338	1813
	TS	mg/L	22,198	3985	37,160	22,416	15,840	38,955
Digestive gas	VS	%	62.2	51.0	78.0	62.0	50.1	70.4
	Solid Content	kg/day	25,309	5679	48,382	26,234	9287	63,341
	Gas production	m <sup>3</sup> /day	99	27	239	80	24	290
	CH <sub>4</sub>	%	13,555	3731	21,391	8679	1607	18,390
	CO <sub>2</sub>	%	62	34	69	64	55	75



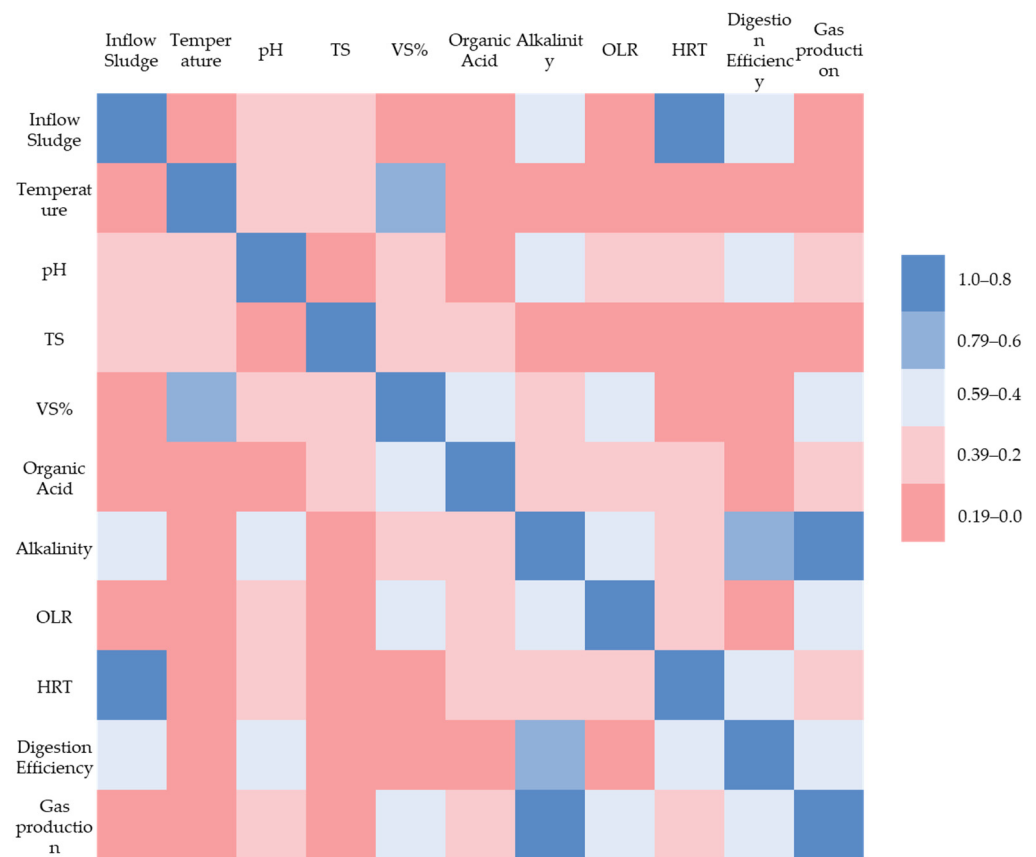
**Figure 5.** The correlation between gas production and digestion efficiency.

### 3.3. Pearson Correlation Analysis Results

Statistical conventions are convention rules in statistics that solve the problems caused by data uncertainty and objectivity. By following these rules, statistical results can be made more objective and reliable. The field is used to analyze data in many areas, including environmental science [50]. In Pearson correlation analysis, statistical conventions need to be followed to ensure the accuracy and validity of the results. For example, data preprocessing checks the normal distribution of data, calculates correlation coefficients, sets significance levels and test statistics for hypothesis testing, etc., to judge whether the statistical results are meaningful. According to statistical conventions, the absolute value of  $r$  is usually divided into the following grades: 0.8–1.0, very strong correlation; 0.6–0.79, strong correlation; 0.4–0.59, moderate correlation; 0.2–0.39, weak correlation 0.0–0.19, very weak or no correlation [51]. According to the results of Pearson's correlation analysis, there is a medium-strength relationship between digestion efficiency and sludge inflow, pH, alkalinity, and HRT, and this has no correlation with other parameters. There is a correlation between gas production and pH, VS%, OLR, HRT, and digestion efficiency and a very strong correlation with alkalinity.

Shuang Zhang et al. [52] studied the response of the semi-continuous anaerobic digestion of food waste to a gradually increasing temperature, including aspects of methanogenic bacterial communities, a correlation analysis, and energy balance. They used Pearson analysis to assess the linear relationship between the different parameters and further investigated the response of the parameters, as well as methane production. Correlation analysis results showed that methane production was positively correlated with temperature, while there was a weak correlation between pH and methane production. These two relationships are not strong, but it does not mean that there is no correlation between the two.

In this study, the relationship between the two variables can be analyzed according to the correlation coefficient in Figure 6. The correlation coefficients between digestion efficiency and influx, pH, and HRT were  $-0.498$ ,  $0.498$ , and  $0.490$ , respectively, indicating that there was a moderate correlation between them. This means that these three variables are related and interrelated to the efficiency of digestion during digestion. The correlation coefficient with alkalinity was  $0.631$ , showing a strong correlation between them, so the effect of alkalinity on digestion efficiency was more significant. The correlation coefficients among VS%, OLR, and digestion efficiency were  $0.464$ ,  $0.508$ , and  $0.590$ , respectively, indicating that there was a moderate correlation among them. This means that these three variables are correlated and interrelated with respect to gas production during digestion. The correlation coefficient with alkalinity was  $0.793$ , indicating that there is a strong correlation between them, so the impact of alkalinity on gas production is more significant. The correlation coefficients between gas production and pH, organic acids, and HRT were  $0.334$ ,  $0.381$ , and  $0.282$ , respectively, indicating that there is a weak correlation between them, which indicates that there is a weak linear relationship between gas production and pH, organic acids, and HRT. Pearson correlation analysis can only detect and describe linear relationships and cannot detect and describe nonlinear relationships. The weak correlation and no correlation in the Pearson correlation analysis indicated that there was a non-linear relationship between digestion efficiency and gas production and other factors. It is worth noting that Pearson correlation analysis can only prove a linear relationship between two variables rather than a causal relationship. In other words, when there is a strong relationship between variable  $a$  and variable  $b$ , it can only show that there is a linear relationship between the two variables. The relationship is strong, rather than an increase or decrease in either variable  $a$  or variable  $b$  resulting in a change in the other.



**Figure 6.** Correlation analysis result.

### 3.4. Comparison between MLR and ANN

In this study, IBM SPSS Statistics 27 was used. Two models, Multiple Linear Regression (MLR) and an Artificial Neural Network (ANN), were compared to determine which model was more suitable for the studied dataset. It should be noted that MLR and the ANN use the same data to build predictive models. During experimental modeling, a multiple linear regression model included digestion efficiency, gas production, etc. In the case of ANN, the accuracy and reliability of the prediction results of ANN were higher (Table 3). The error of the ANN model was lower than that of the MLR model in terms of the percentage error (*PE*), root mean square error (*RMSE*), and mean absolute error (*MAE*). The above four indicators were used to measure and evaluate the accuracy and reliability of the model. The final research results show that the model can accurately predict digestion efficiency and methane production with high prediction accuracy and reliability, which means that the neural network model makes more accurate and reliable predictions for our dataset. While MLR models may perform well on some datasets, this study shows that ANN models are a more reliable and accurate choice when dealing with large amounts of data. The main advantage of the ANN model is that it can handle a large amount of nonlinear data, so it has wide applicability in many practical applications. In addition, the ANN model also has self-learning ability, which can improve the performance of the model through training based on a large number of datasets. To sum up, this study uses the artificial neural network model when processing large amounts of data to obtain more accurate and reliable prediction results. Of course, this is not to say that the MLR model is useless. For some situations with less data and obvious linear relationships, the MLR model is still a viable choice.



**Table 3.** MLR and ANN coefficient and error comparison.

		Gas Production			
	Algorithm	$R^2$	$PE$	$RMSE$	$MAE$
WWTP	MLR	0.722	0.001	121.791	0.001
	ANN	0.794	0.000	104.794	0.000
		Digestion Efficiency			
	Algorithm	$R^2$	$PE$	$RMSE$	$MAE$
WWTP	MLR	0.964	0.000	0.127	0.000
	ANN	0.994	0.000	0.051	0.000

Through Pearson analysis, it can be seen that not all parameters have a linear relationship with digestion efficiency and gas production, and most parameters show a weak correlation or no correlation. That is, through Pearson correlation analysis, it can be concluded that there is a nonlinear relationship between digestion efficiency and gas production, and the advantage of ANN lies in the processing of nonlinear relationships, which can be modeled well through neurons and weights.

### 3.5. ANN Training Test Range

This section details the importance of a multilayer perceptron in training and validation and the impact of min and max unit hidden layers. In experiments, we analyzed the prediction results by setting different training and validation set ratios and min and max unit hidden layers and evaluated the performance of the model. During the research process, it is usually necessary to calculate and analyze multiple input variables to determine their impact on the output variables. This study performed calculations on 17 available input variables using IBM SPSS Statistics 27. These input variables include OLR, HRT, and other parameters that affect anaerobic digestion. The dataset consists of 96 sets of data, by default: 70% for training, 30% for validation, and also includes the minimum number of units in the hidden layer 1 and the maximum number of units, 60.

In learning, the ratio of the training set and verification set is very important; a training set that is too large or too small will affect the performance of the model. Larger training sets make the model too simple to capture complex patterns in the data, while smaller training sets tend to cause the model to overfit and fail to generalize predictions to new data. Therefore, in general, during the training process of the ANN model, the proportion of training set and verification set is kept at 70% and 30%, as well as 80% and 20% [53]. Also, the settings of the maximum and minimum hidden layers affect the performance of the multilayer perceptron. If the number of units is too small, the model will be underfitted, and too many will cause the model to be overfitted, dependent on the training data, and unable to generalize; for the new dataset, the prediction performance is not good, or it cannot be adapted well to the new dataset [54].

In order to evaluate the performance of the model, different proportions of the training set and validation set were set in the experiment, the minimum and maximum unit hidden layers were set respectively, and the performance of the model was evaluated by predicting the results. The test results show that when the training set is 70% and the verification set is 30%, when the minimum unit hidden layer is 5 and the maximum unit hidden layer is 6, the activation function of the hidden layer is the hyperbolic tangent function, and the output layer function is the identity function. Currently, the prediction effect of digestion efficiency and gas production is the best. Therefore, in this experiment, according to the prediction results, in the multi-layer perceptron, the proportion of the training set and validation set should be kept in the range of 70% and 30%, and the minimum and maximum unit hidden layers should be kept at 5 and 6, to obtain the best prediction results.

Wei-Yao Chen et al. [30] manipulated and modeled the operating parameters of biogas in a pre-commercial integrated anaerobic-aerobic bioreactor (IAAB) using ANN techniques. As a result, the COD removal rate increased by 23.3%, and the methane yield increased

by 13.4%. R. Yukesh Kannah et al. [55] predicted biogas production by using an ANN for substrate concentration in a mixed upflow anaerobic digester reactor treating landfill leachate. Different OLRs were used in the study, and the accuracy of the ANN was predicted. As a result, the methane production increased, and the COD removal rate also increased significantly.

### 3.6. Schemes for Increasing Digestion Efficiency and Gas Production

Observations from the actual operation of a WWTP show that the control and adjustment of multiple parameters are crucial for its efficient operation, including temperature, pH, flow rate, etc. There are complex interactions among these parameters, and the correct choice of parameter combinations is of great importance for the WWTP digester control parameters. The combination of these parameters can also be referred to as a vector. In this study, we chose the organic matter load and hydraulic retention time as control vectors. The OLR and HRT are two important parameters affecting digestion efficiency and methane production [56]. Among them, the OLR refers to the amount of organic matter entering the reactor per unit time, and the HRT refers to the average time that wastewater stays in the reactor. When both the OLR and hydraulic retention time are appropriate, the digestion process can be made more stable, improving digestion efficiency and methane production [57]. In the WWTP, the OLR and HRT are two important control parameters that significantly affect the degradation of organic matter by microorganisms in the digester. If the OLR is too high or the HRT is too short, the microorganisms in the digester may not be able to fully degrade the organic matter, resulting in reduced digestion efficiency and even the accumulation of volatile fatty acids, affecting the stability of the entire digestive system. On the contrary, if the OLR is too low or the HRT is too long, the microorganisms in the reactor may lose their vitality, and the digestion efficiency and methane production will also decrease. Therefore, it is necessary to balance the OLR and HRT to maintain the activity and stability of microorganisms in the digester and then increase the degradation rate of organic matter to improve digestion efficiency and gas production.

Normally, the range of the HRT in the sludge anaerobic digestion process is 15 to 30 days, and this range is obtained based on practical experience and scientific research [57,58]. However, due to the different treatment effects and raw water quality of different wastewater treatment plants, the range of the HRT needs to be determined according to the actual situation in specific practice. In this study, the HRT was assigned a range of 10 to 35 days, which is slightly wider than the usual range. This is due to the need to determine the appropriate HRT range in conjunction with the actual situation to ensure sludge stability in order to maximize digestion efficiency and gas production, taking into account the differences in factors, such as the equipment and raw water quality of WWTPs in the actual situation. In addition, the OLR value is also an important parameter in the sludge anaerobic digestion process. In this study, the range of the OLR value was also carefully considered and controlled to achieve a better treatment effect. The control of the OLR value needs to be based on the water quality of the raw water, the treatment capacity of the treatment plant, and other factors that are determined. The range of OLR values was calculated from the quartile values (minimum, first quartile, median, third quartile, and maximum) of the OLR of the wastewater treatment plant.

By sequentially combining the organic matter load in Table 4 with the HRT, the gas production and digestion efficiency were predicted, respectively. The results showed that when the organic load ranged from 1.26 to 1.46 (kg/m<sup>3</sup> day) and the HRT ranged from 10 to 15 days, the gas production was the highest, increasing by 2.8%. This shows that under the combination of organic matter load and the HRT in this range, the gas production of the biogas digester is optimal. When the organic matter load ranged from 0.86 to 1.06 (kg/m<sup>3</sup> day) and the HRT ranged from 21 to 25 days, the predicted value of the digestion efficiency was tested, and the results showed that it was 0.7% higher than the actual value (Table 5). To improve digestion efficiency and gas production at the same time, the organic matter load was finally maintained at 1.26 to 1.46 (kg/m<sup>3</sup> day), and the HRT

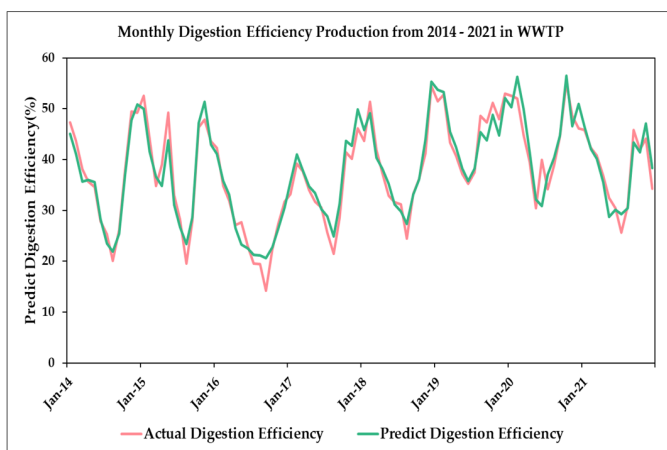
range was 26 to 30 days. The results showed that the gas production increased by 1.3%, and the digestion efficiency increased by 0.5%. (Figures 7 and 8).

**Table 4.** Value range of the HRT and OLR.

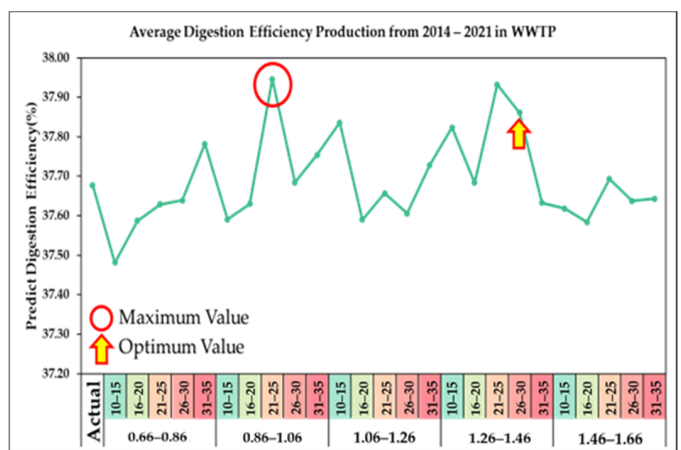
HRT (day)	Organic Loading Rate (kg/m <sup>3</sup> day)
10–15	0.66–0.86
16–20	0.86–1.06
21–25	1.06–1.26
26–30	1.26–1.46
31–35	1.46–1.66

**Table 5.** ANN training and testing time.

Training	Gas Production	Result/Gas Production	Test/Digestion Efficiency	Result/Digestion Efficiency
1	11,955.41	0.0000	37.68	0.0000
2	11,884.45	−0.0059	37.48	−0.0052
3	11,991.42	0.0030	37.59	−0.0024
4	12,063.65	0.0091	37.63	−0.0013
5	11,884.83	−0.0059	37.64	−0.0010
6	11,854.80	−0.0084	37.78	0.0028
7	11,973.67	0.0015	37.59	−0.0023
8	12,057.15	0.0085	37.63	−0.0013
9	11,897.62	−0.0048	37.95	0.0071
10	11,860.79	−0.0079	37.68	0.0002
11	11,914.42	−0.0034	37.75	0.0021
12	12,287.80	0.0278	37.84	0.0042
13	12,031.70	0.0064	37.59	−0.0023
14	11,862.88	−0.0077	37.66	−0.0005
15	12,068.04	0.0094	37.61	−0.0019
Original	11,955.41	-	37.68	-

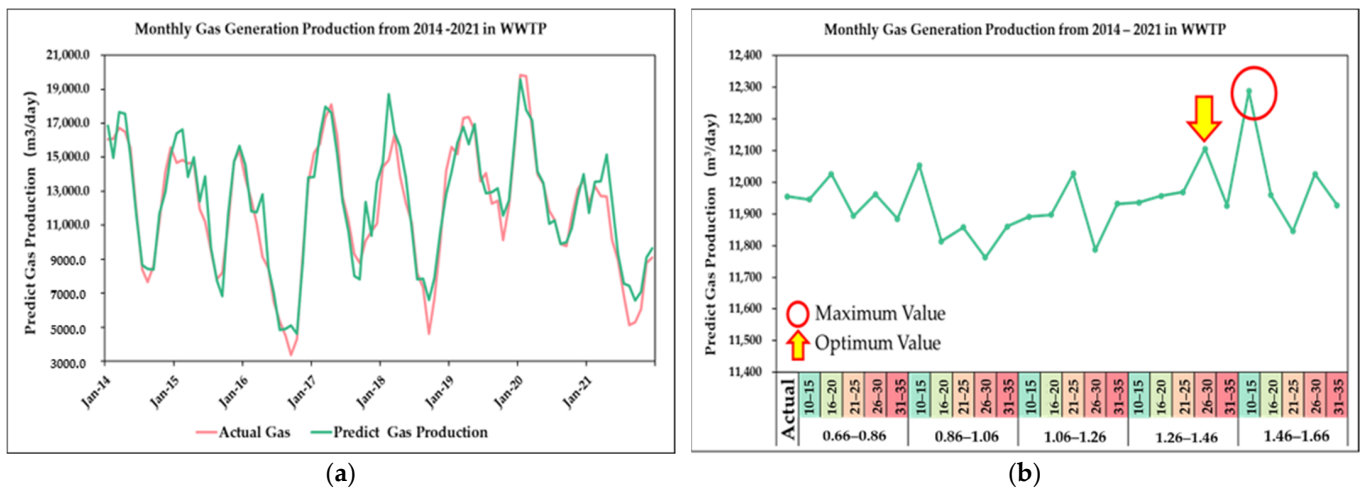


(a)



(b)

**Figure 7.** (a) Comparison of actual digestion efficiency and ANN-predicted digestion efficiency; (b) the circle represents the best prediction value of digestion efficiency, and the arrow represents the best prediction value of gas production and digestion efficiency under the same OLR and HRT conditions.



**Figure 8.** (a) Comparison of actual gas production and ANN-predicted gas production; (b) the circle indicates the best predicted value of digestion efficiency, and the arrow indicates the best predicted value of gas production and digestion efficiency under the same OLR and HRT conditions.

In the study, the influence of the OLR and HRT on the digestion efficiency and gas production of the digester was determined. The results showed that within a certain range of the OLR and HRT, the digestion efficiency and gas production can be optimized, but in some cases, the predicted value will be lower than the actual digestion efficiency, which indicates that the predictive model needs further optimization. It is important to note that these results are based on specific conditions and specific circumstances of the wastewater treatment plant. Therefore, in practical applications, it needs to be adjusted and optimized according to the wastewater treatment plant to obtain better treatment and digestion.

The results in Tables 6 and 7 show that within the studied OLR range of 1.26 to 1.46 ( $\text{kg}/\text{m}^3 \cdot \text{day}$ ) and HRT range of 26 to 30 days, the prediction model can improve digestion efficiency and gas production to a certain extent. Especially in the summer and autumn, the improvement of digestion efficiency and gas production was more significant. These results are of great significance for improving the treatment efficiency and reducing treatment costs of sewage treatment plants. However, although the error between the predicted and actual values was within an acceptable range, the digestion efficiency and gas production decreased in the winter. In addition, in the actual operation process, there may be other factors, such as the actual ambient temperature, which may affect the digestion efficiency and gas production, so further research is needed to determine the specific mechanism of these effects.

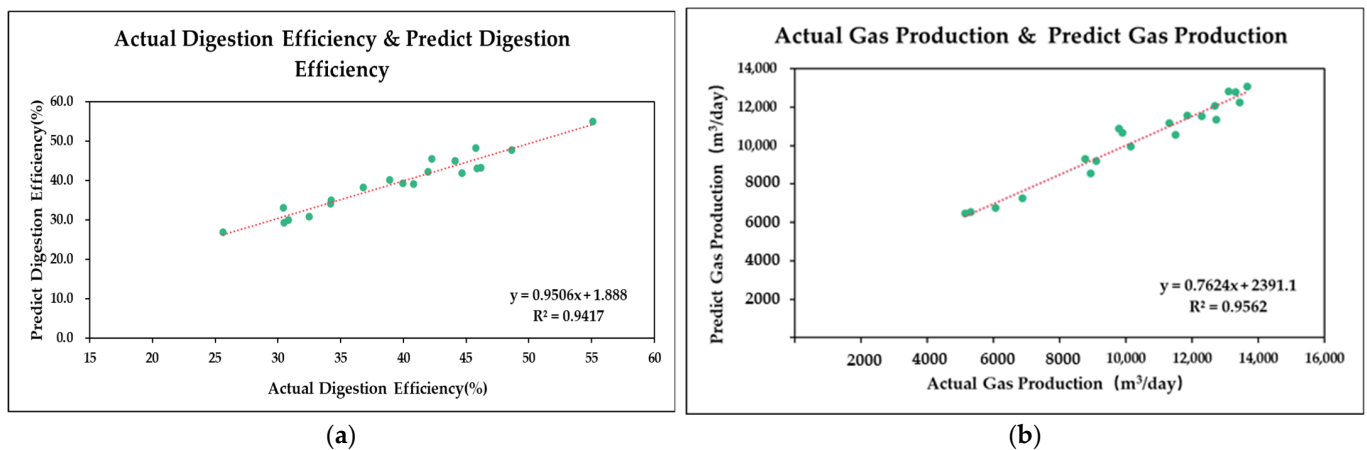
After the artificial neural network model is established, in order to judge the accuracy and reliability of the model, it is necessary to evaluate and verify the model. Among them, an important indicator is the significance between the predicted value and the actual value, which is usually measured based on  $R^2$ . It can be seen from Figure 9 that the  $R^2$  between the predicted value and the actual value of the digestion efficiency was 0.9417, and the  $R^2$  between the predicted value and the actual value of the gas production was 0.9562. This means that the model is highly predictive and fits the data well, and a high  $R^2$  value also means that the model has a high degree of confidence in the prediction, which can be used under different parameter combinations to improve gas production and digestion efficiency.

**Table 6.** Monthly forecasted operating range to improve digestion efficiency and gas production.

Month	Actual				Predicted			
	OLR (kg/m <sup>3</sup> day)	HRT (day)	Gas Production (m <sup>3</sup> /day)	Digestion Efficiency (%)	OLR (kg/m <sup>3</sup> day)	HRT (day)	Gas Production (m <sup>3</sup> /day)	Digestion Efficiency (%)
January	1.1	16.7	15,229	46.1	1.3	27	14,565	45.7
February	1.1	16.3	15,288	45.0	1.3	27.9	15,030	45.1
March	1.1	17.7	15,356	39.2	1.4	27.5	15,086	39.4
April	1.0	18	14,549	36.3	1.3	27.8	14,962	36.3
May	1.0	17.7	13,079	34.5	1.4	27.8	13,359	35.0
June	1.0	18.6	10,965	31.5	1.3	28.6	11,514	31.4
July	1.1	17	9365	28.4	1.4	28.8	9550	28.4
August	1.2	16.3	7984	27.9	1.4	28.4	8206	28.6
September	1.1	16.7	7640	33.5	1.4	28.4	8431	34.1
October	1.2	16.8	8877	41.5	1.4	28.4	9191	41.6
November	1.0	16.9	11,596	43.3	1.3	28.1	11,698	43.9
December	1.1	16.3	13,537	44.8	1.4	27.9	13,675	44.9

**Table 7.** Season forecasted operating range to improve digestion efficiency and gas production.

Season	Actual				Predicted			
	OLR (kg/m <sup>3</sup> day)	HRT (day)	Gas Production (m <sup>3</sup> /day)	Digestion Efficiency (%)	OLR (kg/m <sup>3</sup> day)	HRT (day)	Gas Production (m <sup>3</sup> /day)	Digestion Efficiency (%)
Winter	1.1	16.4	14,685	45.3	1.3	27.6	14,423	45.2
Spring	1.0	17.8	14,328	36.7	1.3	27.7	14,469	36.9
Summer	1.1	17.3	9438	29.3	1.4	28.6	9757	29.4
Autumn	1.1	16.8	9371	39.5	1.4	28.3	9774	39.9



**Figure 9.** (a) Correlation between actual digestion efficiency and predicted digestion efficiency after using the ANN; (b) correlation between actual gas production and predicted gas production after using the ANN.

Using the OLR of 1.26–1.46 (kg/m<sup>3</sup>·day) and the HRT of 26–30 days in Bad Case prediction, it can be seen from Table 7 that both digestion efficiency and gas production have been significantly improved, and for the predicted digestion, the efficiency is 5.2% higher than the actual value, and the predicted gas production is 8.8% higher than the actual value. It can be seen from Table 8. In this study, the range of the OLR and HRT can be applied to abnormal operating conditions at the same time, but the OLR and HRT

need to be adjusted according to the actual situation in operation to ensure the maximum improvement of digestion efficiency and gas production.

**Table 8.** Actual gas production vs. predicted gas production under abnormal conditions.

Parameter		Actual Bad Case	Improve Bad Case
OLR (kg/m <sup>3</sup> day)	Average	1.02	1.36
	Range	0.49–2.10	1.26–1.46
HRT (day)	Average	18.0	28
	Range	10.8–50.9	26–30
Digestion Efficiency (%)	Average	27.92	29.45 (5.2%)
	Range	5.7–68.6	26.8–54.3
Gas Production (m <sup>3</sup> /day)	Average	6766	7426 (8.8%)
	Range	1607–9107	5495–13,679

#### 4. Conclusions

In this research, by adjusting the two parameters of the OLR and HRT in the biogas digester and using artificial neural network technology to establish a prediction model, the digestion efficiency and gas production were successfully predicted and improved. During data processing, we noticed that seasonality is also a factor affecting digestion efficiency and gas production. After comparing multiple regression analysis and the artificial neural network, the ANN was chosen as the predictive model, with the parameters in the biogas digester as the input variables and the digestion efficiency and gas production as the output variables. The experimental results show that when the OLR is 1.26 to 1.46 (kg/m<sup>3</sup> day) and the HRT is 26 to 30 days, the digestion efficiency increases by 0.5% and the gas production increases by 1.3%. Under abnormal operating conditions, the digestion efficiency increased by 0.6% and the gas production increased by 1.6%.

Based on the experimental results, we believe that the artificial neural network model can effectively predict the digestion efficiency and gas production during anaerobic digestion and can significantly improve the effect of anaerobic digestion after parameter optimization. This research result is of great significance for the operation and optimization of biogas digesters and provides a feasible prediction and optimization method for research on other similar processes. Future research can further expand the dataset and optimize the prediction model to obtain more accurate and reliable prediction results.

**Author Contributions:** Conceptualization, methodology, and writing original draft preparation: Y.B.; writing, validation and formal analysis, review, and editing: R.K.; supervision, project administration, and funding acquisition: T.-G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was conducted with the support of the Daegu Green Environment Support Center R&D project in 2022, Daegu Green Environment Center (DEGEC).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This research was supported by the Daegu Green Environment Center (DEGEC) in 2022 (Project No. 22-04-02-10-16).

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Ahmad, T.; Ahmad, K.; Alam, M. Characterization of Water Treatment Plant's Sludge and Its Safe Disposal Options. *Procedia Environ. Sci.* **2016**, *35*, 950–955. [[CrossRef](#)]
2. Tobajas, M.; Verdugo, V.; Polo, A.M.; Rodriguez, J.J.; Mohedano, A.F. Assessment of Toxicity and Biodegradability on Activated Sludge of Priority and Emerging Pollutants. *Environ. Technol.* **2016**, *37*, 713–721. [[CrossRef](#)] [[PubMed](#)]
3. Di Capua, F.; Spasiano, D.; Giordano, A.; Adani, F.; Fratino, U.; Pirozzi, F.; Esposito, G. High-Solid Anaerobic Digestion of Sewage Sludge: Challenges and Opportunities. *Appl. Energy* **2020**, *278*, 115608. [[CrossRef](#)]
4. Martínez, E.; Sotres, A.; Arenas, C.; Blanco, D.; Martínez, O.; Gómez, X. Improving Anaerobic Digestion of Sewage Sludge by Hydrogen Addition: Analysis of Microbial Populations and Process Performance. *Energies* **2019**, *12*, 1228. [[CrossRef](#)]
5. Kim, D.; Phae, C. Analysis of the Effect of Sludge Reduction and Energy Production from Introducing Anaerobic Digest Plant into a Sewage Treatment Plant. *Water* **2022**, *14*, 1821. [[CrossRef](#)]
6. Simcock, N.; Mullen, C. Energy Demand for Everyday Mobility and Domestic Life: Exploring the Justice Implications. *Energy Res. Soc. Sci.* **2016**, *18*, 1–6. [[CrossRef](#)]
7. Kiselev, A.; Magaril, E.; Magaril, R.; Panepinto, D.; Ravina, M.; Zanetti, M.C. Towards Circular Economy: Evaluation of Sewage Sludge Biogas Solutions. *Resources* **2019**, *8*, 91. [[CrossRef](#)]
8. Abdelrahman, A.M.; Kosar, S.; Gulhan, H.; Cicekalan, B.; Ucas, G.; Atli, E.; Guven, H.; Ozgun, H.; Ozturk, I.; Koyuncu, I.; et al. Impact of Primary Treatment Methods on Sludge Characteristics and Digestibility, and Wastewater Treatment Plant-Wide Economics. *Water Res.* **2023**, *235*, 119920. [[CrossRef](#)]
9. Zhang, M.; Li, J.; Wang, Y.; Yang, C. Impacts of Different Biochar Types on the Anaerobic Digestion of Sewage Sludge. *RSC Adv.* **2019**, *9*, 42375–42386. [[CrossRef](#)] [[PubMed](#)]
10. Feng, L.; He, S.; Gao, Z.; Zhao, W.; Jiang, J.; Zhao, Q.; Wei, L. Mechanisms, Performance, and the Impact on Microbial Structure of Direct Interspecies Electron Transfer for Enhancing Anaerobic Digestion—A Review. *Sci. Total Environ.* **2023**, *862*, 160813. [[CrossRef](#)] [[PubMed](#)]
11. Li, G.; Xu, F.; Yang, T.; Wang, X.; Lyu, T.; Huang, Z. Microbial Behavior and Influencing Factors in the Anaerobic Digestion of Distiller: A Comprehensive Review. *Fermentation* **2023**, *9*, 199. [[CrossRef](#)]
12. Luo, L.; Xu, S.; Liang, J.; Zhao, J.; Wong, J.W.C. Mechanistic Study of the Effect of Leachate Recirculation Ratios on the Carboxylic Acid Productions during a Two-Phase Food Waste Anaerobic Digestion. *Chem. Eng. J.* **2023**, *453*, 139800. [[CrossRef](#)]
13. Paranjpe, A.; Saxena, S.; Jain, P. Biogas Yield Using Single and Two Stage Anaerobic Digestion: An Experimental Approach. *Energy Sustain. Dev.* **2023**, *74*, 6–19. [[CrossRef](#)]
14. Khanh Nguyen, V.; Kumar Chaudhary, D.; Hari Dahal, R.; Hoang Trinh, N.; Kim, J.; Chang, S.W.; Hong, Y.; Duc La, D.; Nguyen, X.C.; Hao Ngo, H.; et al. Review on Pretreatment Techniques to Improve Anaerobic Digestion of Sewage Sludge. *Fuel* **2021**, *285*, 119105. [[CrossRef](#)]
15. Jain, S.; Jain, S.; Wolf, I.T.; Lee, J.; Tong, Y.W. A Comprehensive Review on Operating Parameters and Different Pretreatment Methodologies for Anaerobic Digestion of Municipal Solid Waste. *Renew. Sustain. Energy Rev.* **2015**, *52*, 142–154. [[CrossRef](#)]
16. Garlicka, A.; Zubrowska-Sudol, M.; Umiejewska, K.; Roubinek, O.; Palige, J.; Chmielewski, A. Effects of Thickened Excess Sludge Pre-Treatment Using Hydrodynamic Cavitation for Anaerobic Digestion. *Energies* **2020**, *13*, 2483. [[CrossRef](#)]
17. Ferrentino, R.; Langone, M.; Fiori, L.; Andreottola, G. Full-Scale Sewage Sludge Reduction Technologies: A Review with a Focus on Energy Consumption. *Water* **2023**, *15*, 615. [[CrossRef](#)]
18. Moestedt, J.; Rönnberg, J.; Nordell, E. The Effect of Different Mesophilic Temperatures during Anaerobic Digestion of Sludge on the Overall Performance of a WWTP in Sweden. *Water Sci. Technol.* **2017**, *76*, 3213–3219. [[CrossRef](#)] [[PubMed](#)]
19. Appels, L.; Baeyens, J.; Degève, J.; Dewil, R. Principles and Potential of the Anaerobic Digestion of Waste-Activated Sludge. *Prog. Energy Combust. Sci.* **2008**, *34*, 755–781. [[CrossRef](#)]
20. Van, D.P.; Fujiwara, T.; Leu Tho, B.; Song Toan, P.P.; Hoang Minh, G. A Review of Anaerobic Digestion Systems for Biodegradable Waste: Configurations, Operating Parameters, and Current Trends. *Environ. Eng. Res.* **2019**, *25*, 1–17. [[CrossRef](#)]
21. Kumar, G.; Sivagurunathan, P.; Park, J.-H.; Kim, S.-H. Anaerobic Digestion of Food Waste to Methane at Various Organic Loading Rates (OLRs) and Hydraulic Retention Times (HRTs): Thermophilic vs. Mesophilic Regimes. *Environ. Eng. Res.* **2016**, *21*, 69–73. [[CrossRef](#)]
22. Rossi, E.; Pecorini, I.; Iannelli, R. Multilinear Regression Model for Biogas Production Prediction from Dry Anaerobic Digestion of OFMSW. *Sustainability* **2022**, *14*, 4393. [[CrossRef](#)]
23. Xu, F.; Wang, Z.-W.; Li, Y. Predicting the Methane Yield of Lignocellulosic Biomass in Mesophilic Solid-State Anaerobic Digestion Based on Feedstock Characteristics and Process Parameters. *Bioresour. Technol.* **2014**, *173*, 168–176. [[CrossRef](#)]
24. Perendeci, N.A.; Yilmaz, V.; Ertit Taştan, B.; Gökgöl, S.; Fardinpoor, M.; Namli, A.; Steyer, J.P. Correlations between Biochemical Composition and Biogas Production during Anaerobic Digestion of Microalgae and Cyanobacteria Isolated from Different Sources of Turkey. *Bioresour. Technol.* **2019**, *281*, 209–216. [[CrossRef](#)]
25. Park, Y.; Khim, J.; Kim, J.D. Application of a Full-Scale Horizontal Anaerobic Digester for the Co-Digestion of Pig Manure, Food Waste, Excretion, and Thickened Sewage Sludge. *Processes* **2023**, *11*, 1294. [[CrossRef](#)]
26. Kariri, E.; Louati, H.; Louati, A.; Masmoudi, F. Exploring the Advancements and Future Research Directions of Artificial Neural Networks: A Text Mining Approach. *Appl. Sci.* **2023**, *13*, 3186. [[CrossRef](#)]

27. Bahmer, A.; Gupta, D.; Effenberger, F. Modern Artificial Neural Networks: Is Evolution Cleverer? *Neural Comput.* **2023**, *35*, 763–806. [[CrossRef](#)]
28. Chen, J.W.; Chan, Y.J.; Arumugasamy, S.K.; Yazdi, S.K. Process Modelling and Optimisation of Methane Yield from Palm Oil Mill Effluent Using Response Surface Methodology and Artificial Neural Network. *J. Water Process Eng.* **2023**, *52*, 103493. [[CrossRef](#)]
29. Kumar Pandey, A.; Park, J.; Muhorakeye, A.; Morya, R.; Kim, S.-H. Predicting the Impact of Hydraulic Retention Time and Biodegradability on the Performance of Sludge Acidogenesis Using an Artificial Neural Network. *Bioresour. Technol.* **2023**, *372*, 128629. [[CrossRef](#)] [[PubMed](#)]
30. Chen, W.-Y.; Chan, Y.J.; Lim, J.W.; Liew, C.S.; Mohamad, M.; Ho, C.-D.; Usman, A.; Lisak, G.; Hara, H.; Tan, W.-N. Artificial Neural Network (ANN) Modelling for Biogas Production in Pre-Commercialized Integrated Anaerobic-Aerobic Bioreactors (IAAB). *Water* **2022**, *14*, 1410. [[CrossRef](#)]
31. Mafalda, L.; Fernandes, S. *Modeling Anaerobic Digestion with Artificial Neural Networks*; Technical Report; Instituto Superior Tecnico, Universidade de Lisboa: Lisboa, Portugal, 2014.
32. Wu, Y.; Zhang, G.; Zhang, Y.; Wang, B. Design of Biogas Production Prediction Model Based on Improved BP Neural Network and Genetic Algorithm. *J. Phys. Conf. Ser.* **2022**, *2401*, 012076. [[CrossRef](#)]
33. Sakiewicz, P.; Piotrowski, K.; Ober, J.; Karwot, J. Innovative Artificial Neural Network Approach for Integrated Biogas—Wastewater Treatment System Modelling: Effect of Plant Operating Parameters on Process Intensification. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109784. [[CrossRef](#)]
34. Zohdi, T.I. A Machine-Learning Framework for Rapid Adaptive Digital-Twin Based Fire-Propagation Simulation in Complex Environments. *Comput. Methods Appl. Mech. Eng.* **2020**, *363*, 112907. [[CrossRef](#)]
35. Zohdi, T.I. A Digital Twin Framework for Machine Learning Optimization of Aerial Fire Fighting and Pilot Safety. *Comput. Methods Appl. Mech. Eng.* **2021**, *373*, 113446. [[CrossRef](#)]
36. Ying, J.; Liu, J.; Chen, J.; Cao, S.; Hou, M.; Chen, Y. Multi-Scale Fusion Network: A New Deep Learning Structure for Elliptic Interface Problems. *Appl. Math. Model.* **2023**, *114*, 252–269. [[CrossRef](#)]
37. Chen, Y.; Li, Y.; Wu, M.; Lu, F.; Hou, M.; Yin, Y. Differentiating Crohn’s Disease from Intestinal Tuberculosis Using a Fusion Correlation Neural Network. *Knowl.-Based Syst.* **2022**, *244*, 108570. [[CrossRef](#)]
38. Decoster, J.; Claypool, H.M. *Data Analysis in SPSS*; Department of Psychology University of Alabama: Tuscaloosa, Alabama, 2004; pp. 1–59.
39. Ezezi Isaac, O.; Eric Chikweru, A. Test for Significance of Pearson’s Correlation Coefficient (R). *Int. J. Innov. Math. Stat. Energy Policies* **2018**, *1*, 11–23.
40. Gogtay, N.J.; Thatte, U.M. Principles of Correlation Analysis. *J. Assoc. Physicians India* **2017**, *65*, 78–81. [[PubMed](#)]
41. Breiman, L.; Friedman, J.H. Predicting Multivariate Responses in Multiple Linear Regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1997**, *59*, 3–54. [[CrossRef](#)]
42. Parker, S. *Principles and Practice*; OTexts: Melbourne, Australia, 2006; Volume 32, ISBN 9780987507112.
43. Lee, D.J.; Yoon, Y.M.; Choi, I.W.; Bae, J.S.; Seo, D.C. Effect of Seasonal Variations of Organic Loading Rate and Acid Phase on Methane Yield of Food Waste Leachate in South Korea. *Appl. Biol. Chem.* **2017**, *60*, 87–93. [[CrossRef](#)]
44. Sabbir, A.S.M.Y.B.; Saha, C.K.; Nandi, R.; Zaman, M.F.U.; Alam, M.M.; Sarker, S. Effects of Seasonal Temperature Variation on Slurry Temperature and Biogas Composition of a Commercial Fixed-Dome Anaerobic Digester Used in Bangladesh. *Sustainability* **2021**, *13*, 11096. [[CrossRef](#)]
45. Muhayodin, F.; Fritze, A.; Rotter, V.S. A Review on the Fate of Nutrients and Enhancement of Energy Recovery from Rice Straw through Anaerobic Digestion. *Appl. Sci.* **2020**, *10*, 2047. [[CrossRef](#)]
46. Illmer, P.; Gstraunthaler, G. Effect of Seasonal Changes in Quantities of Biowaste on Full Scale Anaerobic Digester Performance. *Waste Manag.* **2009**, *29*, 162–167. [[CrossRef](#)]
47. Sánchez, E.; Borja, R.; Weiland, P.; Travieso, L.; Martín, A. Effect of Substrate Concentration and Temperature on the Anaerobic Digestion of Piggery Waste in a Tropical Climate. *Process Biochem.* **2001**, *37*, 483–489. [[CrossRef](#)]
48. Chae, K.J.; Jang, A.; Yim, S.K.; Kim, I.S. The Effects of Digestion Temperature and Temperature Shock on the Biogas Yields from the Mesophilic Anaerobic Digestion of Swine Manure. *Bioresour. Technol.* **2008**, *99*, 1–6. [[CrossRef](#)]
49. Lin, C.Y.; Noike, T.; Sato, K.; Matsumoto, J. Temperature Characteristics of the Methanogenesis Process in Anaerobic Digestion. *Water Sci. Technol.* **1987**, *19*, 299–300. [[CrossRef](#)]
50. Salais, R. Quantification and Objectivity. From Statistical Conventions to Social Conventions. *Hist. Soc. Res.* **2016**, *41*, 118–134. [[CrossRef](#)]
51. Selvanathan, M.; Jayabalan, N.; Kaur Saini, G.; Supramaniam, M.; Hussin, N. Employee Productivity in Malaysian Private Higher Educational Institutions-Palarch’s. *PalArch’s J. Archaeol. Egypt/Egyptol.* **2020**, *17*, 66–79.
52. Zhang, S.; Ma, X.; Sun, H.; Zhao, P.; Wang, Q.; Wu, C.; Gao, M. Response of Semi-Continuous Anaerobic Digestion of Food Waste to Progressively Increasing Temperature: Methanogen Community, Correlation Analysis, and Energy Balance. *Ind. Crops Prod.* **2023**, *192*, 116066. [[CrossRef](#)]
53. Gholamy, A.; Kreinovich, V.; Kosheleva, O. *Why 70/30 or 80/20 Relation between Training and Testing Sets: A Pedagogical Explanation*; Technical Report; The University of Texas at El Paso: El Paso, Texas, 2018; pp. 1–6.
54. Walczak, S.; Cerpa, N. Artificial Neural Networks. In *Encyclopedia of Physical Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 631–645.

55. Yukesh Kannah, R.; Bhava Rohini, K.; Gunasekaran, M.; Gokulakrishnan, K.; Kumar, G.; Rajesh Banu, J. Prediction of Effective Substrate Concentration and Its Impact on Biogas Production Using Artificial Neural Networks in Hybrid Upflow Anaerobic Sludge Blanket Reactor for Treating Landfill Leachate. *Fuel* **2022**, *313*, 122697. [[CrossRef](#)]
56. Emmanuel Alepu, O.; Li, Z. Effect of Hydraulic Retention Time on Anaerobic Digestion of Xiao Jiahe Municipal Sludge. *Int. J. Waste Resour.* **2016**, *6*, 231. [[CrossRef](#)]
57. Parajuli, A.; Khadka, A.; Sapkota, L.; Ghimire, A. Effect of Hydraulic Retention Time and Organic-Loading Rate on Two-Staged, Semi-Continuous Mesophilic Anaerobic Digestion of Food Waste during Start-Up. *Fermentation* **2022**, *8*, 620. [[CrossRef](#)]
58. Mao, C.; Feng, Y.; Wang, X.; Ren, G. Review on Research Achievements of Biogas from Anaerobic Digestion. *Renew. Sustain. Energy Rev.* **2015**, *45*, 540–555. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.