IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 21, 2021, accepted February 27, 2021, date of publication March 2, 2021, date of current version March 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063231

Soft Sensor for VFA Concentration in Anaerobic Digestion Process for Treating Kitchen Waste Based on SSAE-KELM

YUHONG WANG AND SHENGKUN WANG
College of Control Science and Engineering, China University of Petroleum, Qingdao 266580, China
Corresponding author: Yuhong Wang (y.h.wang@upc.edu.cn)

ABSTRACT Anaerobic digestion technology is the most environmentally friendly approach to treat kitchen waste. Volatile fatty acid (VFA) is an essential quality monitoring indicator in the anaerobic digestion process of treating kitchen waste. In this paper, a soft measurement method of volatile fatty acid (VFA) concentration in the anaerobic digestion process is established based on stacked supervised auto-encoder combine kernel extreme learning machine algorithm (SSAE-KELM) to improve the real-time monitoring level and resource conversion efficiency of the anaerobic digestion process. Given the problems of poor feature extraction and low accuracy and efficiency of the model, a stack supervised autoencoder is proposed to realize nonlinear and deep feature extraction of process data. Simultaneously, using the idea of the extreme learning machine to train the network significantly improves the efficiency of the model. Then, the kernel extreme learning machine is used to realize regression modelling. Besides, a combined feature selection algorithm is presented to select auxiliary variables more accurately. The simulation results demonstrate that the soft sensor model can predict the concentration of volatile fatty acids (VFA) more efficiently and accurately.

INDEX TERMS Anaerobic digestion, soft sensor, autoencoder, extreme learning machine.

I. INTRODUCTION

In recent years, kitchen waste treatment and resource reuse has become a new research hotspot. Compared with waste incineration and landfill, anaerobic digestion technology has significantly environmental and economic advantages. Through the decomposition and metabolism of various microorganisms, anaerobic digestion technology can transform solid organic waste into clean energy such as biogas under certain anaerobic conditions to realize the reduction, harmless, and resource utilization of kitchen waste. However, the anaerobic digestion process is an extremely complex biological process. Considerable parameters in the reaction process are strongly nonlinear and time-varying. Moreover, people's understanding of the mechanism of anaerobic digestion is not comprehensive. Acidification might occur in practical application, leading to the failure of the digestion process. Consequently, it is difficult to maximize the utilization of resources. Meanwhile, the lack of accurate and reliable on-line measuring instruments for key parameters in the anaerobic digestion process limits the application of advanced control in the anaerobic digestion process [1]. The stability of the anaerobic digestion process is affected by various factors. Particularly, volatile fatty acid (VFA) is the key quality indicator for the stable operation of the anaerobic digestion process. Specifically, excessive accumulation of VFA would inhibit the activity of methanogens; it would acidify the system under no effective monitoring and timely treatment, severely influencing the stability of digestion process [2]. Therefore, timely and effective monitoring of VFA concentration is crucial to ensure the stable operation of the anaerobic digestion process. At present, there are few economic and reliable on-line measuring instruments for the monitoring of VFA concentration. Most of them are obtained by off-line laboratory analysis [3], which is time-consuming and high cost and cannot meet the needs of real-time control. Soft sensing technology has the characteristics of low cost, rapid response, and reliable accuracy. The research and development of soft sensing technology for VFA concentration is of great significance to the control and optimization of the digestion process.

At present, researchers have explored on-line monitoring of the anaerobic digestion process. Some researchers use the ADM1 model for mechanism modelling [4]. However,

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Imran Tariq.

due to the lack of prior knowledge, considerable parameters in the mechanism model are uncertain, and the adaptability of the model is poor, resulting in the low accuracy of the model. Since data driven soft sensing modelling uses historical data to build the prediction model, it does not need a lot of prior knowledge nor complex mechanism analysis. It has been well applied in complex industrial processes. With the rapid development of data-driven modelling methods, many statistical methods and machine learning methods have been applied to soft sensor modelling, such as partial least squares (PLS) [5], support vector machine (SVM) [6], and artificial neural network (ANN) [7]. Nevertheless, there are still some problems such as under-fitting, over-fitting, and low prediction accuracy. With the help of the deep neural network, deep learning methods, such as deep belief network (DBN) [8] and stacked auto-encoder network (SAE) [9], can extract the original features into a more abstract high-level representation and effectively extract nonlinear potential variables. It has strong nonlinear fitting ability and excellent feature learning ability, making the deep learning method suitable for soft sensor modelling of anaerobic digestion process with strong nonlinear characteristics. Some researchers have applied a deep learning algorithm and improved algorithm to investigate complex industrial process soft sensing and achieved good results [10]–[14]. Semi-supervised deep learning method is also applied to soft sensing [15], [16]. Nonetheless, most of them have the problems of poor model feature extraction and low model training efficiency. The prediction accuracy of model still needs to be improved. Therefore, a stacked supervised auto-encoder network with extreme learning machine for training is proposed to achieve more effective feature extraction. Then, the kernel extreme learning machine is used for regression prediction.

The selection of auxiliary variables is an essential step in the establishment of soft sensor model and directly related to the final quality of soft sensor model. There are many parameters in the complex industrial process. Selecting appropriate auxiliary variables can not only effectively simplify the soft sensor model and improve the efficiency of the soft sensor model but also contribute to the prediction accuracy of the soft sensor model. At present, various feature selection methods are applied to the selection of auxiliary variables, such as minimum redundancy maximum correlation algorithm (mRMR) [17] and fast filter based on correlation (FCBF) [18]. In the process of anaerobic digestion, there are numerous known or unknown interactions among various variables [19]. Most of the traditional feature selection methods ignore the interaction between features. Therefore, a combined feature selection algorithm is proposed to select the auxiliary variables of the soft sensor to obtain higher quality auxiliary features.

For the anaerobic digestion process of kitchen waste, a deep learning soft sensor model is established in this paper to achieve a more accurate and efficient prediction of volatile fatty acid concentration. First, a combined feature selection algorithm based on the minimum redundancy and maximum correlation algorithm is proposed to select soft sensing auxiliary variables. Next, an improved neural network trained by extreme learning machine based on stacked supervised auto-encoder is presented to extract features. Finally, kernel extreme learning machine is used for regression prediction. The experimental results verify that the improved deep learning soft sensor algorithm can significantly improve the prediction accuracy and efficiency of the model.

II. DEEP LEARNING SOFT SENSOR MODEL

The general idea of deep learning soft sensor modelling is described as follows:

1) Select auxiliary variables to preprocess the raw data.
2) Extract features from the preprocessed data to learn more abstract and effective feature representation in the data.
3) Establish the regression prediction model of the target variables.

A. STACKED AUTO-ENCODER NETWORK(SAE)

Auto-encoder is an artificial neural network with a single hidden layer, composed of input layer x, hidden layer z, and output layer x'. Its network structure is illustrated in Figure 1.

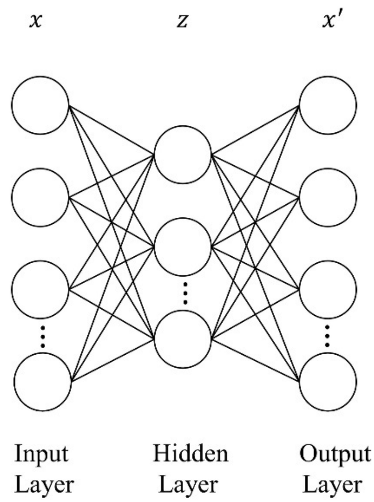

FIGURE 1. The Structure of Auto-encoder.

Auto-encoder has a good ability to extract data feature and the purpose to reconstruct the input data in the output layer. It consists of encoding phase and decoding phase and exhibits a symmetrical structure. The encoder transforms the input data x into the hidden space representation z, as expressed in Eq. (1).

$$z = \sigma \left(w_1 x + b_1 \right) \tag{1}$$

where w_1 and b_1 denote encoding weight and bias, respectively; σ represents activation function. The sigmoid function can be selected.

$$\sigma = \frac{1}{1 + e^{-x}} \tag{2}$$

Besides, the decoder transforms the hidden space representation z into reconstructed data x', as expressed in Eq. (3).

$$x' = \sigma (w_2 z + b_2) \tag{3}$$

where w_2 and b_2 represent decoding weight and bias, respectively.

The loss function can be expressed as Eq. (4), where the first term is reconstruction error, and the second term refers to regularization.

$$J(w, b) = \sum \left\| x' - x \right\|_2^2 + \lambda \left\| w \right\|_2^2 \tag{4}$$

Hidden space representation z is a more abstract hidden layer representation of input data. It can be regarded as a feature extraction layer and applied to feature extraction and data dimension reduction.

Stacked auto-encoder (SAE) is composed of multiple auto-encoders and can extract deep feature representation through layer-by-layer training. Its structure is exhibited in Figure 2.

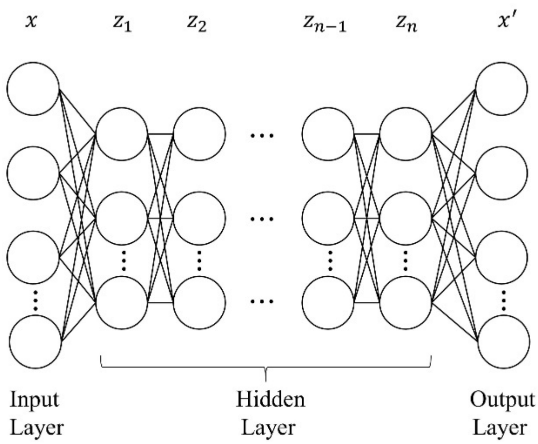

FIGURE 2. The Structure of Stacked Auto-encoder.

The model training is divided into two stages: pre-training and reverse fine-tuning. In the pre-training stage, each auto-encoder is trained separately. After the low auto-encoder training is completed, the low auto-encoder hidden layer is used as the input layer of the high auto-encoder. Next, high auto-encoder is trained. All auto-encoders are trained layer by layer. The weight and bias obtained from pre-training are used as the initial weight and bias of stacked auto-encoder. In the phase of back fine-tuning, the back propagation algorithm is adopted to fine-tune the network parameters to complete the training.

B. STACKED SUPERVISED AUTO-ENCODER(SSAE)

In the regression problem, it is crucial to make full use of label information for the performance of soft sensor model. In the stacked auto-encoder network, features are reconstructed by unsupervised learning in the pre-training stage, and the label information is used for reverse fine-tuning. This method has limited utilization of label information and cannot extract feature information more related to the label. Moreover, the calculation of the reverse fine-tuning stage is complex, and the modelling efficiency is low. Therefore, the stacked auto-encoder network is improved into a stacked supervised auto-encoder network, which realizes better use of label information to extract feature information. This benefits the establishment of the regression model and enhances the efficiency of the model without reverse fine-tuning.

The structure of supervised auto-encoder is provided in Figure 3.

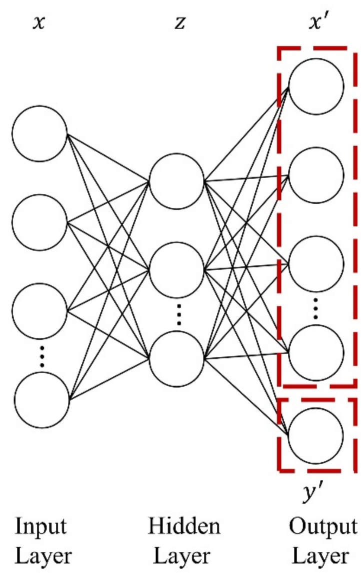

FIGURE 3. The Structure of Supervised Auto-encoder.

Different from the auto-encoder, the decoding stage of the supervised auto-encoder not only reconstructs the input information x and outputs the reconstructed value x' but also predicts the label information y and outputs the predicted value y'. The calculation formula is provided in Eq. (5).

$$[x', y'] = \sigma (w_2 z + b_2) \tag{5}$$

where w_2 and b_2 denote decoding weight and bias, respectively.

The loss function can be expressed as Eq. (6), where the first term is reconstruction error, the second term is prediction error, and the third term refers to regularization.

$$J(w, b) = \sum \left\| x' - x \right\|_2^2 + \sum \left\| y' - y \right\|_2^2 + \lambda \left\| w \right\|_2^2 \tag{6}$$

Similar to the stacking mode of the stacked auto-encoder network, the stacked supervised auto-encoder network is composed of multiple supervised auto-encoders. Its structure is presented in Figure 4.

Besides, the training method of the stacked supervised auto-encoder network is similar to that of the stacked auto-encoder network in the pre-training stage. However, the former can achieve better and more efficient feature extraction without reverse fine-tuning.

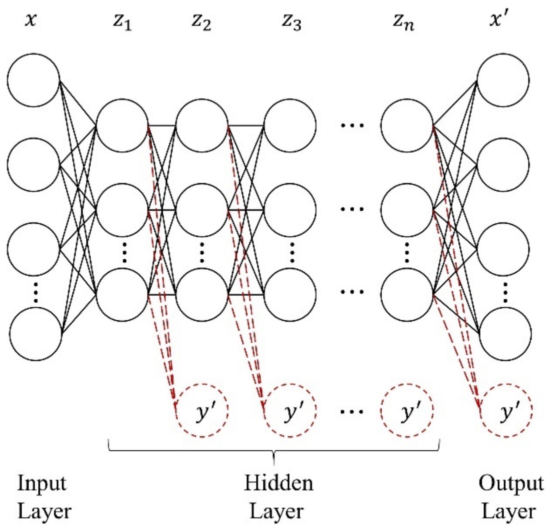

FIGURE 4. The Structure of Stacked Supervised Auto-encoder.

C. SSAE TRAINED BY EXTREME LEARNING MACHINE

For stacked supervised auto-encoder network, the traditional iterative training method based on gradient descent method is relatively inefficient. Its training speed is slow and easy to fall into a local minimum, and its performance is affected by network parameters such as learning rate. Extreme learning machine (ELM) is a learning algorithm for single hidden layer feedforward neural networks. Fortunately, using extreme learning machine algorithm to train supervised stacked auto-encoder neural network needs no iterative optimization and possesses fast learning speed and good generalization performance [20]. In extreme learning machine, the weights and biases of hidden layer neurons are generated randomly. After the number of hidden layer neurons and the weight and bias of hidden layer neurons are determined, the unique optimal solution of output layer parameters can be obtained. Besides, the genetic algorithm can be used to optimize the weights and biases of hidden layer to improve the performance and stability of the model, so as to eliminate the influence of the uncertainty of random weight and bias of hidden layer neurons on model performance.

Regarding the stacked supervised auto-encoder network, suppose that the number of input layer neurons is m, the number of hidden layer neurons is n, the number of samples is N, the input sample feature is $x \in R^{N\times m}$, the sample label is $y \in R^{N\times 1}$, the connection weight between hidden layer z and output layer x is $r_1 \in R^{n\times m}$, the connection weight between hidden layer z and output layer y is $r_2 \in R^{n\times 1}$, output layer weight is $r = [r_1,\ r_2]$, $r \in R^{n\times(m+1)}$, the output layer's expected output is $Y = \left[x',\ y'\right]$, $x' \in R^{N\times m}$, $y' \in R^{N\times 1}$, $Y \in R^{N\times(m+1)}$, and the hidden layer's output is $H \in R^{N\times n}$. Then, the reconstructed value x' and the predicted value y' of the output layer can be written as Eq. (7) and Eq. (8), respectively.

$$x' = Hr_1 \tag{7}$$

$$y' = Hr_2 \tag{8}$$

Set penalty factors C_1 and C_2. Then, the loss function J of the model can be expressed as Eq. (9), where the first term is reconstruction error, the second term is prediction error, and the third term refers to regularization to avoid over-fitting.

$$\begin{aligned} J &= \frac{C_1}{2}\left\|x - x'\right\|_2^2 + \frac{C_2}{2}\left\|y - y'\right\|_2^2 + \frac{1}{2}\|r\|_2^2 \\ &= \frac{C_1}{2}\|x - Hr_1\|_2^2 + \frac{C_2}{2}\|y - Hr_2\|_2^2 + \frac{1}{2}\|r\|_2^2 \end{aligned} \tag{9}$$

The output layer weight r is obtained by minimizing the following unconstrained optimization problems, as presented in Eq. (10).

$$\begin{aligned} \min_{r \in R^{n\times(m+1)}} J = \frac{C_1}{2}\|x - Hr_1\|_2^2 \\ + \frac{C_2}{2}\|y - Hr_2\|_2^2 + \frac{1}{2}\|r\|_2^2 \end{aligned} \tag{10}$$

In Eq. (10), the derivative of the loss function J relative to the weight r of the output layer is calculated and made equal to 0, as expressed in Eq. (11).

$$\nabla J = C_1 H^T (x - Hr_1) + C_2 H^T (y - Hr_2) + r = 0 \tag{11}$$

The output layer weight r is obtained by solving Eq. (11), as written in Eq. (12).

$$r = [(CH^T H - I_{m+1})^{-1} H^T Y]^T \tag{12}$$

where I denotes the identity matrix, and C represents the diagonal matrix.

$$C = \begin{bmatrix} C_1 & & & \\ & \ddots & & \\ & & C_1 & \\ & & & C_2 \end{bmatrix}_{(m+1)\times(m+1)} \tag{13}$$

D. KERNEL EXTREME LEARNING MACHINE REGRESSION PREDICTION

With the output of the last hidden layer of the stacked supervised auto-encoder network as the input of the regression model, the kernel extreme learning machine regression model is established to predict the target variables. The neural network model is illustrated in Figure 5.

Back propagation (BP) neural network is easy to fall into local minimum in training. ELM trains network by computing the generalized inverse matrix, can be used as regression model. Compared with BP neural network, ELM has better training efficiency and prediction accuracy.

The hidden layer parameters of the regression model established by traditional extreme learning machine are randomly mapped. Kernel extreme learning machine (KELM) introduces kernel function into the extreme learning machine, replaces random mapping with kernel mapping, enhances the generalization ability of extreme learning machine, and avoids the problem of poor stability caused by random hidden layer parameters. Moreover, it can be solved by kernel function point multiplication without setting the dimension of hidden layer feature mapping.

For the extreme learning machine, suppose the weight of hidden layer is a, the bias of hidden layer is b, the output of hidden layer is H, the number of hidden layer neurons is n, the weight of output layer is r, the input of input layer is x, the output of output layer is y, and the penalty factor is C. Then, the output of hidden layer and output layer is expressed as Eq. (14) and Eq. (15).

$$H(x) = \sigma(ax + b) \tag{14}$$

$$y(x) = Hr \tag{15}$$

where σ denotes the activation function of the hidden layer, and sigmoid function can be selected.

The loss function J is written as Eq. (16).

$$\begin{aligned} J &= \frac{C}{2}\|y-x\|_2^2 + \frac{1}{2}\|r\|_2^2 \\ &= \frac{C}{2}\|y\text{-}Hr\|_2^2 + \frac{1}{2}\|r\|_2^2 \end{aligned} \tag{16}$$

The output layer weight r can be obtained by solving the minimization unconstrained optimization problem of Eq. (17).

$$\min_{r\in R^{n\times m}} J = \frac{C}{2}\|y\text{-}Hr\|_2^2 + \frac{1}{2}\|r\|_2^2 \tag{17}$$

The output weight r is presented in Eq. (18).

$$r = H^T\left(\frac{I_n}{C} + H^T H\right)^{-1} y \tag{18}$$

The output function of the extreme learning machine is expressed as Eq. (19).

$$f(x) = h(x) H^T\left(\frac{I_n}{C} + H^T H\right)^{-1} y \tag{19}$$

If the kernel function is introduced into the extreme learning machine, the output function of the kernel extreme learning machine can be expressed as Eq. (20).

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}\left(\frac{I_n}{C} + H^T H\right)^{-1} y \tag{20}$$

where N represents the dimension of the input layer, and $K(x_i, x_j)$ indicates a kernel function that can be selected, as presented in Eq. (21).

$$K(x_i, x_j) = x_i^T x_j \tag{21}$$

III. COMBINED FEATURE SELECTION ALGORITHM

The feature combination method and the maximum information coefficient are used to improve the minimum redundancy and maximum correlation criterion and select the combined features as the auxiliary variables of soft sensing. This algorithm is named as FC-mRMR. It can fully mine feature information and is more suitable for the regression prediction task.

The steps of combination feature selection algorithm are introduced as follows: 1) extract effective combination features; 2) remove irrelevant features based on feature ranking; 3) remove the redundant features according to the minimum redundancy and maximum correlation criterion of incremental search.

Suppose the number of features in the feature space is n, the feature set is F $= \{x_1, x_2, \ldots, x_n\}$, and the prediction target variable is y. The specific process of the combined feature selection algorithm is detailed as follows.

1) Feature combination

① Use linear and nonlinear functions to represent the interaction between features.

Suppose x is the feature, $F = \{x_1, x_2, \ldots, x_n\}$ indicates the feature space with n features, and I represents the n-dimensional identity matrix. Then, the linear feature combination matrix φ_1 and nonlinear feature combination matrix φ_2 are constructed by Eq. (22) and Eq. (23), respectively.

$$\varphi_1 = [x_1 I + F, x_2 I + F, \ldots, x_n I + F] \tag{22}$$

$$\varphi_2 = [x_1 F, x_2 F, \ldots, x_n F] \tag{23}$$

② Generate effective feature combination set P.

Suppose the feature combination generated by x_i and x_j is x_k. Then, the maximum information coefficient (MIC) of the feature and the prediction target y can be calculated. If Eq. (24) and Eq. (25) are satisfied, x_k is the effective feature combination and put into the effective feature combination set P.

$$MIC(x_k, y) > MIC(x_i, y) \tag{24}$$

$$MIC(x_k, y) > MIC(x_j, y) \tag{25}$$

③ Combine the effective feature combination set P and the original feature set F to generate the complete feature set Q.

2) Remove irrelevant features.

The maximum information coefficient $MIC(x_i, y)$ between all features in set Q and prediction target y is calculated. Then, the features are sorted according to $MIC(x_i, y)$. The correlation threshold is set to α. If $MIC(x_i, y) < \alpha$, x_i is removed from the feature set.

3) Select features by incremental search

① In feature set, the feature x_i, which has the largest maximum information coefficient between feature x_i and prediction target y, is selected as the first feature of the feature subset.

$$S_1 = x_i, x_i = \arg\max_{x_i\in X} MIC(y; x_i) \tag{26}$$

② The correlation and redundancy of the remaining features in the feature set are analyzed, and the features satisfying Eq. (27) are added to the feature subset.

$$\begin{aligned} x_k = arg \max_{x_i\in Q-S_m-1} [&MIC(x_i; y) \\ &- \frac{1}{m-1}\sum_{x_j\in S_{m-1}} MIC(x_i; x_j)] \end{aligned} \tag{27}$$

$$S_k = S_{k-1} \cup x_k \tag{28}$$

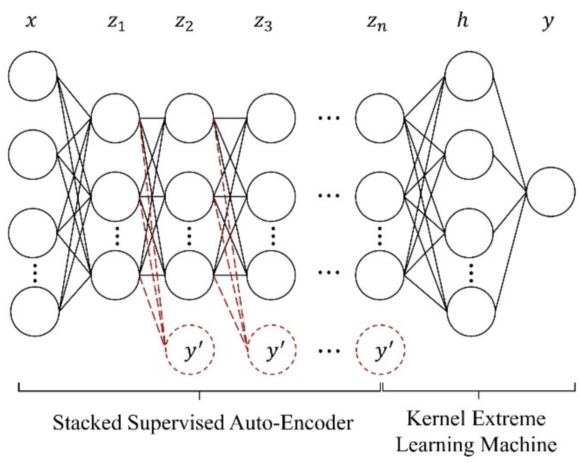

FIGURE 5. The Structure of SSAE-KELM Model.

③ Step ② is repeated until Eq. (29) reaches the maximum to obtain the final feature subset S.

$$\Phi = \frac{1}{|S|}\sum_{x_i \epsilon S} MIC\,(x_i; y) - \frac{1}{|S|^2}\sum_{x_i \epsilon S} MIC\left(x_i; x_j\right) \quad (29)$$

IV. RESULTS AND DISCUSSION

Considering the anaerobic digestion process of kitchen waste, a soft sensing model of VFA concentration based on deep learning is established. Firstly, the combined feature selection algorithm is used to select auxiliary variables of the VFA concentration soft sensing model. Then, the soft sensing model of VFA concentration is established based on SSAE-KELM neural network. This model can achieve better feature extraction, faster model training, and more accurate VFA concentration prediction ability. The steps of soft sensing model establishment are illustrated in Figure 6.

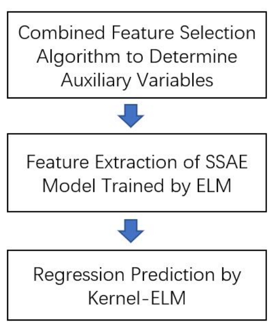

FIGURE 6. The Establishment Process of Soft Sensor Model.

According to the known mechanism of the anaerobic digestion process and the actual situation of industrial field, nine variables (solid content percentage (TS), pH value, volatile suspended solid content percentage (VS), average flow rate, alkalinity (ALK), daily gas production, CH_4 content percentage, CO_2 content percentage, and the laboratory test value of VFA concentration of the previous day) were selected as the primary auxiliary variables of VFA concentration soft sensing model.

TABLE 1. Partial data.

TS	pH	VS	Flow rate	ALK	biogas	CH_4	CO_2	VFA
2.5	7.9	51.1	175.6	10483	4153	61.2	39.4	5496
2.2	7.8	47.6	203.6	11018	5460	66.2	35.8	7119
2.1	7.9	52.3	205.7	10720	4269	55.1	47.3	6579
2.3	7.7	47.5	212.4	9529	5367	61.2	38.2	5772
1.8	7.6	58.7	142.9	6789	3634	62.1	38.7	7644

TABLE 2. Feature dimension of data set.

Number	Raw data	FC-mRMR
Feature dimension	9	4

In the actual industrial field of anaerobic digestion of kitchen waste, the data of anaerobic digestion process, including auxiliary variable and VFA concentration value, were obtained through field device collection and laboratory analysis. The abnormal data were removed and the original data were normalized to eliminate the influence of different dimensions. After data preprocessing, 464 groups of data were obtained. Partial data are listed in table 1. The characteristics of data are auxiliary variable value, and the label of data is VFA concentration value.

The results of the feature dimension preserved by combined feature selection algorithm are provided in Table 2. The four features retained in the combined feature selection algorithm are pH value, a nonlinear combination of solid content percentage (TS) and volatile suspended solid content percentage (VS), a linear combination of average flow rate and VFA concentration of the previous day, and a linear combination of CH_4 and VFA concentration of the previous day. The results of feature selection indicate that the feature dimension after combined selection is significantly less, contributing to significantly simplifying the complexity of the model and improving the modelling efficiency of the soft sensing model.

Root mean square error (RMSE) is selected as the evaluation index of the prediction performance of the model. Specifically, the smaller the RMSE, the higher the prediction accuracy of soft sensor model. The formula of RMSE is expressed as Eq. 30.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i^p - y_i^e\right)^2}{n}} \quad (30)$$

where y_i^p is the predicted value, y_i^e is the actual value, and n is the number of samples.

Moreover, BP neural network, stacked auto-encoder (SAE), extreme learning machine (ELM), and SSAE-KELM are used to establish the soft sensing model of VFA concentration, respectively, so as to compare and verify the effect of

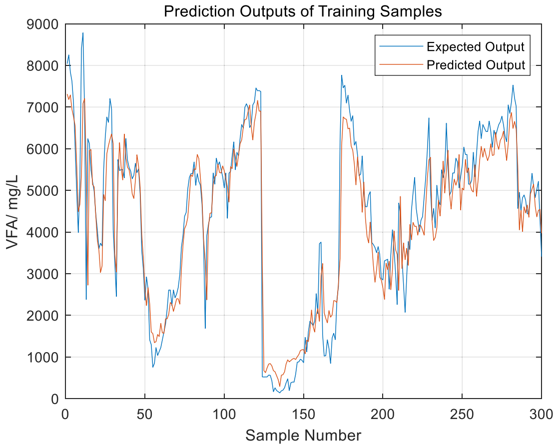

FIGURE 7. Prediction Result of BP (Training Data).

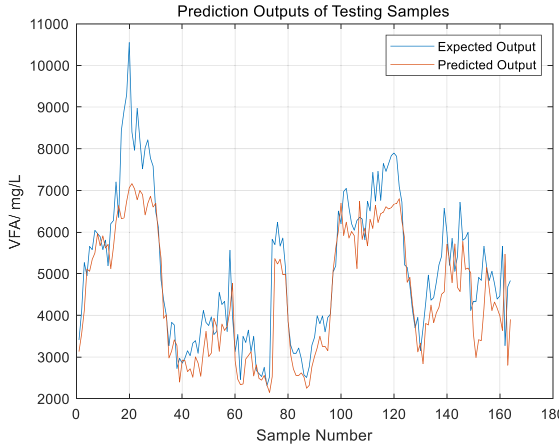

FIGURE 8. Prediction Result of BP (Test Data).

TABLE 3. The RMSE error of each prediction model.

RMSE	BP	SAE	ELM	SSAE-KELM
training	679.0970	711.9940	681.8462	762.4782
test	907.7576	834.0483	865.5735	684.0403

TABLE 4. The average relative error of each prediction model.

Model	BP	SAE	ELM	SSAE-KELM
error	17.2%	13.9%	15.4%	10.8%

the model. Compared with each method, BP and ELM has no feature extraction, SAE has deep feature extraction and SSAE-KELM has feature extraction combined with label. High quality feature extraction can help to improve the prediction accuracy of the model to a certain extent. The feature extraction ability of model can be reflected by the prediction accuracy.

Among 464 groups of actual data samples, 300 groups of data and 164 groups of data were taken as the training set and the test set, respectively. The input of neural network is auxiliary variable data, and the output is prediction value of VFA concentration. SSAE-KELM model is a five-layer neural network structure. The number of neurons in input layer is 4, the number of neurons in two hidden layers of SSAE is 7-5, and the number of neurons in output layer is 1. The parameter C_1 and C_2 of loss function are 1 and 1.4, respectively. As the comparison model, the model structure of BP neural network is 4-3-1, the model structure of SAE is 4-7-5-1 and the model structure of ELM is 4-7-1. The experimental results of the training set and test set of each soft sensing model are presented in Table 3. The average relative error of each soft sensing model are shown in Table 4. The training effect of each soft sensor model is exhibited in Figure 7-14. The average relative error of SSAE-KELM are shown in Figure 15.

As demonstrated in Figs.7-15 and Table 3-4, the soft sensor model based on SSAE-KELM algorithm has the best fitting

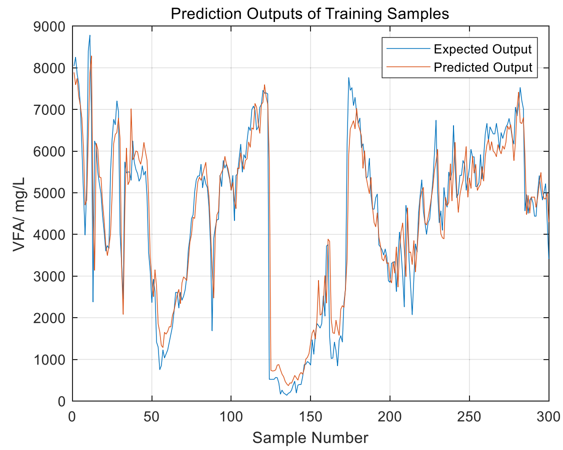

FIGURE 9. Prediction Result of SAE (Training Data).

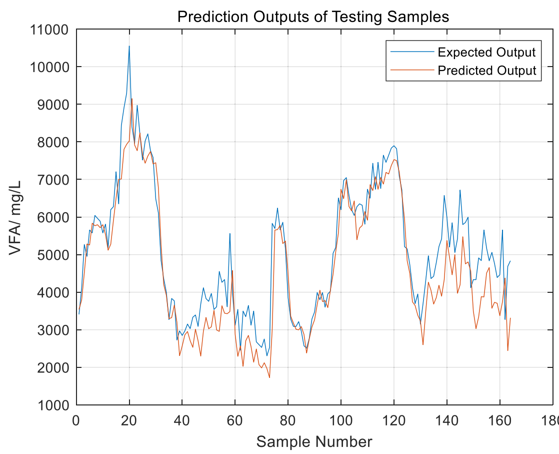

FIGURE 10. Prediction Result of SAE (Test Data).

effect, lower RMSE than the two comparison models, and the best prediction performance. Although the error of the training set based on SSAE-KELM model is slightly larger than that of the two comparison models, the error of the test set that is more essential for the soft sensor model is

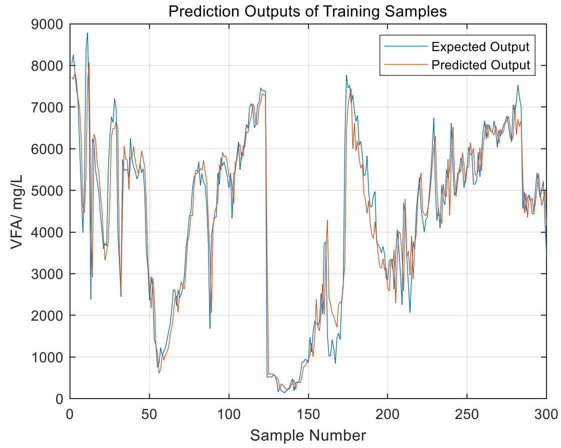

FIGURE 11. Prediction Result of ELM (Training Data).

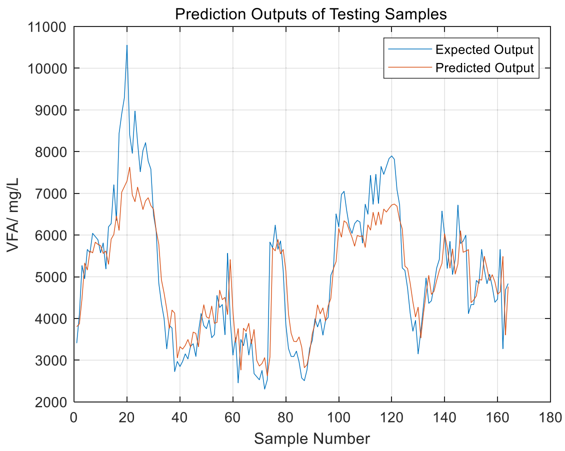

FIGURE 12. Prediction Result of ELM (Test Data).

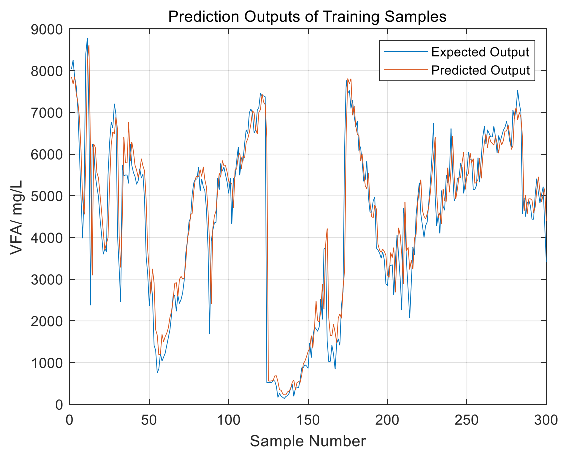

FIGURE 13. Prediction Result of SSAE-KELM (Training Data).

significantly smaller than that of the two comparison models. Compared with the BP model, the error of the test set of the SAE model decreases by 8.1%. Compared with the BP model, SAE model, and the ELM model, the error of the test set of SSAE-KELM model decreases by 24.6%, 18.0%, and 21.0%,

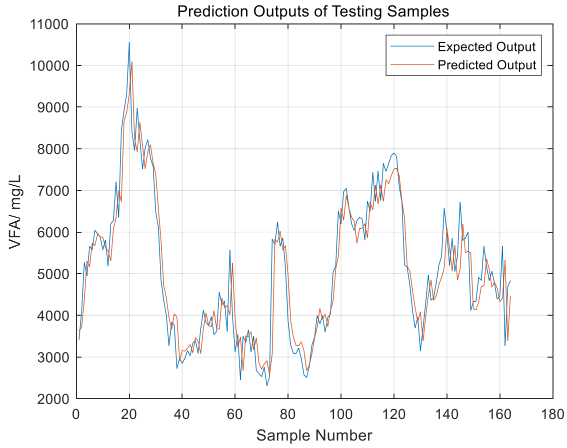

FIGURE 14. Prediction Result of SSAE-KELM (Test Data).

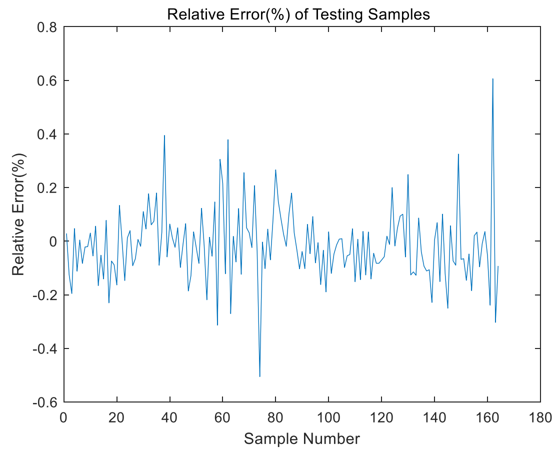

FIGURE 15. Relative Error of SSAE-KELM.

respectively. For the average relative error, SSAE-KELM has the lowest error, and the prediction effect of most samples is acceptable.

In terms of training efficiency, the calculation time of BP, SAE, ELM and SSAE-KELM model is 2.6s, 7.5s, 1.2s and 2.9s, respectively. The training mode of extreme learning machine improves the training efficiency of neural network.

It can be observed from the training effect chart that the BP model has the insufficient fitting ability for the test set owing to the limited learning ability of the shallow neural network. Particularly, the prediction error is significantly larger in the 15th-30th samples and 110th-120th samples of the test set. Due to the deep neural network structure, the SAE model has stronger learning ability and better fitting effect on the test set than the BP model. However, the prediction accuracy in some sample intervals is still insufficient. Moreover, the calculation of model training is large, and the training time is relatively long. The extreme learning machine model directly uses the raw sample data modelling and does not extract deep feature information, exhibiting a low prediction accuracy. Besides, the SSAE-KELM model has stronger generalization ability, can extract more feature information that is conducive to regression prediction, has a better fitting ability on test sets, and improves the prediction accuracy of the model. Additionally, the training method of the extreme learning machine reduces the calculation amount of the model, shortens the training time, and significantly improves the training efficiency of the model.

V. CONCLUSION

Considering the problem of low prediction accuracy and efficiency of VFA concentration soft sensing model in the anaerobic digestion process, a soft sensing model of VFA concentration with higher prediction accuracy and higher training efficiency was established based on the SSAE-KELM algorithm and feature selection algorithm. The results demonstrate that the combined feature selection algorithm can considerably simplify the model. Additionally, the SSAE-KELM algorithm can better extract features and improve the training efficiency. Furthermore, the established VFA concentration soft sensor model has higher prediction accuracy, with broad application prospects in the field of monitoring and controlling the anaerobic digestion process of treating kitchen waste.

REFERENCES

[1] D. Gaida, C. Wolf, and M. Bongards, "Feed control of anaerobic digestion processes for renewable energy production: A review," *Renew. Sustain. Energy Rev.*, vol. 68, pp. 869–875, Feb. 2017.

[2] M. Worwg and A. Kwarciak-Kozowska, "Volatile fatty acid (VFA) yield from sludge anaerobic fermentation through a biotechnological approach," *Ind. Municipal Sludge*, pp. 681–703, Jan. 2019.

[3] S. Wu, W. Gu, C. Pang, R. Dong, and Y. Deng, "Design and test of alkalinity and VFA online monitoring system in anaerobic fermentation," *Trans. Chin. Soc. Agricult. Machinery*, vol. 44, no. 10, pp. 172–176, Oct. 2013.

[4] D. Gaida, C. Wolf, C. Meyer, A. Stuhlsatz, J. Lippel, T. Bäck, M. Bongards, and S. McLoone, "State estimation for anaerobic digesters using the ADM1," *Water Sci. Technol.*, vol. 66, no. 5, pp. 95–1088, Jul. 2012.

[5] M. Zhu, T. Shi, Y. Chen, S. Luo, T. Leng, Y. Wang, C. Guo, and M. Xie, "Prediction of fatty acid composition in camellia oil by 1H NMR combined with PLS regression," *Food Chem.*, vol. 279, pp. 339–346, May 2019.

[6] C. Shang, X. Gao, F. Yang, and D. Huang, "Novel Bayesian framework for dynamic soft sensor based on support vector machine with finite impulse response," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1550–1557, Jul. 2014.

[7] K. Sun, J. Liu, J.-L. Kang, S.-S. Jang, D. S.-H. Wong, and D.-S. Chen, "Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote," *J. Process Control*, vol. 24, no. 7, pp. 1068–1075, Jul. 2014.

[8] W. Yuhong, D. Kesong, Z. Shan, S. Chao, and H. Dexian, "Melt index prediction of polypropylene based on DBN-ELM," *CIESC J.*, vol. 67, no. 12, pp. 5163–5168, 2016.

[9] H.-I. Suk, S.-W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Function*, vol. 220, no. 2, pp. 841–859, Mar. 2015.

[10] Y. Wang and X. Li, "Soft measurement for VFA concentration in anaerobic digestion for treating kitchen waste based on improved DBN," *IEEE Access*, vol. 7, pp. 60931–60939, 2019.

[11] P. Yan, B. Shen, and Y. Wang, "Soft sensor for VFA concentration in anaerobic digestion process for treating kitchen waste based on DSTHELM," *IEEE Access*, vol. 8, pp. 223618–223625, 2020.

[12] Z. He, H. Shao, P. Wang, J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105313.

[13] W. Shao, S. Chen, and C. J. Harris, "Adaptive soft sensor development for multi-output industrial processes based on selective ensemble learning," *IEEE Access*, vol. 6, pp. 55628–55642, 2018.

[14] W. Xie, J. S. Wang, C. Xing, S. S. Guo, M. W. Guo, and L. F. Zhu, "Extreme learning machine soft-sensor model with different activation functions on grinding process optimized by improved black hole algorithm," *IEEE Access*, vol. 8, pp. 25084–25110, 2020.

[15] P. Zhu, X. Liu, Y. Wang, and X. Yang, "Mixture semisupervised Bayesian principal component regression for soft sensor modeling," *IEEE Access*, vol. 6, pp. 40909–40919, 2018.

[16] B. Lu and L. Chiang, "Semi-supervised online soft sensor maintenance experiences in the chemical industry," *J. Process Control*, vol. 67, pp. 23–34, Jul. 2018.

[17] B.-Q. Li, L.-L. Hu, L. Chen, K.-Y. Feng, Y.-D. Cai, and K.-C. Chou, "Prediction of protein domain with mRMR feature selection and analysis," *PLoS ONE*, vol. 7, no. 6, Jun. 2012, Art. no. e39308.

[18] A. M. Kowshalya, R. Madhumathi, and N. Gopika, "Correlation based feature selection algorithms for varying datasets of different dimensionality," *Wireless Pers. Commun.*, vol. 108, no. 3, pp. 1977–1993, Oct. 2019.

[19] B. Eryildiz and M. J. Taherzadeh, "Effect of pH, substrate loading, oxygen, and methanogens inhibitors on volatile fatty acid (VFA) production from citrus waste by anaerobic digestion," *Bioresource Technol.*, vol. 302, Apr. 2020, Art. no. 122800.

[20] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, Jan. 2018.

YUHONG WANG received the B.Eng. and M.Eng. degrees from the China University of Petroleum at Qingdao, Qingdao, China, in 1992 and 1997, respectively, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, in 2004. He is currently a Professor with the College of Control Science and Engineering, China University of Petroleum at Qingdao. His current research interests include advanced process control, soft sensing, hybrid systems, and production planning optimization.

SHENGKUN WANG received the B.Eng. degree from the Tianjin University of Technology and Education, in 2018. He is currently pursuing the M.Eng. degree in control engineering with the College of Control Science and Engineering, China University of Petroleum, Qingdao, China. His current research interests include soft sensor and optimization for the anaerobic digestion process of kitchen waste-based on deep learning.