

Article

Application of Various Machine Learning Models for Process Stability of Bio-Electrochemical Anaerobic Digestion

Ain Cheon ¹, Jwakyung Sung ², Hangbae Jun ¹, Heewon Jang ³, Minji Kim ¹ and Jungyu Park ^{3,*}

¹ Department of Environmental Engineering, Chungbuk National University, Cheongju 28644, Korea; ain722@naver.com (A.C.); jhbcbe@cbnu.ac.kr (H.J.); qhfhfwkd12@naver.com (M.K.)

² Department of Crop Science, Chungbuk National University, Cheongju 28644, Korea; jksung73@chungbuk.ac.kr

³ Department of Advanced Energy Engineering, Chosun University, Gwangju 61457, Korea; hwjang@chosun.ac.kr

* Correspondence: jp@chosun.ac.kr; Tel.: +82-62-230-7119; Fax: +82-62-210-7111

Abstract: The application of a machine learning (ML) model to bio-electrochemical anaerobic digestion (BEAD) is a future-oriented approach for improving process stability by predicting performances that have nonlinear relationships with various operational parameters. Five ML models, which included tree-, regression-, and neural network-based algorithms, were applied to predict the methane yield in BEAD reactor. The results showed that various 1-step ahead ML models, which utilized prior data of BEAD performances, could enhance prediction accuracy. In addition, 1-step ahead with retraining algorithm could improve prediction accuracy by 37.3% compared with the conventional multi-step ahead algorithm. The improvement was particularly noteworthy in tree- and regression-based ML models. Moreover, 1-step ahead with retraining algorithm showed high potential of achieving efficient prediction using pH as a single input data, which is plausibly an easier monitoring parameter compared with the other parameters required in bioprocess models.



Citation: Cheon, A.; Sung, J.; Jun, H.; Jang, H.; Kim, M.; Park, J.

Application of Various Machine Learning Models for Process Stability of Bio-Electrochemical Anaerobic Digestion. *Processes* **2022**, *10*, 158. <https://doi.org/10.3390/pr10010158>

Academic Editor: Shaoan Cheng

Received: 30 August 2021

Accepted: 11 January 2022

Published: 14 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

Keywords: machine learning; bio-electrochemical anaerobic digestion; methane yield; pH; process stability

1. Introduction

Anaerobic digestion (AD) is gaining attention as a promising technology for biogas production from various organic wastes, such as food waste, waste activated sludge, livestock manure, and landfill leachate [1]. However, AD performances are often affected by substrate characteristics, organic loading rate (OLR), accumulated volatile fatty acids (VFAs) concentration, pH, alkalinity, ammonia concentration, and toxic substances [2,3]. Therefore, AD reactors occasionally exhibit unstable methane production and inefficient organic degradation rate [4,5]. In particular, highly concentrated and easily biodegradable organic matter, such as food waste, interrupts efficient methane production and fast stabilization by accelerating VFA accumulation and pH decrease, resulting in an imbalance between acidogenesis and methanogenesis [6].

Bio-electrochemical anaerobic digestion (BEAD) is gaining attention as an advanced technology that improves microbial activity and growth rates as well as organic removal efficiency and biogas productivity by supplying low voltage (0.2~1.0 V) through bio-electrodes in an AD reactor [7,8]. BEAD systems are superior to AD systems with respect to organic substances removal and biogas production, and that a decrease in pH and VFA accumulation has a low inhibitory effect on methane production [9–11]. Previous lab-scale studies have sufficiently demonstrated the superiority of BEAD through basic studies such as reaction mechanism identification, changes in microbial community structure, electrode configuration, and material suitability [12–14].

Operational stability should be examined as the next step to enhance the applicability of BEAD because operational stability and maintainability of BEAD are important parameters for its application in full-scale processes. This can be achieved by predicting the performance based on the long-term performance of BEAD. In BEAD processes, the analytical parameters are nonlinear in nature [15]. Various methods for forecasting process performance have been researched to improve operational stability by analyzing nonlinear patterns.

Machine learning (ML), a statistical forecasting method, is gaining significant attention for forecasting performance and preventing operational risks. ML can be successfully applied into process models because of its capability to interpret the nonlinear relationships that might be produced among variables (multi input/output) in a complex system [16]. Compared with the AD models, ML can model and predict complex relationships between dependent and independent variables associated with the AD process, without requiring detailed mechanisms of anaerobic processes [17]. In addition, ML models contain a class of generic nonlinear regression models that learn the arbitrary mapping of the input data to the output data to obtain computational models with high predictive accuracy [18]. Hence, an extensive understanding in process model is not required in ML modeling [19]. This suggests that ML can support the long-term process stability of BEAD by applying some operational parameters as input data. BEAD is proven techniques that could achieve a higher process stability than that of conventional AD, supporting bio-electrochemically active microorganism and preventing various inhibitions that cause failure of a reaching steady state [8,20]. Based on these advantages of BEAD, various ML models could be applied into long-term operation of BEAD process for supporting operational stability and accelerating biogas production. However, in-depth study results supporting the long-term process stability of BEAD have not yet been reported, highlighting the need for studying ML applications of BEAD.

Conventional ML models focused on raw data collected during specific operational periods for prediction of future performances by using simultaneous prediction method [21]. Although that method was widely applied to continuously operated bio-process, simultaneous prediction has a limitation in applying new input data that is continuously accumulated. A 1-step ahead algorithm showed a possibility of continual training which contributes an achieving a higher adoption to bio-process. Previous study clearly showed that the 1-step ahead with retraining algorithm was suitable for the practical application by predicting performances derived from continuously operated bio-process [21].

Therefore, a practical application of ML to BEAD for treating food waste was suggested in this study using a long-term evaluation of the effects of operational parameters on BEAD reactors. Various ML models with multi-step and 1-step ahead algorithms were applied to forecast the performance and achieve high operational stability of the BEAD. Moreover, pH was applied as a single input data to evaluate the possibility of real-time prediction and practical applications. The 1-step ahead method, which utilized prior data of BEAD performances, could enhance the prediction accuracy. In addition, 1-step ahead with the retraining algorithm could achieve high prediction accuracy when pH was used as a single input parameter.

2. Materials and Methods

2.1. Data Preprocessing

The data used in this study were collected from a lab-scaled single-chamber BEAD reactor (effective volume: 20 L) treating food waste. The BEAD reactor was operated for 1086 days under various organic loading rates (OLRs) based on the input chemical oxygen demand (COD) concentration. The details of the BEAD reactor have been published in previous studies [22,23]. The pH, alkalinity, and COD removal efficiency were used as the input parameters, and the input COD based methane yield ($L\text{-CH}_4/g\text{-COD}$) was used as the output parameter. The input parameters were chosen in accordance with the variable importance analysis results. When pH, alkalinity, and COD removal efficiency were applied as independent variable, the highest R^2 value was calculated. The lab-scaled

BEAD reactor was operated by supplying voltage of 0.3 V under gradually increased OLRs (Table 1). During stage 1, the BEAD reached intermediate and final steady states after 98 and 250 days, respectively, of operation and continued stable methane production by stage 5. Stable methane yields in the BEAD reactor at the final steady state of S1–S5 (2.0–10.0 kg/m³·d) were 0.35 ± 0.02, 0.36 ± 0.04, 0.36 ± 0.04, 0.36 ± 0.02, and 0.36 ± 0.02 L-CH₄/g-COD, respectively. More details on BEAD performance are presented in Table 1.

Table 1. Methane production and yield in BEAD reactor during the total operation periods.

Item	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Operation period (days)	0–365	366–598	599–795	796–950	951–1086
OLR (kg-COD/m ³ ·d)	2.5 ± 0.6	1.0 ± 0.2	6.0 ± 0.3	8.0 ± 0.3	10.0 ± 0.4
pH	7.7 ± 0.3	8.0 ± 0.2	8.1 ± 0.1	8.1 ± 0.1	8.2 ± 0.1
Alkalinity (g/L as CaCO ₃)	7.6 ± 0.9	10.1 ± 0.8	13.9 ± 0.8	14.8 ± 0.7	15.3 ± 0.7
Total VFAs (mg/L)	2.6 ± 0.9	3.1 ± 0.2	3.9 ± 0.2	4.6 ± 0.3	5.3 ± 0.3
COD removal efficiency (%)	67.8 ± 7.2	71.4 ± 2.5	73.5 ± 3.0	75.1 ± 2.3	76.3 ± 1.7
CH ₄ production (L/day)	15.7 ± 4.6	33.9 ± 3.9	51.2 ± 6.3	63.4 ± 3.9	74.7 ± 3.4
CH ₄ yield (L-CH ₄ /g-COD)	0.32 ± 0.07	0.35 ± 0.04	0.35 ± 0.04	0.36 ± 0.02	0.36 ± 0.01

BEAD: bio-electrochemical anaerobic digestion, OLR: organic loading rate, VFA: volatile fatty acid, COD: chemical oxygen demand.

2.2. Statistical Analysis

2.2.1. Principal Component Analysis (PCA)

The PCA analysis was conducted using pH, alkalinity, COD removal efficiency, and methane yield of the BEAD reactor as principal components. The axes of principal components presenting eigenvalues of 1.0 that showed the dispersion size of orthogonal data were considered when the number of principal components was determined [24]. The varimax rotation method that can explain the relationships between variables and components was used to rotate the axis [21]. The Bartlett's sphericity test and the Kaiser Meyer Olkin (KMO) test were applied to determine validity of preprocessed data for the PCA. The KMO test results reveal the degree of covariance between the variables used in the analysis and the components inherent in the data. As the degree of covariance approaches 1, the validity of the analysis is high, and the analysis can be performed only when it is at least 0.5 [25]. Statistical analysis was performed using four variables that satisfied the standard value of KMO. The KMO-value and *p*-value of four variables which consist of pH, alkalinity, COD removal efficiency, and methane yield were 0.73 and less than 0.01, respectively.

2.2.2. Variable Importance Analysis

Input data that was properly selected simplifies the model algorithm and improves its applicability to full scale processes. Therefore, Recursive feature elimination (RFE) was used to remove low important variables, one at a time. The lowest RMSE of 0.2382 L-CH₄/g-COD and the highest R² of 0.971 were obtained when the three independent variables (ranked as follows: pH > COD removal efficiency > alkalinity) were applied. Therefore, these three parameters were used as input data in ML models used in this study.

2.3. ML

2.3.1. Prediction Models

The input layer treats all the input data by communicating with the external environment that provides significant pattern [26]. These input data are transferred to the hidden layer, and every input neuron could show independent variables that can affect to the outputs of the neural network (Figure 1a). The hidden layer collects those neurons that include applied activation function. Because hidden layer processes the inputs obtained from previous layer, it is responsible for extracting the required features from the input data [27]. The output layer collects and transmits information according to a designated method.

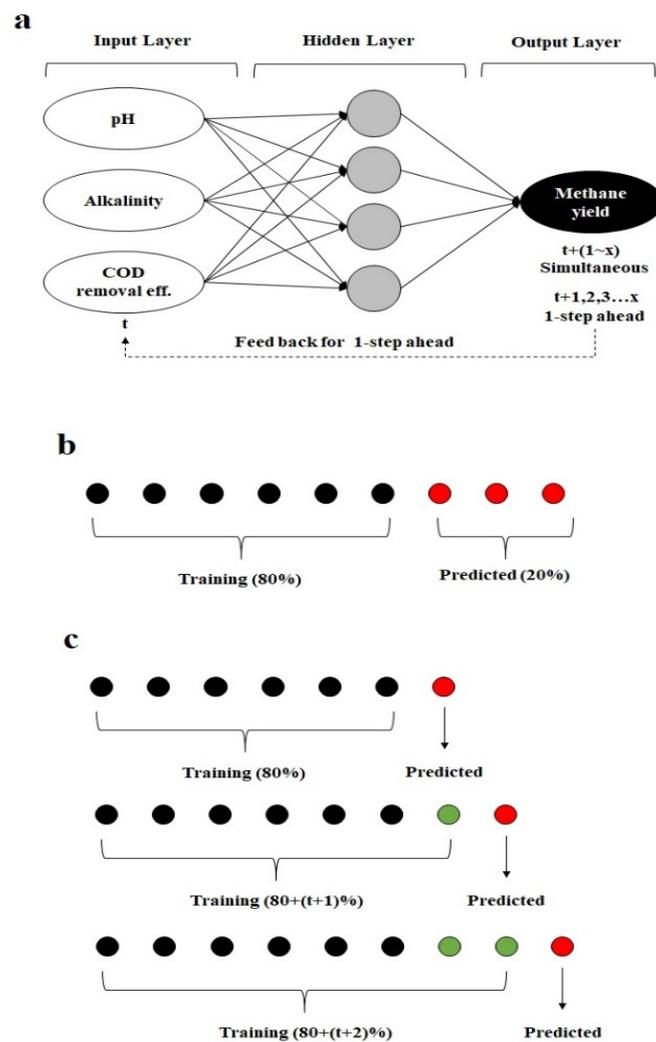


Figure 1. Schematic diagrams for understanding the (a) machine learning algorithm, (b) multi-step ahead method, and (c) 1-step ahead with retraining method [21].

The following five ML models were applied to predict the methane yield of BEAD reactor: random forest (RF), extreme gradient boosting (XGboost), support vector regression (SVR), long short-term memory (LSTM), and recurrent neural network (RNN). A neural network algorithm has three different layers: input, hidden, and output [26,27].

Each ML was modeled by using multi-step ahead method and 1-step ahead with retraining method. More detailed fundamentals of each method are presented in Section 2.3.3 and Section 2.3.4, respectively. This study used R program (version 3.5.1), which is a software for statistical analysis, ML modeling, and graphics formation. The R program packages used for each ML model are listed in Table 2.

Table 2. R program packages used for the prediction of methane yield.

ML Models	Packages
Random Forest (RF)	Package "randomForest"
Extreme gradient boosting (XGboost)	Package "rxgboost"
Support Vector Regression (SVR)	Package "e1071"
Long Short-Term Memory (LSTM)	Package "rnn" and "keras"
Recurrent Neural Networks (RNN)	Package "rnn"

2.3.2. Validations and Model Accuracy Calculation

Determining the optimal model parameters is important for improving the prediction accuracy of ML models [27]. Cross validation was introduced to determine the optimal combinations of hyperparameters. Learning rate, number of hidden nodes, batch size for LSTM and RNN, C and sigma for SVR, and number of trees for the RF and XGboost were considered as hyperparameters to optimize each model [21]. The 10-fold cross-validation was repeated three times to prevent overfitting and evaluate the prediction performance. The data was divided into a training set and a test set, which were used for the model construction and evaluation of prediction accuracy.

Based on the continuously accumulated operation data of the BEAD reactor during operational stages 1–5, 80% of the total time-series data were provided as training data, and the posterior 20% was provided as test data (Figure 1b,c). For predicting final methane yield, pH, alkalinity, and COD removal efficiency were used as input parameters and amount of training and predicting samples were 312ea (80% of operation period) and 78ea (20% of operation period), respectively. To compare the prediction accuracies of each ML model, the RMSEs of all ML model results were evaluated in this study, using Equation (1):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

2.3.3. Multi-Step Ahead Method

The raw data (see Supplementary Materials) obtained by BEAD reactor operation for 3 years was divided into training and test datasets (Figure 1). In this study, 80% of the raw data were used for training, and the remaining 20% of raw data were used for testing. The multi-step ahead method was applied using split-sample experiments [21]. After modeling was finished, the prediction accuracy was evaluated by comparing with predicted values and known data. Therefore, multi-step ahead prediction was performed by using only 80% past data of raw dataset as inputs for training process.

2.3.4. 1-Step Ahead with the Retraining Method

In contrast with multi-step ahead method, 1-step ahead with the retraining method considers the previous learning contents required in the time-series data analysis and updates the inputs sequentially for the retraining process. In the 1-step ahead with the retraining method, the network trained up to past time step n th is retrained to predict the outputs for the next time step, that is, the $(n + 1)$ th step (Figure 1c) [28,29]. Cumulative 1-step ahead retraining and learning were performed as follows: a model using the data at time point t was constructed and the future value at time $t + 1$ was predicted. After adding the data at time $t + 1$, a new model was built to retrain data at $[1, \dots, t + 1]$ to predict the value at time step $t + 2$. After repeating this process and when predicting the value after time N elapsed, the model is constructed using data from the time step $[1, \dots, t + 1, \dots, t + n]$, and the value is predicted at time step $t + n + 1$ (Figure 2b) [27]. In this study, when time $t - 3$ was included, the prediction accuracy was the highest. Therefore, input parameters and predicted outputs at $t - 3$ step was applied for retraining process of each 1-step ahead ML model.

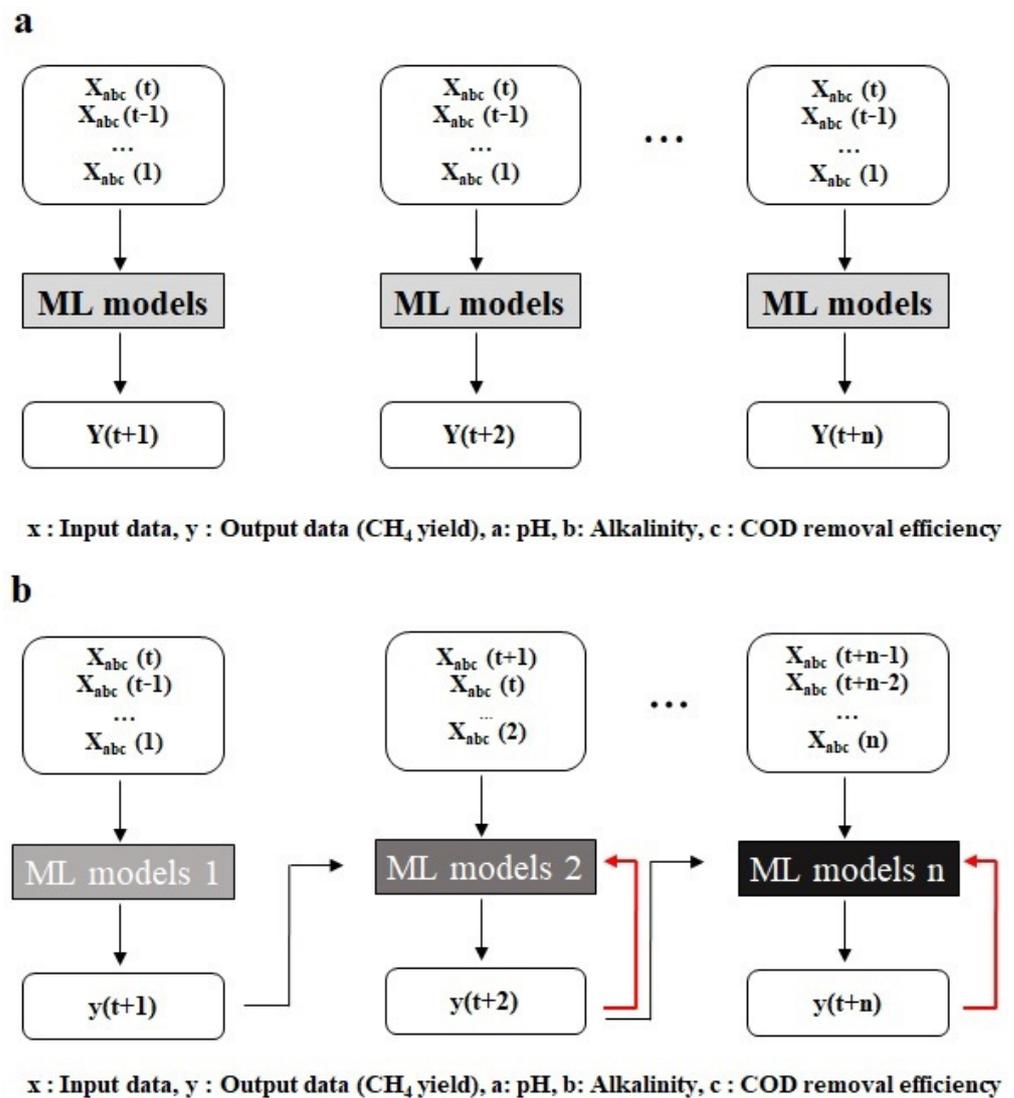


Figure 2. Fundamentals of the (a) multi-step ahead and (b) 1-step ahead with retraining methods [21].

3. Results

3.1. Statistical Analysis

Figure 3a shows the results of PCA analysis when methane yield, pH, alkalinity, and COD removal efficiency of the BEAD reactor were applied as variables. The methane yield of the BEAD reactors shows positive correlations with the pH, alkalinity, and COD removal efficiency. The decreases in pH, alkalinity, and COD removal efficiency affected the decrease in the final methane yield [30]. In particular, pH had the highest correlation with the methane yield (BEAD reactor: 0.80), suggesting that rapidly overcoming the inhibition caused by a pH decrease could contribute to stable methane production (Figure 3b) [31]. The methane yield of the BEAD reactor showed no correlation with the VFAs, which did not satisfy the baseline value of KMO.

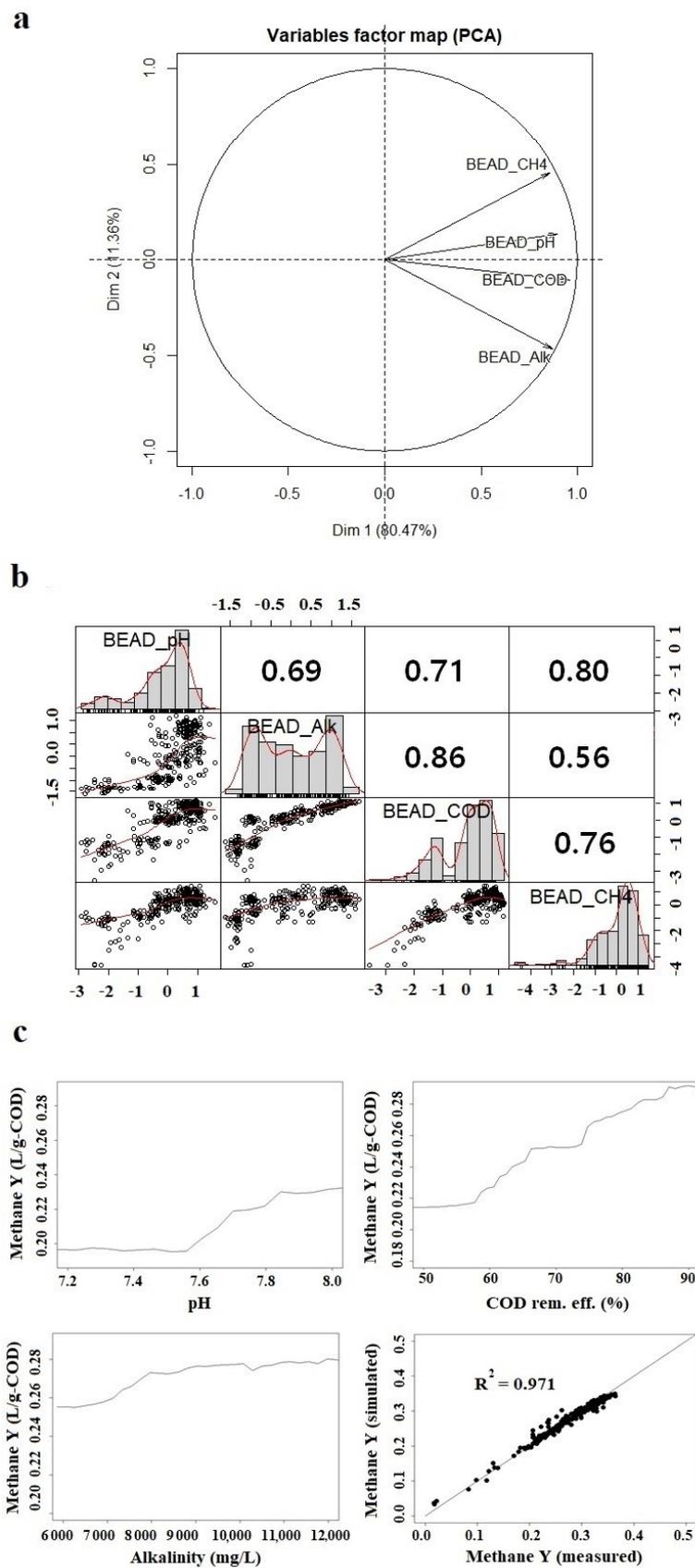


Figure 3. (a) Principal component analysis (PCA) loading plots and (b) scatter plots matrix for pairwise correlations, and (c) partial dependents correlations between independent variables (pH, alkalinity, and COD removal efficiency) and subordination variable (methane yield) of bio-electrochemical anaerobic digestion (BEAD).

When the three independent variables were used, the BEAD reactor showed the lowest RMSE values in the RFE-RE model. The variables were in the order of pH > COD removal efficiency > alkalinity for the BEAD reactor. The R^2 and RMSE of the BEAD reactor were 0.971 and 0.2382 L-CH₄/g-COD, respectively. This explained why the methane production of the BEAD reactor was affected by COD and H⁺ consumption rates [22,32,33]. As shown in the partial correlations, the correlations were nonlinear and complex. The methane yield of BEAD decreased when the pH, alkalinity, and COD removal efficiencies were lower than 7.6, 8000 mg/L as CaCO₃, and 60%, respectively. The results of partial dependents correlations (Figure 3c) clearly showed that the methane yield of BEAD had non-linear relationships with pH, alkalinity, and COD removal efficiency, respectively, and clarified needs of enhanced prediction models for achieving high process stability in the BEAD operation.

3.2. Multi-Step Ahead ML Models

The RMSE value of the prediction result using the RNN method was 0.025 L-CH₄/g-COD, showing the best prediction efficiency (Figure 4). In addition, the RMSEs of RF, XGboost, LSTM, and SVR were 0.041, 0.053, 0.055, and 0.056 L-CH₄/g-COD, respectively. For the BEAD reactor, the prediction accuracy of the RNN method, which was effective for time-series prediction, was the highest. Therefore, RNN could reflect the characteristics of daily data appropriately, thereby showing a high prediction accuracy [34]. In cases of the BEAD reactor using the decision tree-based RF and XGboost, the prediction result was overestimated for the instantaneous methane yield decrease at the initial operation in each stage. This implied that the prediction accuracy was low for data that deviated significantly from the mean value of the regression calculated through learning [35]. Furthermore, the prediction efficiency of regression-based SVR, which assumed a linear combination of variables, was low in biological reactions with complex nonlinear relationships of various factors. For efficient operation and management of real BEAD reactors, it would be more effective to use the RNN method based on the accumulated time-series data when predicting the methane yield of the BEAD reactors with nonlinear relationships with time [36].

3.3. 1-Step Ahead ML Models

Reportedly, 1-step ahead prediction methods can predict and analyze time-series data with high accuracy and prediction efficiency [37,38]. Figure 5 shows the results of the 1-step ahead prediction using various ML models. In case of the BEAD reactor, the RMSE value of the prediction result using the RNN method was 0.017 L-CH₄/g-COD, showing the best prediction efficiency. The RMSEs of SVR, LSTM, RF, and XGboost were 0.021, 0.022, 0.028, and 0.030 L-CH₄/g-COD, respectively. In every ML models, The 1-step ahead with retraining method showed a higher RMSEs than the RMSEs of the multi-step ahead method shown earlier. This indicated that the 1-step ahead method which continuously retrains previous prediction values could more efficiently predict the methane yield of the BEAD reactor based on data that have nonlinear relationships with time [39]. In other words, because operation data are accumulated continuously in BEAD reactor that is operated continuously, the 1-step ahead method that facilitates learning by applying them in stages can be effectively applied [40]. In particular, the prediction accuracies of RF, XGboost, and SVR, which were not appropriate for time-series prediction, were increased through the 1-step ahead method, and they were not significantly different from the RMSE value of the RNN method. These results suggest that the prediction can be performed indirectly for the time-series data analysis using 1-step ahead method. Of note, the prediction value that deviates greatly from the regression section in the multi-step ahead prediction of decision tree-based RF and XGboost can be corrected based on the time-series learning and prediction through the 1-step ahead method. Therefore, the usability of the decision tree-based model can be increased in the prediction of nonlinear data over time [41,42].

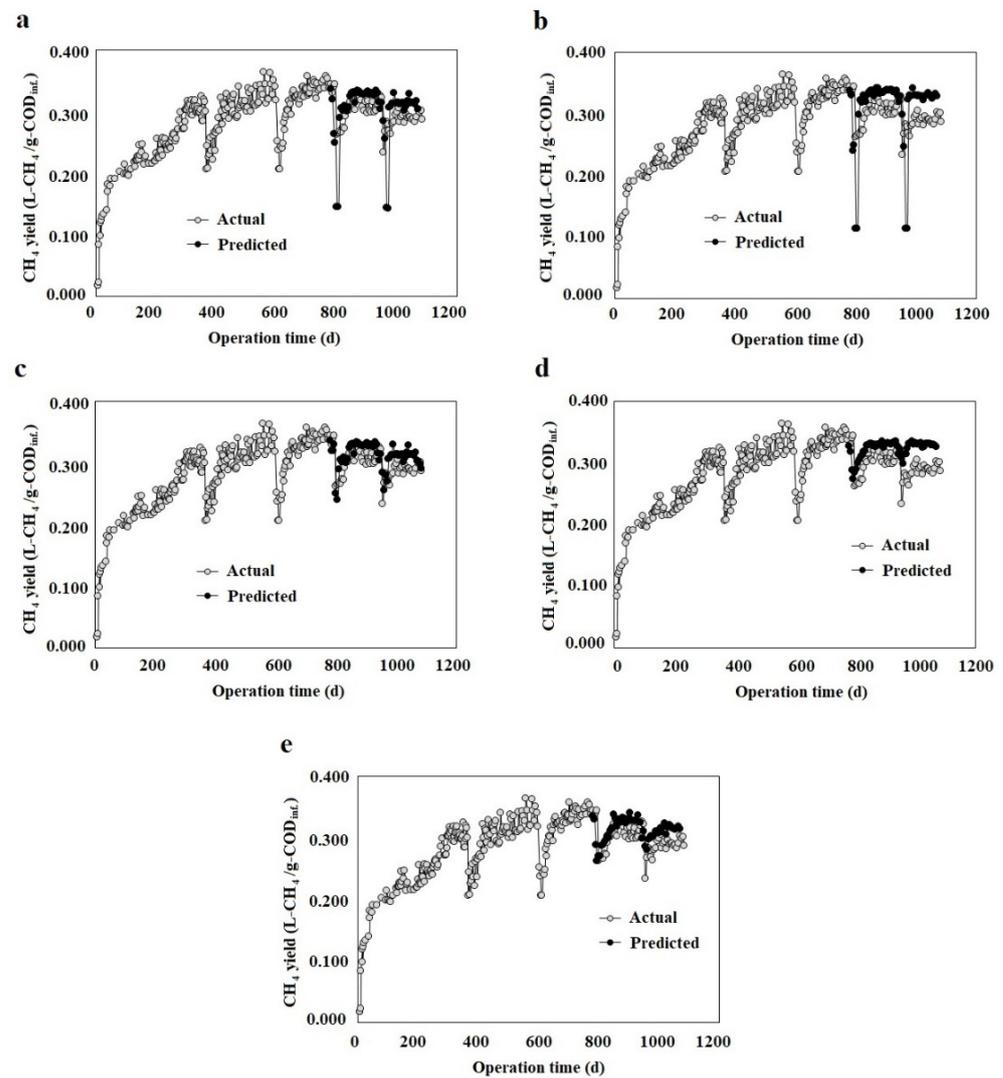


Figure 4. Results of the multi-step ahead predictions prediction of bio-electrochemical anaerobic digestion (BEAD) reactor using (a) random forest (RF), (b) extreme gradient boosting (XGboost), (c) support vector regression (SVR), (d) long short-term memory (LSTM), and (e) recurrent neural network (RNN) models.

3.4. Prediction of Methane Yield Using pH as Single Input Data

The 1-step ahead model using pH, alkalinity, and COD removal efficiency as input data was found to enable the effective prediction of time-series data. However, these input data are not available for real-time prediction because of the difficulty of prompt measurement in the full-scale BEAD process [43]. pH is the easiest parameter for monitoring full-scale BEAD processes using portable instruments and is one of the most important factors that directly affects methanogenic microorganism activity [44–46]. Therefore, the effect of pH as a single input data point on the prediction of methane yield was evaluated in this study. For the BEAD reactor, the prediction efficiency of the RNN method, which was effective for time-series prediction, was the highest. The 1-step ahead method of every ML model showed a higher prediction accuracy than the multi-step ahead prediction efficiency shown earlier. This indicated that the 1-step ahead method that facilitates learning by considering previous prediction values continuously could more efficiently predict the methane yield in the full-scale BEAD process based on the pH as a single input data [39]. Figure 6 shows the RMSE values of BEAD resulting from the multi-step and 1-step ahead RNN models that achieved the highest prediction efficiency. For the multi-step-ahead RNN model, the RMSE value of the BEAD reactor was 0.032 L-CH₄/g-COD. For the 1-step ahead RNN model, the

RMSE value of the BEAD reactor was 0.017 L-CH₄/g-COD. These results show that the methane yield could be effectively predicted by pH as a single input data and suggest the possibility of applying BEAD to a full-scale process [46].

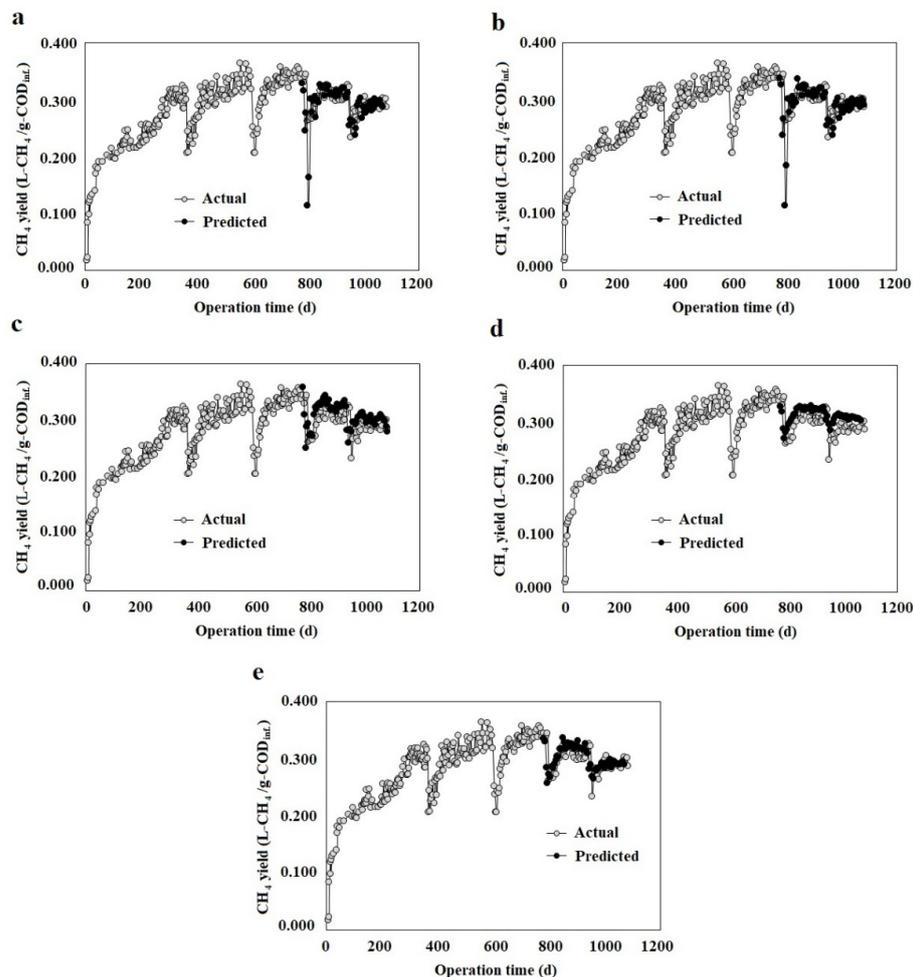


Figure 5. Results of 1-step ahead prediction of bio-electrochemical anaerobic digestion (BEAD) reactor using (a) random forest (RF), (b) extreme gradient boosting (XGboost), (c) support vector regression (SVR), (d) long short-term memory (LSTM), and (e) recurrent neural network (RNN).

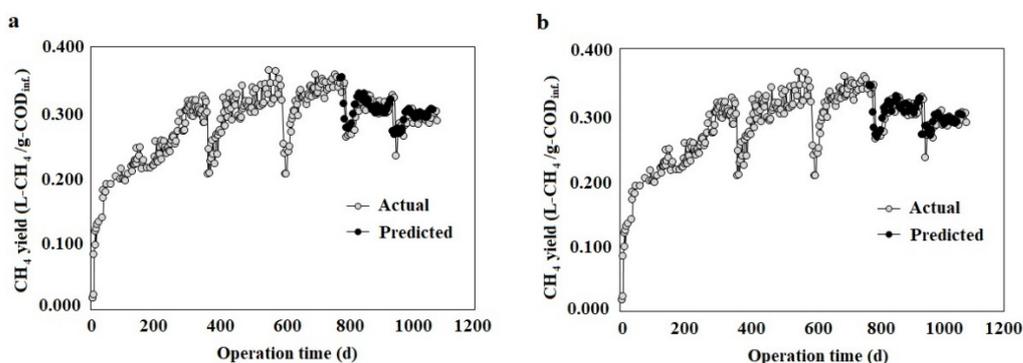


Figure 6. Results of multi-step ahead (a) and 1-step ahead (b) prediction of bio-electrochemical anaerobic digestion (BEAD) reactor using recurrent neural network (RNN) model with pH as a single input data.

4. Discussion

Results from the PCA showed that pH had the highest correlation with the methane yield in the BEAD reactor, which meant that quickly overcoming the inhibition caused

by a pH decrease could contribute to stable methane production. 1-step ahead prediction method could predict and analyze time-series data with high accuracy and prediction efficiency (Table 3). In other words, because operational performance data are continuously accumulated in the BEAD reactor, the 1-step ahead with retraining method that facilitates learning by applying them in stages can be effectively applied. The capability of the 1-step ahead with the retraining method could realize real-time monitoring and prediction of BEAD performance simultaneously. These potentials would be useful to achieve the stable operation of the full-scale BEAD process, especially when BEAD is faced with unexpected status, causing a loss of economic and energy production.

Table 3. RMSE values of BEAD reactor for multi-step ahead and 1-step ahead predictions using various machine learning models.

Parameters		RMSE (L-CH ₄ /g-COD)				
		RF	XGboost	SVR	LSTM	RNN
BEAD	Multi-step ahead	0.041	0.053	0.056	0.055	0.025
	1-step ahead	0.028	0.030	0.021	0.022	0.017

RMSE: root mean square error, RF: random forest, COD: chemical oxygen demand, XGboost: extreme gradient boosting, SVR: support vector regression, LSTM: long short-term memory, RNN: recurrent neural network, BEAD: bio-electrochemical anaerobic digestion.

While alkalinity, COD removal efficiency, VFAs, and others could be also used as input parameters for ML models, they are not suitable for real-time predictions in full-scale BEAD processes due to time-consuming disadvantages and uneconomic applicability [21,43]. However, pH can be quickly analyzed by sensor-based portable detectors. Furthermore, pH is the most sensitive factor that directly affects methanogenic microorganism activity and methane yield [44–47], and change of pH showed the highest relationship with BEAD performance in the statistic analysis of this study (Figure 3). Thus, the result of prediction using pH as a single input data showed that the methane yield could be effectively predicted by pH data and implied the possibility of practical application of BEAD, which could maintain optimum pH values via bio-electrochemical reactions (Table 4).

Table 4. RMSE values of BEAD reactor for multi-step ahead and 1-step ahead predictions using various machine learning models with pH as a single input data.

Parameters		RMSE (L-CH ₄ /g-COD)				
		RF	XGboost	SVR	LSTM	RNN
BEAD	Multi-step ahead	0.020	0.023	0.022	0.021	0.019
	1-step ahead	0.019	0.022	0.019	0.019	0.017

RMSE: root mean square error, RF: random forest, COD: chemical oxygen demand, XGboost: extreme gradient boosting, SVR: support vector regression, LSTM: long short-term memory, RNN: recurrent neural network, BEAD: bio-electrochemical anaerobic digestion.

This study could show that the various ML models would be able to help BEAD achieves a higher process stability than AD. Moreover, 1-step ahead with the retraining methods could provide realizable applicability of various ML models to real world bio-processes. The pH could be realizable parameter as a single input data and its applicability was proven in this study. This possibility implies more detailed and scientific algorithm should be developed and modeled in the future.

5. Conclusions

This study confirmed that the 1-step ahead with the retraining method applied to various ML models was able to improve prediction accuracy of BEAD performance by retraining the prior state performances in the time series data. Notably, 1-step ahead with the retraining method significantly improved prediction accuracies during the OLR transition periods in the tree-based RF and regression-based SVR models. Another important finding of 1-step ahead method was that pH as only input parameter could be efficiently

used for real-time prediction of BEAD performance. The ML models using pH as a single input parameter were less accurate than those using multiple input parameters. However, pH was more efficient for monitoring than the other parameters, offering advantages in achieving real-time performance predictions for time-series full-scale operations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pr10010158/s1>, Table S1: Input and output data for various ML models.

Author Contributions: Original draft preparation: A.C.; Conceptualization, data curation, and methodology: H.J. (Hangbae Jun) and J.S.; formal analysis and visualization: M.K. and H.J. (Heewon Jang); Conceptualization, review, and editing: J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R111A3044486) and was supported by the Korea Ministry of Environment as Waste to Energy-Recycling Human Resource Development Project [YL-WE-19-001].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Adekunle, K.F.; Okolie, J.A. A review of biochemical process of anaerobic digestion. *Adv. Biosci. Biotechnol.* **2015**, *6*, 205–212. [[CrossRef](#)]
2. Wang, H.; Zhang, Y.; Angelidaki, I. Ammonia inhibition on hydrogen enriched anaerobic digestion of manure under mesophilic and thermophilic conditions. *Water Res.* **2016**, *105*, 314–319. [[CrossRef](#)]
3. Rajagopal, R.; Massé, D.I.; Singh, G. A critical review on inhibition of anaerobic digestion process by excess ammonia. *Bioresour. Technol.* **2013**, *143*, 632–641. [[CrossRef](#)] [[PubMed](#)]
4. Moset, V.; Bertolini, E.; Cerisuelo, A.; Cambra, M.; Olmos, A.; Cambra-López, M. Start-up strategies for thermophilic anaerobic digestion of pig manure. *Energy* **2014**, *74*, 389–395. [[CrossRef](#)]
5. Chow, W.; Chong, S.; Lim, J.; Chan, Y.; Chong, M.; Tiong, T.; Chin, J.; Pan, G. Anaerobic co-digestion of wastewater sludge: A review of potential co-substrates and operating factors for improved methane yield. *Processes* **2020**, *8*, 39. [[CrossRef](#)]
6. Kazemi, P.; Steyer, J.; Bengoa, C.; Font, J.; Giral, J. Robust data-driven soft sensors for online monitoring of volatile fatty acids in anaerobic digestion processes. *Processes* **2020**, *8*, 67. [[CrossRef](#)]
7. Liu, W.; Cai, W.; Guo, Z.; Wang, L.; Yang, C.; Varrone, C.; Wang, A. Microbial electrolysis contribution to anaerobic digestion of waste activated sludge, leading to accelerated methane production. *Renew. Energy* **2016**, *91*, 334–339. [[CrossRef](#)]
8. Park, J.; Kwon, H.; Sposob, M.; Jun, H. Effect of a side-stream voltage supplied by sludge recirculation to an anaerobic digestion reactor. *Bioresour. Technol.* **2020**, *300*, 122643. [[CrossRef](#)]
9. Feng, Y.; Zhang, Y.; Chen, S.; Quan, X. Enhanced production of methane from waste activated sludge by the combination of high-solid anaerobic digestion and microbial electrolysis cell with iron-graphite electrode. *Chem. Eng. J.* **2015**, *259*, 787–794. [[CrossRef](#)]
10. An, Z.; Feng, Q.; Zhao, R.; Wang, X. Bioelectrochemical methane production from food waste in anaerobic digestion using a carbon-modified copper foam electrode. *Processes* **2020**, *8*, 416. [[CrossRef](#)]
11. De Vrieze, J.; Gildemyn, S.; Arends, J.B.; Vanwonterghem, I.; Verbeke, K.; Boon, N.; Verstraete, W.; Tyson, G.W.; Hennebel, T.; Rabaey, K. Biomass retention on electrodes rather than electrical current enhances stability in anaerobic digestion. *Water Res.* **2014**, *54*, 211–221. [[CrossRef](#)]
12. Zhang, X.; Li, X.; Zhao, X.; Li, Y. Factors affecting the efficiency of a bioelectrochemical system: A review. *RSC Adv.* **2019**, *9*, 19748–19761. [[CrossRef](#)]
13. Escapa, A.; Mateos, R.; Martínez, E.J.; Blanes, J. Microbial electrolysis cells: An emerging technology for wastewater treatment and energy recovery. From laboratory to pilot plant and beyond. *Renew. Sustain. Energy Rev.* **2016**, *55*, 942–956. [[CrossRef](#)]
14. Beegle, J.R.; Borole, A.P. Energy production from waste: Evaluation of anaerobic digestion and bioelectrochemical systems based on energy efficiency and economic factors. *Renew. Sustain. Energy Rev.* **2018**, *96*, 343–351. [[CrossRef](#)]
15. Nair, V.V.; Dhar, H.; Kumar, S.; Thalla, A.K.; Mukherjee, S.; Wong, J.W. Artificial neural network based modeling to evaluate methane yield from biogas in a laboratory-scale anaerobic bioreactor. *Bioresour. Technol.* **2016**, *217*, 90–99. [[CrossRef](#)]

16. Antwi, P.; Li, J.; Boadi, P.O.; Meng, J.; Shi, E.; Deng, K.; Bondinuba, F.K. Estimation of biogas and methane yields in an UASB treating potato starch processing wastewater with back propagation artificial neural network. *Bioresour. Technol.* **2017**, *228*, 106–115. [[CrossRef](#)]
17. Antwi, P.; Li, J.; Meng, J.; Deng, K.; Koblah Quashie, F.; Li, J.; Opoku Boadi, P. Feedforward neural network model estimating pollutant removal process within mesophilic upflow anaerobic sludge blanket bioreactor treating industrial starch processing wastewater. *Bioresour. Technol.* **2018**, *257*, 102–112. [[CrossRef](#)]
18. Pandey, D.S.; Das, S.; Pan, I.; Leahy, J.J.; Kwapinski, W. Artificial neural network based modelling approach for municipal solid waste gasification in a fluidized bed reactor. *Waste Manag.* **2016**, *58*, 202–213. [[CrossRef](#)]
19. Ismail, S.; Elsamadony, M.; Fujii, M.; Tawfik, A. Evaluation and optimization of anammox baffled reactor (AnBR) by artificial neural network modeling and economic analysis. *Bioresour. Technol.* **2019**, *271*, 500–506. [[CrossRef](#)]
20. Park, J.; Lee, B.; Kwon, H.; Jun, H. Contribution analysis of methane production from food waste in bulk solution and on bio-electrode in a bio-electrochemical anaerobic digestion reactor. *Sci. Total Environ.* **2019**, *670*, 741–751. [[CrossRef](#)]
21. Park, J.; Jun, H.; Heo, T. Retraining prior state performances of anaerobic digestion improves prediction accuracy of methane yield in various machine learning models. *Appl. Energy* **2021**, *298*, 117250. [[CrossRef](#)]
22. Park, J.; Lee, B.; Tian, D.; Jun, H. Bioelectrochemical enhancement of methane production from highly concentrated food waste in a combined anaerobic digester and microbial electrolysis cell. *Bioresour. Technol.* **2018**, *247*, 226–233. [[CrossRef](#)]
23. Park, J.; Lee, B.; Park, H.; Jun, H. Long-term evaluation of methane production in a bio-electrochemical anaerobic digestion reactor according to the organic loading rate. *Bioresour. Technol.* **2019**, *273*, 478–486. [[CrossRef](#)] [[PubMed](#)]
24. Bern, C.; Walton-Day, K.; Naftz, D. Improved enrichment factor calculations through principal component analysis: Examples from soils near breccia pipe uranium mines, Arizona, USA. *Environ. Pollut.* **2019**, *248*, 90–100. [[CrossRef](#)] [[PubMed](#)]
25. Jung, S.Y.; Kim, I.K. Analysis of water quality factor and correlation between water quality and Chl-a in middle and downstream weir section of Nakdong River. *J. Korean Soc. Environ. Eng.* **2017**, *39*, 89–96. [[CrossRef](#)]
26. Suzuki, K. (Ed.) *Artificial Neural Networks: Methodological Advances and Biomedical Applications*; BoD—Books on Demand: Janeza Trdine, Croatia, 2011.
27. Shin, Y.; Kim, T.; Hong, S.; Lee, S.; Lee, E.; Hong, S.; Lee, C.; Kim, T.; Park, M.S.; Park, J.; et al. Prediction of chlorophyll-a concentrations in the Nakdong River using machine learning methods. *Water* **2020**, *12*, 1822. [[CrossRef](#)]
28. Xiao, H.; Huang, D.; Pan, Y.; Liu, Y.; Song, K. Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model. *Chemom. Intell. Lab. Syst.* **2017**, *161*, 96–107. [[CrossRef](#)]
29. Jain, V.K.; Banerjee, A.; Kumar, S.; Kumar, S.; Sambhi, S.S. Predictive modeling of an industrial UASB reactor using NARX neural network. In Proceedings of the IREC2015 The Sixth International Renewable Energy Congress, Sousse, Tunisia, 24–26 March 2015; pp. 1–6.
30. Shi, X.; Yuan, X.; Wang, Y.; Zeng, S.; Qiu, Y.; Guo, R.; Wang, L. Modeling of the methane production and pH value during the anaerobic co-digestion of dairy manure and spent mushroom substrate. *Chem. Eng. J.* **2014**, *244*, 258–263. [[CrossRef](#)]
31. Zhai, N.; Zhang, T.; Yin, D.; Yang, G.; Wang, X.; Ren, G.; Feng, Y. Effect of initial pH on anaerobic co-digestion of kitchen waste and cow manure. *Waste Manag.* **2015**, *38*, 126–131. [[CrossRef](#)]
32. Hwang, M.H.; Jang, N.J.; Hyun, S.H.; Kim, I.S. Anaerobic bio-hydrogen production from ethanol fermentation: The role of pH. *J. Biotechnol.* **2004**, *111*, 297–309. [[CrossRef](#)]
33. Wang, K.; Yin, J.; Shen, D.; Li, N. Anaerobic digestion of food waste for volatile fatty acids (VFAs) production with different types of inoculum: Effect of pH. *Bioresour. Technol.* **2014**, *161*, 395–401. [[CrossRef](#)]
34. Ifaei, P.; Karbassi, A.; Lee, S.; Yoo, C. A renewable energies-assisted sustainable development plan for Iran using techno-economic-socio-environmental multivariate analysis and big data. *Energy Convers. Manag.* **2017**, *153*, 257–277. [[CrossRef](#)]
35. De Clercq, D.; Jalota, D.; Shang, R.; Ni, K.; Zhang, Z.; Khan, A.; Wen, Z.; Caicedo, L.; Yuan, K. Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. *J. Clean. Prod.* **2019**, *218*, 390–399. [[CrossRef](#)]
36. Camberos, S.U.A.; Gurubel, K.J.; Sanchez, E.N.; Aguirre, S.A.; Perez, R.G. Neuronal modeling of a two stages anaerobic digestion process for biofuels production. *IFAC-PapersOnLine* **2018**, *51*, 408–413. [[CrossRef](#)]
37. Zuluaga, C.D.; Álvarez, M.A.; Giraldo, E. Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison. *Appl. Energy* **2015**, *156*, 321–330. [[CrossRef](#)]
38. Wang, D.; Luo, H.; Grunder, O.; Lin, Y. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. *Renew. Energy* **2017**, *113*, 1345–1358. [[CrossRef](#)]
39. Sadeghassadi, M.; Macnab, C.J.B.; Gopaluni, B.; Westwick, D. Application of neural networks for optimal-setpoint design and MPC control in biological wastewater treatment. *Comput. Chem. Eng.* **2018**, *115*, 150–160. [[CrossRef](#)]
40. Das, L.; Kumar, G.; Rani, M.D.; Srinivasan, B. A novel approach to evaluate state estimation approaches for anaerobic digester units under modeling uncertainties: Application to an industrial dairy unit. *J. Environ. Chem. Eng.* **2017**, *5*, 4004–4013. [[CrossRef](#)]
41. Zhou, P.; Li, Z.; Snowling, S.; Baetz, B.W.; Na, D.; Boyd, G. A random forest model for inflow prediction at wastewater treatment plants. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 1781–1792. [[CrossRef](#)]
42. Zhou, P.; Li, Z.; Snowling, S.; Goel, R.; Zhang, Q. Short-term wastewater influent prediction based on random forests and multi-layer perceptron. *J. Environ. Inform. Lett.* **2019**, *1*, 87–93. [[CrossRef](#)]

43. Xie, S.; Hai, F.I.; Zhan, X.; Guo, W.; Ngo, H.H.; Price, W.E.; Nghiem, L.D. Anaerobic co-digestion: A critical review of mathematical modelling for performance optimization. *Bioresour. Technol.* **2016**, *222*, 498–512. [[CrossRef](#)] [[PubMed](#)]
44. Nguyen, D.; Gadhamshetty, V.; Nitayavardhana, S.; Khanal, S.K. Automatic process control in anaerobic digestion technology: A critical review. *Bioresour. Technol.* **2015**, *193*, 513–522. [[CrossRef](#)] [[PubMed](#)]
45. Latif, M.A.; Mehta, C.M.; Batstone, D.J. Influence of low pH on continuous anaerobic digestion of waste activated sludge. *Water Res.* **2017**, *113*, 42–49. [[CrossRef](#)]
46. Boe, K.; Batstone, D.J.; Steyer, J.P.; Angelidaki, I. State indicators for monitoring the anaerobic digestion process. *Water Res.* **2010**, *44*, 5973–5980. [[CrossRef](#)] [[PubMed](#)]
47. Park, J.; Jiang, D.; Lee, B.; Jun, H. Towards the practical application of bioelectrochemical anaerobic digestion (BEAD): Insights into electrode materials, reactor configurations, and process designs. *Water Res.* **2020**, *184*, 116214. [[CrossRef](#)]