ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Yield and Properties Prediction Based on the Multicondition LSTM Model for the Solvent Deasphalting Process

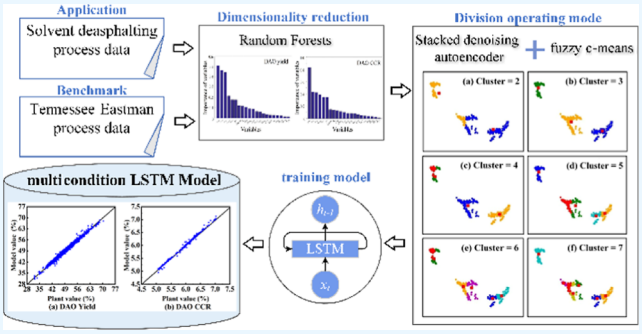
Jian Long, Yifan Chen, Dengke Cao, Pengyu Chen, and Minglei Yang*

Cite This: *ACS Omega* 2023, 8, 5437−5450

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Solvent deasphalting (SDA) is a complex multiscale continuous process. The operation mode of the SDA process is not considered in the related data-driven model. Therefore, this paper proposes a time lag process prediction model with multiple operation modes to solve the above problem. First, based on random forests, the relative importance of initial input variables in the SDA process on DAO yield and Conradson carbon residual are studied and features are selected according to the results. Then, the stack denoising autoencoder (SDAE) is used to reconstruct the data and obtain the nonlinear mapping information of hidden layers of SDAE and achieve feature dimension reduction. SDAE can improve clustering accuracy of fuzzy c-means, and the operation mode of SDA process is accurately divided. Long short-term memory (LSTM) is used to establish a multicondition LSTM model. Compared with the traditional LSTM model, the multicondition LSTM model has a higher prediction accuracy with $R^2 > 0.95$. The sensitivity analyses of the properties of feed and operating conditions on DAO yield are consistent with the principle of two-phase countercurrent extraction in the SDA process. In addition, the benchmark test of the Tennessee Eastman process shows that the proposed method is also effective in the fault detection of other processes. Because the multicondition LSTM can predict the future process measurement data according to operating mode, it can better avoid the false alarm problem and predict the fault earlier.



1. INTRODUCTION

The petroleum industry which is a highly regulated and capital-intensive industry plays a pivotal role in meeting the world's energy demand.[1] With the development and change of global issues such as climate change and environmental protection, the standards of petrochemical production processes and products have become more stringent. Therefore, modeling and simulation in a cheap and rapid way are crucial and useful to engineering design and optimization.[2] Through the industrial process forecast to guide the industrial production processes can timely adjust the industrial equipment temperature, pressure, and other operating conditions to achieve the purpose of reducing resource consumption and improving product yield. In recent years, solvent deasphalting (SDA) has attracted more and more attention in heavy oil processing due to its good adaptability of raw materials,[3,4] especially when combined with the hydrotreating process, the catalytic cracking deasphalted oil (DAO) production process, and the hydrogen production process of deoiled asphalt (DOA).[5] However, the establishment of the SDA mechanism model requires extensive process knowledge research. A data-driven approach can provide a cheaper and faster solution for process modeling. Whether for fault detection or target prediction, data-driven methods are becoming increasingly popular in industries.[6]

Data-driven methods such as principal component analysis (PCA), independent component analysis, and k-nearest neighbor are commonly used in data analysis, which can realize fault detection and diagnosis in the industrial production process.[7] Support vector machine (SVM) and artificial neural network (ANN) can be used to solve nonlinear problems in industrial data.[8] Combining PCA with a quantitative operational risk assessment model can solve the problem of false alarms.[9] The use of Kalman filtering can also avoid false alarms.[10] Hu et al. established a fracturing pressure prediction model by using the locally weighted linear regression approach in a data-driven way and proposed a delicate early warning scheme of fracturing screenout event(s) for practical application in the field.[11] Kumari et al. used the selected attributes to develop an ANN model to predict the causes of an incident. For each cause, another ANN model was

Received: October 14, 2022
Accepted: January 24, 2023
Published: February 3, 2023

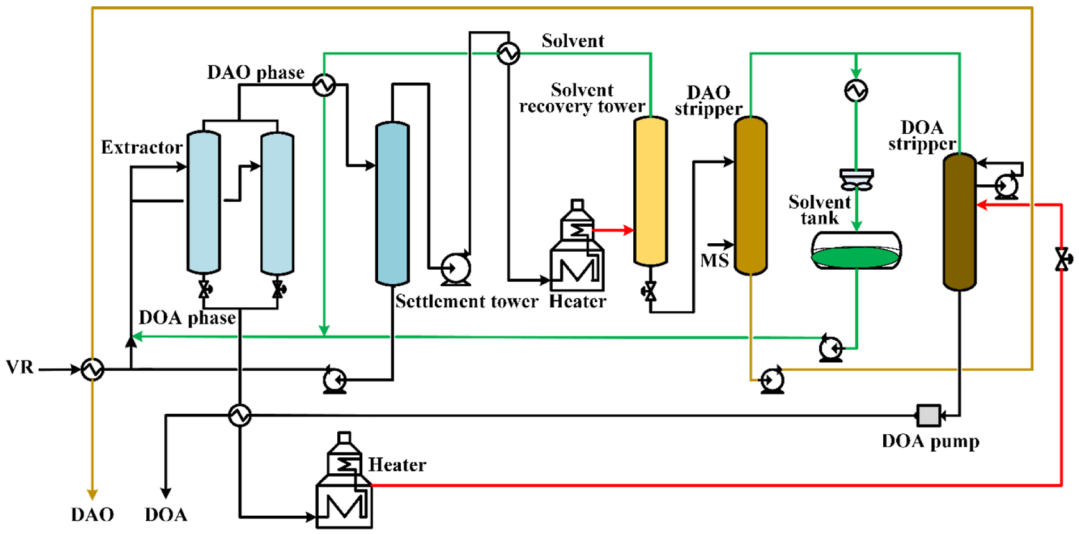


Figure 1. Industrial flowchart of SDA.

developed to predict the sub-cause of the incident. The integration of cause and sub-cause prediction models enabled the efficient causation prediction of incidents.[12]

Some models solve industrial production problems by combining multiple data-driven approaches, such as a real-time hybrid method based on PCA and Bayesian belief network,[13] and limited diagnostic information from kernel PCA, other online fault detection and diagnostic tools, and process knowledge combined through Bayesian belief network.[14] Amin proposed an integrated method for process fault detection and diagnosis and propagation pathway identification, which can detect the fault earlier and ensure the correctness of the diagnosis. The methodology was developed by combining the early fault detection capacity of the multivariate exponentially weighted moving average PCA and robust diagnosis and causality representability of the Bayesian network.[15] The existing models did not consider the existence of cycles that are common in chemical processes, Kumari et al. proposed a modified Bayesian network in this work that accounted for cyclic loops to improve the accuracy of root cause diagnosis.[16] Chaudhari et al. adopted multiple feature selection methods to select descriptors through a data-driven method and compared the prediction accuracy of the model based on random forests (RFs), neural network, and genetic function approximation.[17] Arunthavanathan et al. used the convolutional neural network (CNN)—long short term memory (LSTM) approach to forecast the system parameters for future sampling windows' recognition and an unsupervised one-class-SVM (OC-SVM) was used for fault symptoms' detection using forecasted data window. The results confirmed that the proposed method effectively detected potential fault conditions in multivariate dynamic systems by detecting the fault symptoms early as possible.[18]

Based on the above, the data-driven method has been widely used in fault detection and process prediction in the industrial production process. However, in industrial production under different operating modes, there are complex and different nonlinear relations between the production index and the operating conditions, and there is always a strong coupling between the operating conditions. Changes in these operating conditions affect coupling strength and may result in different operating modes. In addition, the coupling between operating conditions results in redundant variables in the data. Therefore, this paper aims to study a process prediction method of multicondition division to solve the problem of working condition change in industrial processes. Compared with other methods, the proposed method can make a more accurate prediction according to different working conditions, and it can provide guidance for the prediction and fault detection of any industrial production process with the change of working conditions. First, the RF is used to select the important factors affecting the industrial production process. Then, the stacked denoising autoencoder (SDAE)—Fuzzy c-means (FCM) algorithm is used to divide the working conditions. Finally, the LSTM neural network is used to establish the prediction model of SDA-related yield under multiple working conditions.

The main contributions of this paper are as follows:

(1) RFs is used to evaluate the importance of variables. More than 60 variables are voted by out-of-bag (OOB) error, and 21 key variables are finally selected as the input of the model.

(2) A new framework for operation mode classification is proposed. SDAE is used to reduce the dimension to improve the clustering accuracy and then FCM is used to identify the operation mode of the data.

(3) On the basis of operating mode division, LSTM is used to establish a multimode integration model based on the voting method. Also, the sensitivity analyses of the properties of feed and operating conditions on DAO yield are consistent with the extraction principle of the SDA process.

(4) The benchmark Tennessee Eastman process is used to test efficacy of multicondition LSTM and confirm that multicondition LSTM can reduce false alarms and effectively detect potential fault conditions by detecting the fault symptoms earlier.

This paper is organized as follows: the background and current research studies of this paper are described in Section 1. Section 2 introduces the SDA process and related data-driven methods. Section 3 describes the modeling process in detail including feature selection, time delay analysis, operation mode division, and model comparison. Moreover, the

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

established model is benchmarked by Tennessee Eastman process data. Finally, conclusions are mentioned in Section 4.

2. PROPOSED METHODOLOGY

2.1. Study Area. *2.1.1. Process Principle.* The vacuum residue containing high concentration of asphaltenes (As) can be deasphalted by the SDA process which can make certain DAO have the characteristics of good cracking performance and low impurity content. As shown in Figure 1, the overall study area of this paper is a 5×10^5 tons/year UOP's DEMEX SDA industrial unit, which is a quite popular industrial SDA unit due to high solvent recovery, energy-saving capability, and high yield of DAO. It is mainly composed of sub-critical solvent extraction, supercritical solvent recovery, and stripping.

In the industrial flowchart, VR and mixed C_4 (solvent volume composition: 0−3% C_3, 0−3% $C_3^=$, 13−20% *n*-C_4, 29−50% *i*-C_4, 28−49% $C_4^=$, and C_5 < 2%) are adopted to be the feed and solvent, respectively. After being thoroughly mixed, they are fed into the extractor from the upper and lower parts of the column, respectively. The oil is solubilized in the solvent while the insoluble pitch will precipitate out of the mixed feedstock as As. The solvent extractor is designed to separate the DAO phase (the extract phase) and the pitch phase (DOA phase, the raffinate phase) under subcritical conditions. They are taken from the bottom and top of the column, respectively. The settlement tower is used to separate the resin in the DAO phase at a higher temperature than the extraction tower. The solvent recovered from DAO separator (supercritical solvent recovery tower) under high pressure is combined with solvent condensation recovered from DOA and DAO strippers under low pressure to recycled back to initial stage. DAO and DOA after solvent removal are sent out of the plant. DAO is normally used as feedstocks of hydrocracking or fluid catalytic cracking because of its low contents of metal (Ni + V). DOA is usually the raw material for gasification or asphalt production.[19]

The choice of the extraction process or optimization of process parameters in the SDA process will determine the properties and yield of DAO. Under a certain solvent composition and pressure, the lower temperature of the extractor, the higher the DAO yield and the lower the DAO properties. While the DAO yield remains constant with an increase in solvent, the degree of separation of individual components will be improved, then a better quality (properties) of DAO will be obtained. Therefore, an accurate SDA process model is beneficial to analyze the influence of operation variables and raw material content in the production process on DAO yield.[20]

2.1.2. Importance of Yield and Properties of DAO. The SDA process plays a vital role in upgrading heavy oils and removes impurities that can reduce the impurity and improve the cracking performance content of DAO. Carbon residue is an important index for evaluating its coking tendency. The composition of saturates (S), aromatics (A), resins (R), and As dictate its cracking performance. The content of the impurities in DAO such as sulfur, nickel, and vanadium are important evaluation parameters of process operation for the SDA process. The SDA process is an intermediate link in the whole refining process, and the yield and quality of DAO will affect the production efficiency of the next link. Among the many properties affecting DAO cracking performance, CCR, S, A, R, and As are the most important properties of DAO, where the calculation of DAO's product yield and the solvent ratio refers to eqs 1 and 2.

$$\text{the DAO yield} = \frac{\text{DAO flow}}{\text{total feed flow}} \times 100\% \quad (1)$$

$$\text{the solvent ratio} = \frac{\text{total solvent flow}}{\text{total feed flow}} \times 100\% \quad (2)$$

2.2. RFs for Variable Importance Measurements. RFs is an ensemble learning method. RFs has higher accuracy in feature importance evaluation due to its strong robustness.[21] In the industrial production process, the characteristic of high dimension of industrial data often affects the complexity and prediction performance of the model. A forest-based variable selection algorithm can analyze the distribution of variables to measure the importance of each variable, and it has been proved in several cases that the algorithm can identify important information in different applications.[22] There are also cases comparing and analyzing the performance of different methods based on RFs, such as recursive feature elimination, Boruta, and VarSelRF algorithm.[23] The feature screening process of RFs is shown in Figure 2. The most important variable is screened out through importance evaluation of each variable.

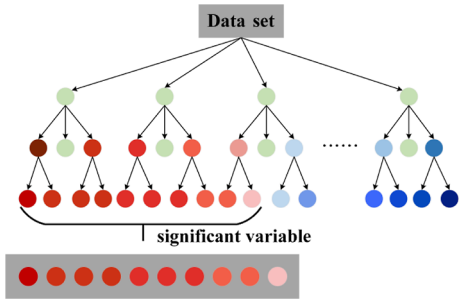


Figure 2. Sequential logic architecture of RFs.

2.3. Stacked Denoising Autoencoder. Autoencoder (AE) is an unsupervised neural network that has powerful nonlinear expression ability to achieve feature dimension reduction.[24] Deep structure AE can more accurately deal with the problem of nonlinear distribution. In image processing, such as thermal image, the process of extracting features from the model can be revealed by layer-by-layer visualization.[25] DAE is a variant of AE. The first is to corrupt original input by adding some disturbance like Gaussian noise and Mask noise. Then, the parameters of DAE are trained for reconstructing the clean input. The objective of DAE is to learn denoised features which can be decoded back into the clean input, which can make the encoded and extracted features have strong robustness.[26] The process of DAE extracting features is shown as Figure 3. First, the original input x is destroyed by matrix of noise q^d to get x' and then the latent variable y is obtained by encoding operator f_θ. Finally, the data are decoded and restored by operator g_θ to get h. $L(x,h)$ is the reconstruction error function, which aims to update the network weight so that the restored data h is as consistent as possible with the original input x.[27] Where, the latent variable y is the feature to be proposed.

SDAE forms a deep network structure by stacking DAE. Through greedy layer-wise training, each self-coding layer is trained separately without supervision. The input data are

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

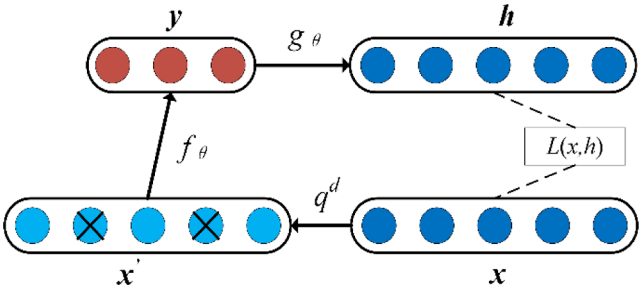


Figure 3. Structure of DAE.

reconstructed, and the likelihood function between the input and reconstructed data are minimized to obtain the information of the hidden layer in the SDAE network.

2.4. Fuzzy C-Means. Fuzzy c-means is a kind of fuzzy clustering algorithm with membership function, which has strong robustness. The membership function can express the degree to which a sample x belongs to the set A, and the range of its independent variables is all possible objects belonging to the set A. This is the fuzzy idea in FCM algorithm. Objective function of FCM is essentially the sum of Euclidean distances (the sum of squared errors) from each point to each class.[28] Suppose a data set A is divided into C classes, then there are C class centers, and μ_{ij}^m represents the membership degree of sample x_i to clustering center c_j. $\|x_i - c_j\|^2$ is the square of the Euclidean distance between each sample point x_i and the current class center point. Therefore, our objective function J_m is equivalent to making each sample point have a corresponding class, so that the distance from the sample point to the class center is the shortest. The process of clustering is the process of minimizing the objective function.[29] When the objective function converges, the final clustering result can be obtained. The objective function refers to eq 3.

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \|x_i - c_j\|^2, \qquad 1 \le m < \infty \tag{3}$$

eqs 4 and 5 are formulations of membership degree and cluster center, respectively. The clustering principle of the highest intra-group similarity and the lowest inter-group similarity can be ensured by minimizing the objective function by finding the clustering center c_j of each group.[30]

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{2/m-1}} \tag{4}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{5}$$

eq 6 is the termination condition, where t is the number of iteration steps and ε is a small constant representing the error threshold. u_{ij} and c_j are updated iteratively until the maximum change of membership degree does not exceed the error threshold. This process eventually converges to local minima or saddle points of J_m.

$$\max_{ij}\{|u_{ij}^{(t+1)} - u_{ij}^{(t)}\} < \varepsilon \tag{6}$$

Steps of FCM algorithm:

(1) Select the number of categories C and the appropriate m to initialize the matrix u_0 determined by the membership function (initialized between random values [0,1]).
(2) Calculate the central value of clustering c_j.
(3) Calculate the new membership matrix u_{ij}.
(4) Compare $u_{ij}^{(t+1)}$ and $u_{ij}^{(t)}$. If the change between them is less than a certain threshold, stop the algorithm; otherwise, turn to (2).

2.5. Modeling Method for Prediction of Yield and Properties of DAO. *2.5.1. Long Short-Term Memory.* LSTM[31] is a variant of recurrent neural network, whose internal structure includes memory gates that can save part of information. Therefore, LSTM partly overcomes the problems of vanishing and exploding gradients of the traditional recurrent neural network. LSTM selects the information to be remembered and the information to be forgotten through memorizing (input gate, forgetting gate, and output gate) and two states (cellular state and hidden state), so as to make long-term dependence of network learning.[32] Due to the memory function of LSTM, LSTM is gradually applied to industrial time series analysis and prediction. A machine learning method combining CNN and LSTM can accurately predict the entire particle trajectory and surface erosion profile.[33] LSTM can effectively predict the CH_4 leakage source of chemical processes, and it is found that the model with larger time steps has better performance and stronger generalization ability.[34] A maximum correntropy criterion-based LSTM (MCC-LSTM) neural network is proposed to develop a reliable soft sensor model for quality prediction, which can identify the outliers and reduce their negative effects on the prediction in some extent.[35]

The structure of LSTM cell is shown in Figure 4. The LSTM network includes four memory gates (for a sample i as

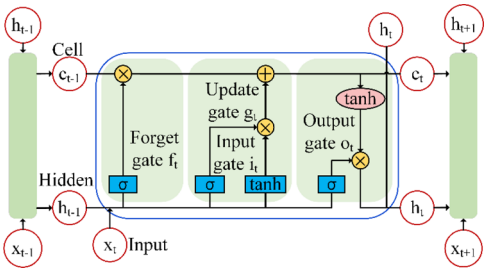


Figure 4. Structure of LSTM cell.

$\boldsymbol{c}_t^i \in \mathbb{R}^{d_h \times 1}$, input $\boldsymbol{i}_t^i$, forget $\boldsymbol{f}_t^i$, update $\boldsymbol{g}_t^i$, and output $\boldsymbol{o}_t^i$ gates $\in \mathbb{R}^{d_h \times 1}$),[36] which can selectively memorize and forget the relevant information. The forget gate through eq 7 chooses to forget some information from the past moment and input gate (refer to eqs 8 and 9) and remember some information from the present moment. The update gate combines past and present information to generate new information by eq 10. The output gate can determine the information in the state to be routed to the output and the calculation as shown in eqs 11 and 12.

$$\boldsymbol{f}_t^i = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{X}_t^i, \boldsymbol{h}_{t-1}^i] + \boldsymbol{b}_f) \tag{7}$$

$$\boldsymbol{i}_t^i = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{X}_t^i, \boldsymbol{h}_{t-1}^i] + \boldsymbol{b}_i) \tag{8}$$

$$\boldsymbol{c}_t^i = \boldsymbol{f}_t^i \odot \boldsymbol{c}_{t-1}^i + \boldsymbol{i}_t^i \odot \boldsymbol{g}_t^i \tag{9}$$

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

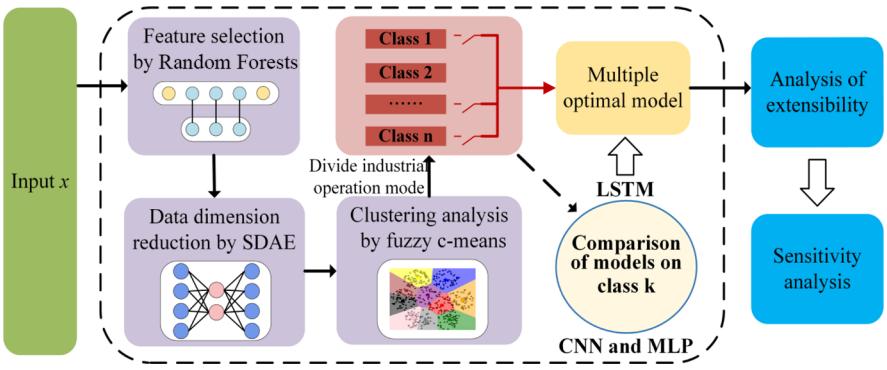


Figure 5. Workflow of the predictor of DAO properties.

$$\boldsymbol{g}_t^i = \tanh(\boldsymbol{W}_g \cdot [\boldsymbol{X}_t^i, \boldsymbol{h}_{t-1}^i] + \boldsymbol{b}_g) \tag{10}$$

$$\boldsymbol{o}_t^i = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{X}_t^i, \boldsymbol{h}_{t-1}^i] + \boldsymbol{b}_o) \tag{11}$$

$$\boldsymbol{h}_t^i = \boldsymbol{o}_t^i \odot \tanh(\boldsymbol{c}_t^i) \tag{12}$$

where $\sigma()$ is the sigmoid activation function; tanh() is the hyperbolic tangent function; $\boldsymbol{w}_i$, $\boldsymbol{w}_f$, $\boldsymbol{w}_g$, and $\boldsymbol{w}_o$ are the weights of each gating unit; $\boldsymbol{b}_i$, $\boldsymbol{b}_f$, $\boldsymbol{b}_g$, and $\boldsymbol{b}_o$ are the bias of each gating unit; and $\odot$ is the Hadamard product. Additional LSTM details are in Piyush Agarwal et al.[37]

2.5.2. Convolutional Neural Network. CNN is a deep neural network with strong feature extraction ability. Its main characteristic is the use of a shared parameter filter to scan the previous feature graph, which can significantly reduce the size of the parameter space.[38] The convolution layer extracts local features of input data through convolution operation. The size and strides of the convolution kernel determine the feature extraction effect of the convolution layer.[39] The results of convolution are down-sampled by the pooling layer to obtain local significant feature information, which aim to reduce network parameters, reduce computation, and avoid overfitting to some extent.[40] The full connection layer is usually located in the last several layers of CNN, which is used to collect information and obtain the value of the predicted target.

2.5.3. Multilayer Perceptron. Multilayer perceptron (MLP) is a widely used ANN with multilayer feedforward structure and has nonlinear system modeling capability.[41] The MLP contains at least one hidden layer, which is fully connected to each other.[42] MLP uses activation functions to find mathematical relationships between inputs and outputs. In the data set of nonlinear relationship, nonlinear activation functions such as logistic function (Sigmoid) and hyperbolic tangent function (tanh) are often used.[43]

2.6. Model Training Strategy and Testing. The data set used in this paper is from the production data of the SDA industrial device. Because samples have different physical meanings and dimensions, they cannot be directly used as input data for network training. SDA data were normalized to [0,1] through eq 13. RFs is used for feature selection. After the data are reduced to two dimensions by DAE, FCM clustering algorithm is used for cluster analysis to realize the division of SDA process conditions. LSTM sub-models under different operating modes are established. In the training process, the appropriate loss function and optimizer need to be selected. Also, the grid method is used to optimize the network structure and parameters for each model. During the model testing phase, the output data are inversely normalized through eq 14. The mean absolute percentage error (MAPE), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and *R*-square (R^2) are adopted to test models and the evaluation function is shown in eqs 15−19.

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

$$x = X \times (x_{\max} - x_{\min}) + x_{\min} \tag{14}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{16}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{17}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{\sum_{i=1}^{n}(\overline{Y}_i - Y_i)^2} \tag{19}$$

where Y_i is the actual value, $\hat{Y}_i$ is the predicted value, $\overline{Y}_i$ is the average value of the actual value, and n is the number of test samples. The lower the values of MAPE, RMSE, and MRE, the better the performance of the model. When the value of R^2 is closer to 1, the prediction effect of the model is more accurate.

Finally, through the comparison of model performance, the superiority of the model is proved. The workflow of the predictor of DAO properties is shown in Figure 5.

3. RESULTS AND DISCUSSION

3.1. Data and Preprocessing. In this paper, the data were collected from the laboratory information management system of the SDA process of a petrochemical company in China and were processed as follows: (1) Due to noise, environment, and other reasons, the sensor cannot detect the value or the detected value is abnormal. Therefore, it is necessary to eliminate the data of large fluctuation period and instrument abnormal period and fill in the missing value. First, some outliers were eliminated through expert experience. Second, the remaining outliers are eliminated by using the 3σ principle. Missing values were processed by means of mean filling. (2)

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

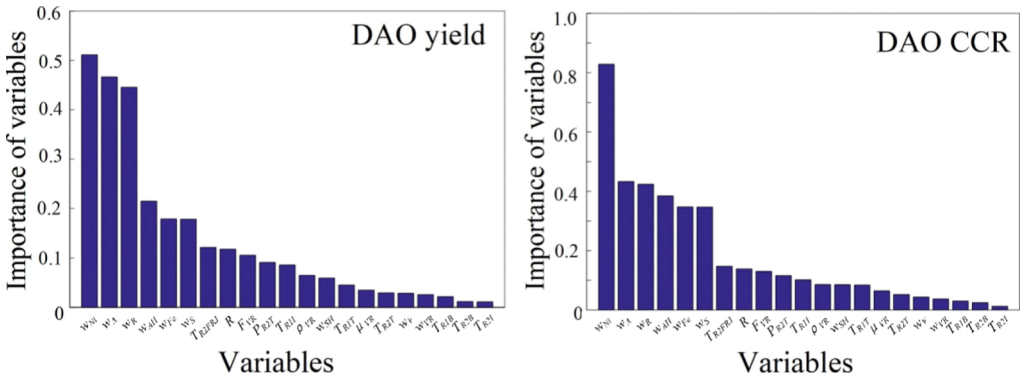


Figure 6. Relative importance of the initial input variables in the SDA models.

Compared with the frequent changing and sufficient operational data in the DCS, the update period of the properties data in the laboratory information management system was lengthy which was generally 8, 24 h or even several days and the data were insufficient. At the same sampling frequency, there may be missing data for oil properties. Therefore, interpolation was used to process the samples to integrate the properties data and operating condition data according to the sample timestamp, which made the sample points continuous in time. (3) The calculation of DAO's product yield and solvent ratio refers to eqs 1 and 2. This paper collected over 25,000 historical data accumulated during the operation period from April 2017 to July 2017. The data set included a total of 65 variables such as raw materials and operational variables introduced, and the data acquisition interval was 5 min. Through the preprocessing of missing values and outliers, as well as the calculation of relevant yields, 25,000 samples were finally retained as the raw data of the model. In addition, our data were continuous numerical data, so there was no need to transform the data. In this paper, there raw data will be used to establish the prediction models of yield, CCR, A, As, R, and S of DAO in the SDA process.

3.2. Importance of Variables. Because the high dimension of the original data will increase the complexity of the model and make the model easy to overfit, using RFs for feature selection can reduce the number of variables. RFs was used to divide the raw data into training data and OOB data and take sample from training data by bootstrapping procedure. Then, the importance of each feature was calculated and the features were voted on using the OOB error to determine the selected variable. Through experiments, 21 variables which consisted of 11 properties items and 10 operating conditions were selected as the input variables of the prediction models. Figure 6 shows the relative importance of the top 21 initial input variables in the models of DAO yield and DAO CCR. As shown, the properties of feed such as S, A, R, As content, and metals' content are very important in predicting the yield and CCR content of DAO. The temperatures of extraction tower (T_{R1T} and T_{R1B}) have a significant influence on DAO yield. Table 1 shows the detailed information of model input variables after feature selection.

3.3. Analysis of the Time Lag Effect between Yield and Properties of DAO and Operating Conditions. The complexity, multiscale characteristics, and continuity of the SDA process result in many variables affecting product yield and performance. At the same time, these characteristics also lead to significant differences of production data in frequency,

Table 1. Input Features of the SDA Process Model

feature type	Description	unit
feed properties	density (ρ_{VR})	kg/m^3
	kinematic viscosity (μ)	mPa/s
	Conradson carbon residue content (w_{CCR})	ω %
	sulfur content (w_S)	ω %
	iron content (w_{Fe})	mg/kg
	nickel content (w_{Ni})	mg/kg
	vanadium content (w_V)	mg/kg
	saturated hydrocarbons (w_{SH})	ω %
	aromatic hydrocarbons (w_{AH})	ω %
	resin content (w_R)	ω %
	asphaltene content (w_A)	ω %
	mass flow (F_{VR})	t/h
operating variable	inlet temperature of extraction tower (T_{R1I})	°C
	top temperature of extraction tower (T_{R1T})	°C
	bottom temperature of extraction tower (T_{R1B})	°C
	inlet temperature of settlement tower (T_{R2I})	°C
	top pressure of settlement tower (P_{R2T})	Mpa
	top temperature of settlement tower (T_{R2T}	°C
	bottom temperature of settlement tower (T_{R2B})	°C
	diluent solvent temperature (T_{R2FRJ})	°C
	the ratio of solvent to VR (R_A)	

diversity, and relationship time lag. Traditional ANN cannot effectively extract the information on the time lag relationship between model input (various operational variables and raw material properties data) and model output (yield and properties of DAO). Therefore, in this paper, the LSTM neural network was used to analyze the time lag effect of related operational variables on yield and properties of DAO in this paper. In the process of time lag analysis, the production process was divided into a mixing unit of residue and solvent, an extraction tower, and a settlement tower according to the flow sequence of materials in the device. The initial prediction model that took DAO yield as the prediction target was constructed, and the time lag relationship between operating conditions (in three units) and DAO yield was investigated in different time window sizes. The MAPE between the actual value and the predicted value of DAO yield was taken as the judgment standard. Finally, three timing analysis results were obtained, as shown in Figure 7. When the time window sizes were 25, 20, and 15 min, respectively, the MAPE of the model results reached the minimum. That is, through the existing data acquisition method, the DAO yield data had a lag effect of 25,

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

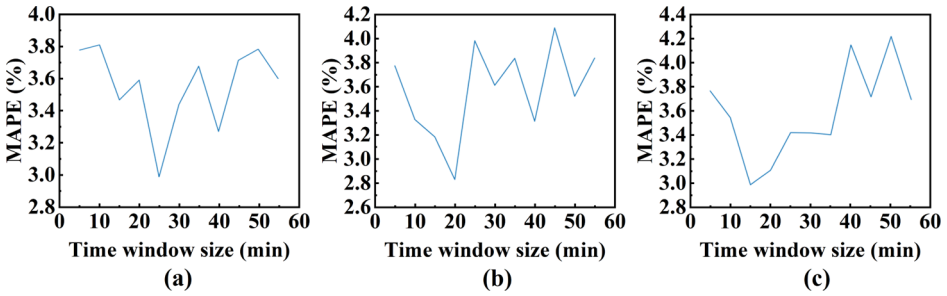


Figure 7. Time lag effect of operating conditions on DAO yield under time window sizes relative to (a) residue and solvent mixing unit, (b) extraction tower, and (c) settling tower.

20, and 15 min, respectively, relative to the mixing unit of oil and solvent, extraction column, and sedimentation column. The results of this time lag analysis also conformed to the production principle of the SDA process, that is, after mixing residuum and solvent, it entered the extraction tower for two-phase reverse extraction and further settled the R in residuum.

3.4. Model Development and Assessment. *3.4.1. Effect of Dimensionality Reduction on Clustering Results.* Due to different operating modes, the feature distribution of sample points is different. In the process of clustering, sample points with similar feature distribution will form clusters as the same category. However, the high-latitude characteristics of industrial data will increase the computational amount and complexity of clustering and affect the clustering results, so dimensionality reduction is conducive to improving the accuracy of clustering. In addition, different dimensionality reduction methods also have different effects on the clustering effect. As shown in Figure 8, iris, wine, abalone, and glass were

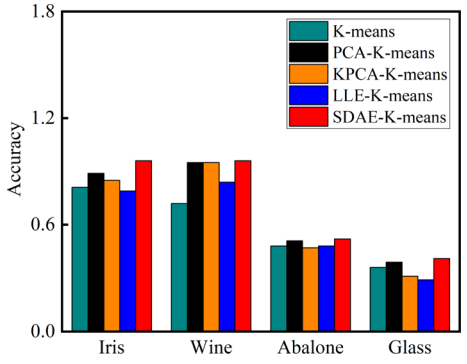


Figure 8. Effects of different dimensionality reduction methods on clustering results.

used for verification in this paper. Taking K-means clustering algorithm as an example, the effects of PCA, KPCA, LLE, and SDAE on clustering accuracy were investigated (all dimensionality reduction methods were reduced to two dimensions). Compared with the clustering accuracy without dimension reduction, KPCA and SDAE with nonlinear expression ability can improve the clustering accuracy in the four data sets, and SDAE-K-means has the highest clustering accuracy in the four data sets.

Clustering in actual production often does not have very clear boundary division. Different from traditional hard clustering methods such as K-means, FCM is a flexible clustering method, which determines the degree to which each sample belongs to each cluster center through the concept of fuzzy membership degree. Therefore, in order to make the clustering results more fit, SDAE-FCM was used in this paper to cluster the data of the SDA process to achieve the purpose of working condition division. SDAE was used to reduce the data dimension from 21 to 2. The number of hidden layer nodes was 10, 5, and 2, the activation function was the sigmoid function, the noise selection randomly set 40% of input features to 0, the learning rate was set to 0.007, and the number of training rounds of each layer was set to 3000. FCM was used to cluster the data after the dimension reduction of SDAE, and the clustering effect was evaluated by eq 20, where, $tr(\cdot)$ is the trace of the matrix and n is the number of sample points.

$$\frac{tr(UU^{T})}{n} \tag{20}$$

The number of clusters was set to an integer ranging from 2 to 7. The visual clustering effect is shown in Figure 9. When the number of clusters was 2, 4, 5, 6, and 7, the samples in a certain region can be divided into two categories. When the number of clusters was 3, the clustering result was significantly better, and the value of FPC was the highest. Therefore, the SDA process was divided into three operating modes.

3.4.2. Optimization and Validation of the LSTM Model. According to the operating mode division result as mentioned in Section 3.4.1, this paper selected 5000 sets of data under a certain operating mode and were normalized by eq 13. The data were divided into training set and test set according to the time series, in which the training set (validation set was also extracted 20% of training set according to time series) and test set was 80 and 20% of the 5000 groups of data, respectively. The established model performance by LSTM with different structures (single layer, double layer, and three layer) and numbers of neurons will be evaluated by using the grid method to obtain optimized structures of the LSTM neural network. According to Table 2, the LSTM model with three hidden layers had minimum RMSE on DAO yield (30 neurons in the first layer, 50 neurons in the second layer, and 50 neurons in the third layer) and was finally adopted in this paper. In addition, the performance of LSTM models with different time window sizes were evaluated by MAPE evaluation function, and the optimal time window size was finally obtained as 25 min. Figure 10 shows that the convergence trend of the DAO yield model on the training set and the validation set, and the convergence result of the model on the validation set is close to the convergence result on the training set. Figure 11 shows the fitting effect and prediction error of training samples and test samples on the model, which indicates that there is no overfitting in the training of the model.

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

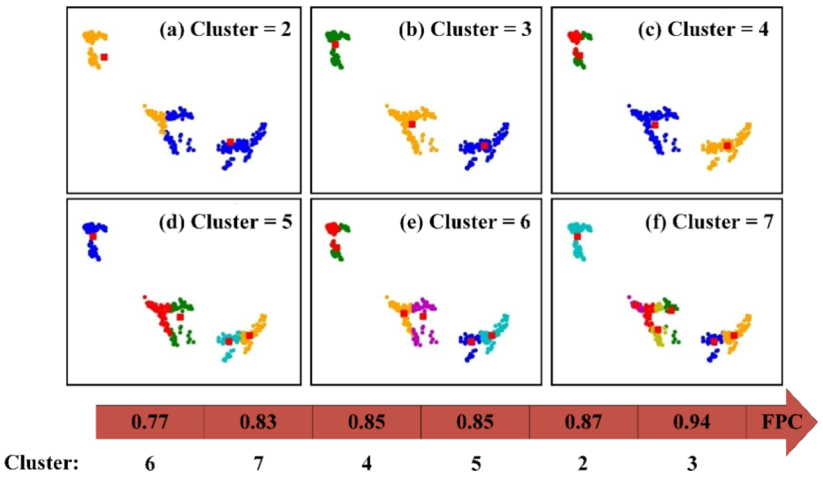


Figure 9. When the number of clusters is different, the clustering diagram of FCM and the quantification of clustering effect of the fuzzy partition coefficient are shown. (a−f) clustering effects of different clustering quantities, respectively. The better the clustering effect, the higher the FPC.

Table 2. Performance Comparison of LSTM Models with Different Network Structures

model	number of hidden layers	parameter	setting hyperparameters	selected hyperparameter	RMSE
LSTM	1	number of neurons in the first layer	[40, 45, 50, 55, 60, 65, 70]	60	0.613
		activation	[tanh, sigmoid, relu]	relu	
	2	number of neurons in the first layer	[30, 35, 40, 45, 50, 55, 60]	50	0.595
		number of neurons in the second layer	[40, 45, 50, 55, 60, 65, 70]	65	
		activationr	[tanh, sigmoid, relu]	relu	
	3	number of neurons in the first layer	[20, 25, 30, 35, 40, 45, 50]	30	0.589
		number of neurons in the second layer	[40, 45, 50, 55, 60, 65, 70]	50	
		number of neurons in the third layer	[40, 45, 50, 55, 60, 65, 70]	50	
		activationr	[tanh, sigmoid, relu]	relu	

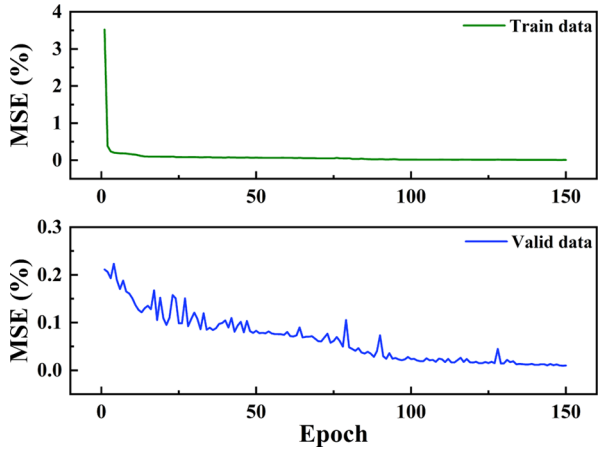


Figure 10. Relationship between DAO yield MSE value and iteration times in the LSTM model.

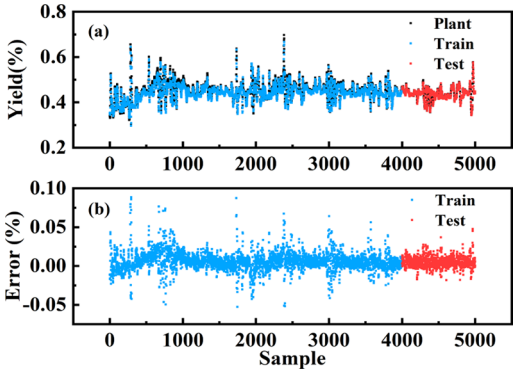


Figure 11. Comparison between the actual value of DAO yield in the LSTM model and the predicted value of the LSTM model. (a) Comparison of the actual value of DAO yield with the model value and (b) absolute error between the actual DAO yield value and the model value.

3.4.3. Comparison of Models Established by Different Methods. To investigate the performance of the LSTM model, the LSTM model was compared with the optimized CNN and MLP models. In CNN and MLP models, data were not divided into training and test sets according to time series after normalization, but were randomly sampled. The ratios of the training set and test set were consistent with that of the LSTM model. The network structure and parameters of CNN and MLP models were optimized by gridding, and the results are shown in Table 3. The performance of these six models on yield, CCR, S, A, R, and As of DAO are shown in Figure 12. The MLP model had the worst performance when evaluated by RMSE, MAPE, and MAE. For the yield prediction model of DAO, LSTM was obviously better than CNN. In the prediction of CCR, S, A, R, and As of DAO, LSTM was also better than CNN, but the performance of the two models was similar.

During the training process of the model, the weight matrix of the neural network obtained by each training was different and had a certain contingency. Therefore, the three models were trained and predicted 20 times. Using MAE as an evaluation function, the results of models were analyzed. Figure 13 shows the MAE distribution of the four models in 20 times. MAE of the MLP model are scattered and are generally

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

Table 3. Network Structure and Parameters of the Optimized Models of CNN and MLP

model	parameter	setting hyperparameters	selected hyperparameter
CNN	number of conv_filters	[16, 32, 64, 96]	64
	conv_kernel size	[1, 2, 3, 4]	2
	activation	[relu, tanh]	relu
	conv_strides	[1, 2, 3]	1
	pool_size	[1, 2]	2
	number of neurons in the fully connected network	[30, 40, 50, 60, 70]	60
MLP	solver	[adam, sgd, lbfgs]	lbfgs
	number of neurons in the first layer	[5, 10, 15, 20, 25, 30]	20
	number of neurons in the second layer	[5, 10, 15, 20, 25]	10
	alpha		1e-4
	activation	[tanh, sigmoid, relu]	relu

larger than models of LSTM and CNN in terms of yield, CCR, S, A, R, and As of DAO. Compared with CNN, the models established by LSTM for yield, CCR, S, A, R, and As of DAP had better performance.

3.4.4. Performance of the Multicondition LSTM Model based on SDAE-FCM. In Section 3.4.1, it has been proved that dimension reduction through SDAE can improve the accuracy of clustering, and combined with FCM clustering algorithm, SDA process data can be accurately divided into three operating modes. The sub-samples of each operation mode were further divided into training set, validation set, and test set. The network structure and parameters of each LSTM sub-model were optimized by the grid method. In the prediction stage, the established multicondition LSTM model is predicted by switch selection. The operation mode of input data was identified by SDAE-FCM, and the corresponding LSTM sub-model was selected to output the prediction results. Compared with the ordinary LSTM model, the multicondition LSTM model can output the corresponding prediction results according to the different operation modes. Figure 14 shows the performance comparison between the multicondition LSTM model and the ordinary LSTM model. Although the R^2 of both models is more than 0.9, the performance of the multicondition LSTM model is significantly better than that of the ordinary LSTM model. More SDA industrial process data were collected and used to further verify the predictive ability

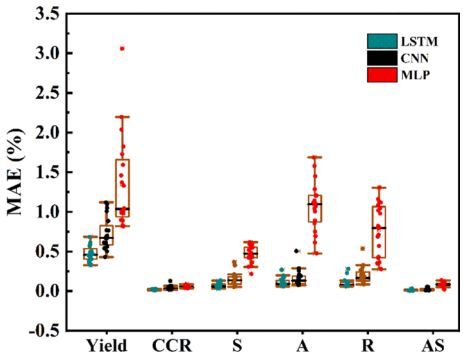


Figure 13. Comparison between the LSTM model, CNN, and the MLP model on MAE.

of the multicondition LSTM model. Figure 15 shows the prediction effect of the multicondition LSTM model on yield, CCR, S, A, R, and As of DAO. Prediction results indicate that the errors were very small, and R^2 were all greater than 0.95.

3.5. Sensitivity Analysis of Operating Variables. Through sensitivity analysis, the effects of some important raw material properties and operating conditions on the yield of the product were investigated, such as μ_{VR}, W_{CR}, W_S, T_{R1T}, and P_{R2T}. Figure 16 shows the influence of the properties of various raw materials and operating conditions on the yield. The results show that the DAO yield decreases with the increase of μ_{VR}, W_{CR}, W_S, nickel and vanadium content (W_{Ni+V}), and W_R, while the increase of W_{SH} and W_{AH} contributes to the DAO yield. In terms of operating conditions, the increase of T_{R1I}, T_{R1T}, T_{R2T}, T_{R2FRJ}, and R_A is beneficial to improve DAO yield, while the increase of T_{R2I}, P_{R2T}, and T_{R2B} will decrease DAO yield. In addition, μ_{VR}, W_{Ni+V}, and W_R of raw materials have great influence on DAO yield. The T_{R1I}, T_{R1T}, T_{R2I}, and R_A also have great influence on DAO yield. It can be seen that the effects of various raw material properties and operational variables predicted by the LSTM model on DAO yield all conform to the principle of two-phase countercurrent extraction in the SDA process.

3.6. Benchmarking of the Method in the Tennessee Eastman Process. The multicondition LSTM in this paper is benchmarked with Tennessee Eastman process data. Multi-condition LSTM is used to predict the future process measurement, and OC-SVM is used for fault detection. The model needs to be tested with a predefined fault margin

$$\text{margin level} = 2 \times \max(\text{anomaly}_{\text{count}})$$

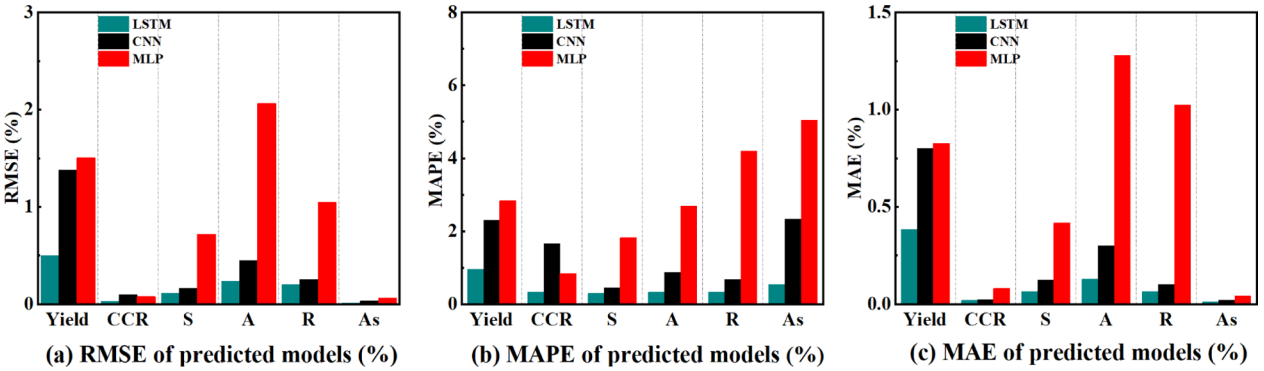


Figure 12. Performance comparison of CNN and MLP models with LSTM models at yield, CCR, S, A, R, and As of DAO. (a−c) Comparison of model evaluation of RMSE, MAPE, and MAE respectively.

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

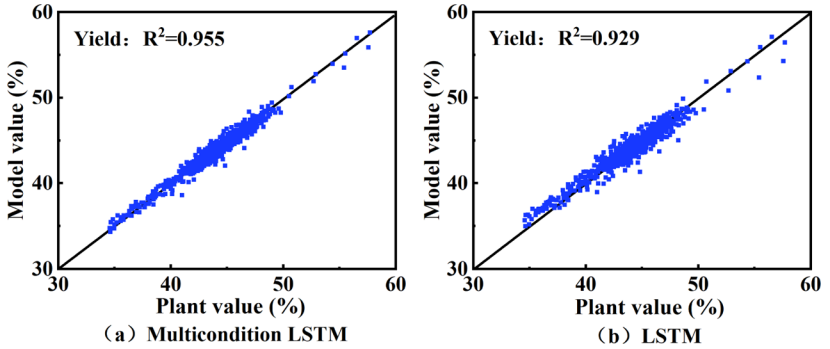


Figure 14. (a,b) Prediction accuracy of the multicondition LSTM model and the LSTM model at DAO yield.

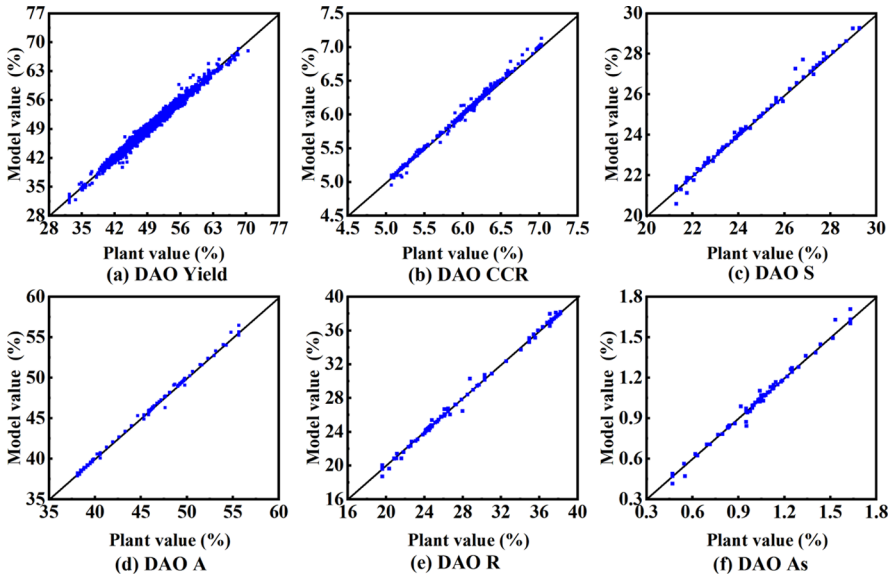


Figure 15. (a–f) Prediction effect of the multicondition LSTM model on yield, CCR, S, A, R, and As of DAO.

If the number of outliers exceeds the marginal level, the tested window is predicted to be in an abnormal condition. Otherwise, it is defined as in a normal operating condition. This paper compares multicondition LSTM OC-SVM with CNN-LSTM OC-SVM.[18] Among them, the sampling window size (window size is 100) and OC-SVM parameters of the two methods are consistent. Tennessee Eastman process faults are shown in Table 4, and the fault detection effect of multicondition LSTM OC-SVM is shown in Figure 17. The blue line in the figure is the fault margin. When the data anomaly points exceed the blue line, a fault is considered to have occurred. Compared with CNN-LSTM OC-SVM, multicondition LSTM OC-SVM is not easy to appear as a false alarm in the fault prediction of IDV4, IDV6, and IDV11 and can predict the fault earlier in the fault prediction of IDV1, IDV6, and IDV12. The reason may be that the division of working conditions makes the data more conducive to fault detection. The performance of multicondition LSTM OC-SVM in IDV5 fault prediction is not as good as that of CNN-LSTM OC-SVM. It may be that IDV5 is insensitive to changes in operating conditions. Compared with CNN-LSTM OC-SVM, multicondition LSTM OC-SVM has better performance in some fault prediction scenarios, including false alarms and earlier fault prediction, which can provide better guidance for the safety and effective control of the industrial process in time.

4. CONCLUSIONS

A time series model is applied to the SDA industrial process effectively. Before modeling, RFs is used to analyze the importance of each variable of SDA process data, and 21 variables with the greatest influence on yield are selected. On this basis, the time lag effect of operating conditions of residue and solvent mixing unit, extraction tower, and settling tower on DAO yield is analyzed, and the results indicate that the model needs to have the ability to deal with time lag data. The effects of different dimensionality reduction methods on clustering results are also investigated. The results show that SDAE has the ability of accurate dimensionality reduction. At the same time, the FCM clustering algorithm is combined to realize the operating modes division of the SDA process. LSTM is used to establish the model with the data of one of the operating modes on yield, CCR, S, A, R, and As of DAO for dealing with time lag relationship of the SDA process, and compare with models of CNN and MLP. In the process of comparison, the performance of each model is analyzed through MAE, MAPE, RMSE, and the stability of each model on yield, CCR, S, A, R, and As of DAO is further investigated through 20 training and prediction models. It is concluded that the LSTM model has the best predictive performance. Then, based on the results of SDAE-FCM, a multicondition LSTM model is established. Compared with the ordinary LSTM model, the prediction accuracy of multicondition LSTM is higher. Through more

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

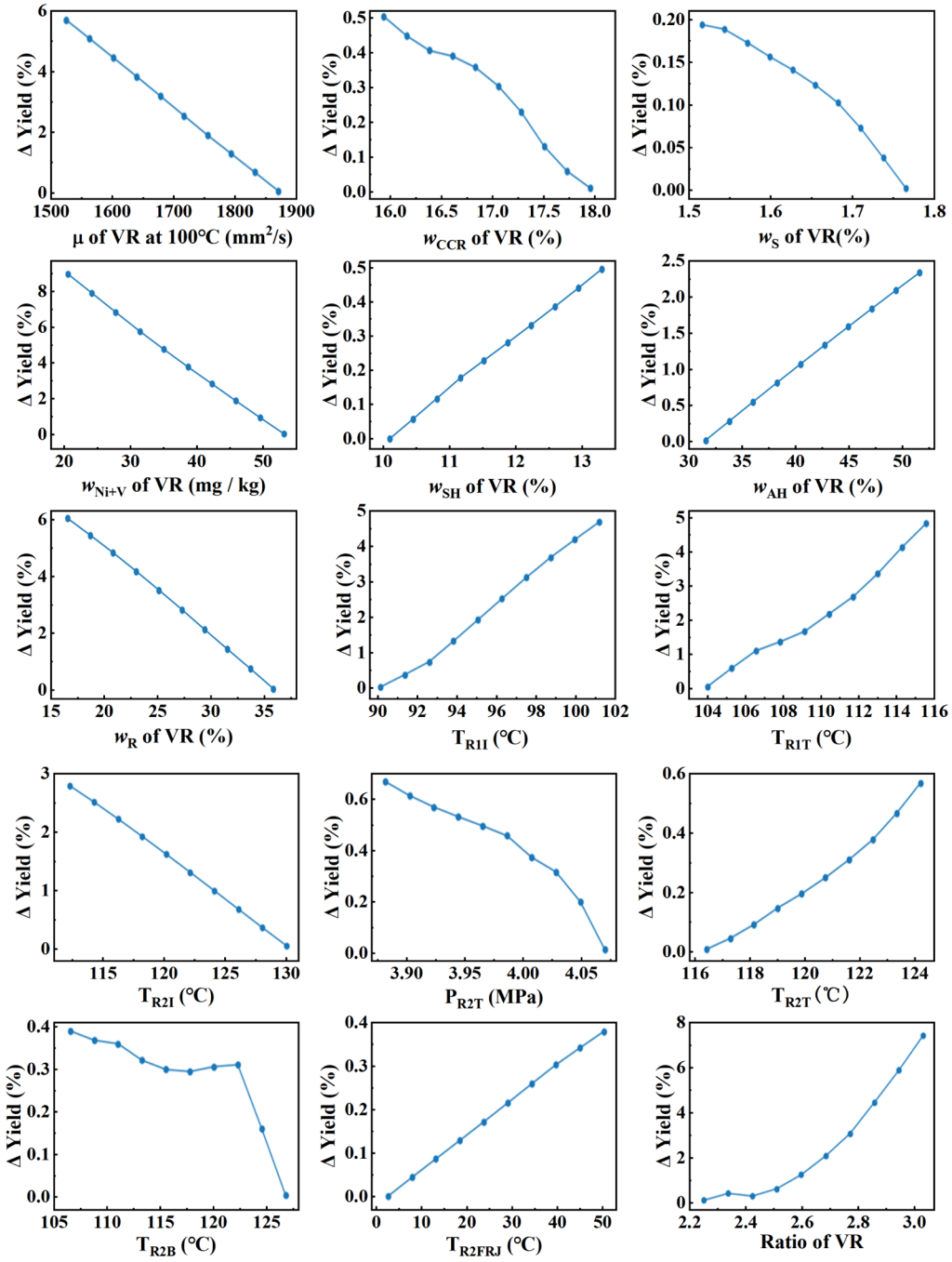


Figure 16. Effects of feed properties and operating conditions on DAO yield (ΔYield is the change value of DAO yield after changing the properties of raw materials or operating conditions).

Table 4. Tennessee Eastman Process Faults

fault ID	description	type
IDV1	A/C feed ratio, B composition constant (stream 4)	step
IDV4	reactor cooling water inlet temperature	step
IDV5	condenser cooling water inlet temperature	step
IDV6	A feed loss (stream 1)	step
IDV11	reactor cooling water inlet temperature	random variation
IDV12	condenser cooling water inlet temperature	random variation

SDA process data verification, the multicondition LSTM model still has a good prediction effect. Finally, the effects of the properties of raw materials and operating variables on the yield are analyzed by sensitivity analysis, and the results are consistent with the principle of two-phase countercurrent extraction in the SDA process. In this paper, the proposed methods are benchmarked with Tennessee Eastman process data. Through comparison, it is found that the proposed methods have better results in fault detection. Because the proposed methods can reduce the possibility of false alarms and predict the faults in advance, it can effectively prevent the system from being in fault condition and ensure production safety. Because the method in this paper can well solve the

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

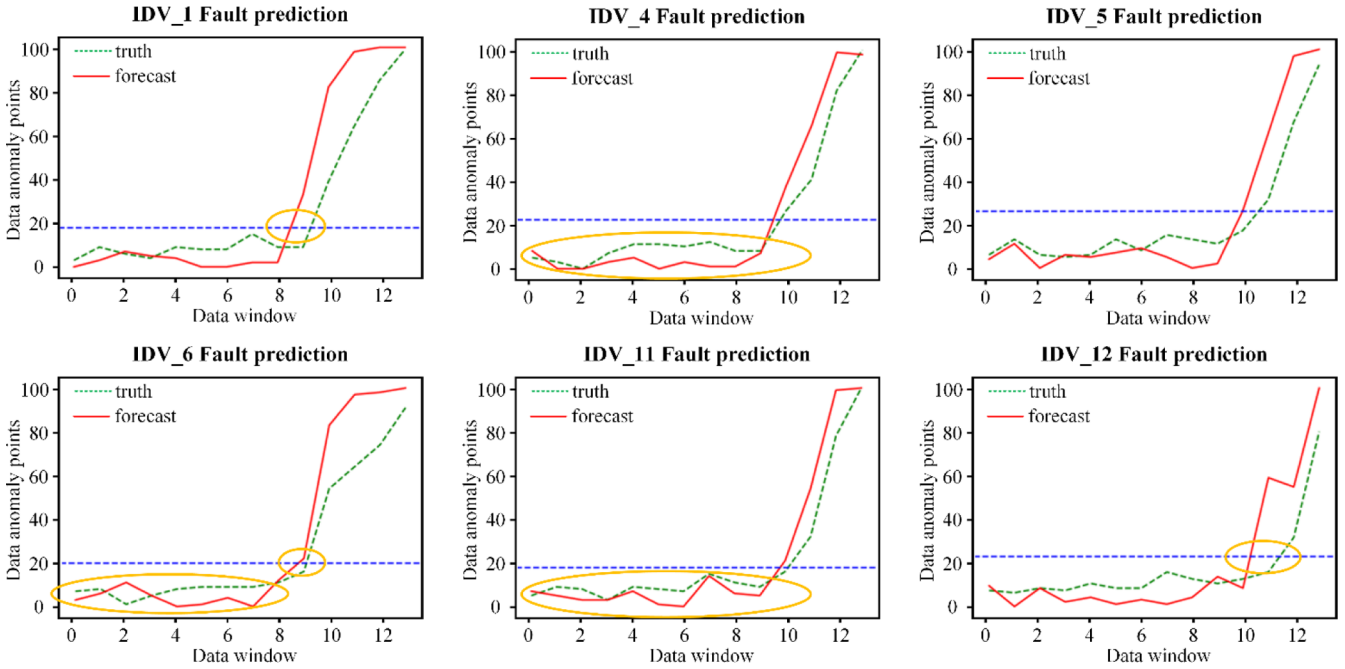


Figure 17. Test result of Tennessee Eastman early fault detection.

problems of the industrial process variable, the variable working conditions, and the time series of process data, this method can not only establish accurate yield prediction models in the application of the SDA process but can also predict the fault earlier and less false alarm in the fault detection of the Tennessee Eastman process.

AUTHOR INFORMATION

Corresponding Author

Minglei Yang − *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*; Email: mlyang@ecust.edu.cn

Authors

Jian Long − *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*; orcid.org/0000-0002-3221-309X

Yifan Chen − *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*

Dengke Cao − *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*

Pengyu Chen − *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c06624

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program—Intergovernmental International Science and Technology Innovation Cooperation Project (2021YFE0112800), National Natural Science Foundation of China (61973124, 62073142), Fundamental Research Funds for the Central Universities (222202317006), and Shanghai AI Lab.

NOMENCLATURE

c_j	the central value of clustering
u_{ij}	membership matrix
i	the sample
$\boldsymbol{i}_t^i$	input gate
$\boldsymbol{g}_t^i$	the update gate
$\boldsymbol{c}_t^i$	the cell state
$\boldsymbol{o}_t^i$	the output gate
$\boldsymbol{h}_t^i$	the output value
t	the input at time
σ	the element-wise sigmoid
tanh	the hyperbolic tangent functions
$\boldsymbol{W}_O$, $\boldsymbol{W}_f$, $\boldsymbol{W}_i$, $\boldsymbol{W}_g$	the input weight matrices
$\odot$	Hadamard product
$\boldsymbol{b}_f$, $\boldsymbol{b}_i$, $\boldsymbol{b}_g$, $\boldsymbol{b}_o$	the bias parameters
n	the number of test samples
Y_i	the actual value of the model
$\hat{Y}_i$	the predicted value of the model
$\overline{Y}_i$	the average value of the model
ρ_{VR}	density
μ_{VR}	kinematic viscosity
w_{VR}	carbon residue content
w_s	sulfur content
w_{Fe}	iron content
w_{Ni}	nickel content
w_V	vanadium content
w_{SH}	saturated hydrocarbons
w_{AH}	aromatic hydrocarbons

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

w_R	resin content
w_A	asphaltene content
F_{VR}	mass flow
T_{R1I}	inlet temperature of extraction tower
T_{R1T}	top temperature of extraction tower
T_{R1B}	bottom temperature of extraction tower
T_{R2I}	inlet temperature of settlement tower
P_{R2T}	top pressure of settlement tower
T_{R2T}	top temperature of settlement tower
T_{R2B}	bottom temperature of settlement tower
T_{R2FRJ}	diluent solvent temperature
R_A	the ratio of solvent to VR
w_{Ni+V}	nickel and vanadium content

■ ACRONYMS

A	aromatics
As	asphaltenes
CCR	Conradson carbon residue
DAO	deasphalted oil
DCS	distributed control system
DOA	deoiled asphalt
FCM	Fuzzy c-means
LIMS	laboratory information management system
LSTM	long short-term memory
MAE	mean absolute error
MAPE	mean absolute percentage error
MSE	mean square error
Ni	nickel
R	resins
R^2	*R*-square
RFs	Random Forests
RMSE	root mean square error
S	composition of saturates
SDA	solvent deasphalting
SDAE	stacked denoising autoencoder
V	vanadium
VR	vacuum residuum

■ REFERENCES

(1) Long, J.; Jiang, S.; Liu, T.; Wang, K.; He, R.; Zhao, L. Modified Hybrid Strategy Integrating Online Adjustable Oil Property Characterization and Data-Driven Robust Optimization under Uncertainty: Application in Gasoline Blending. *Energy Fuels* **2022**, *36*, 6581−6596.

(2) Long, J.; Li, T.; Yang, M.; Hu, G.; Zhong, W. Hybrid Strategy Integrating Variable Selection and a Neural Network for Fluid Catalytic Cracking Modeling. *Ind. Eng. Chem. Res.* **2019**, *58*, 247−258.

(3) Long, J.; Shen, B.; Ling, H.; Zhao, J.; Lu, J. Novel Solvent Deasphalting Process by Vacuum Residue Blending with Coal Tar. *Ind. Eng. Chem. Res.* **2011**, *50*, 11259−11269.

(4) Long, J.; Shen, B.; Ling, H.; Zhao, J. Nonconventional Vacuum Residue Upgrading Blended with Coal Tar by Solvent Deasphalting and Fluid Catalytic Cracking. *Ind. Eng. Chem. Res.* **2012**, *51*, 3058−3068.

(5) Magomedov, R. N.; Pripakhaylo, A. V.; Dzhumamukhamedov, D. S.; Maryutina, T. A. Solvent deasphalting of vacuum residue using carbon dioxide-toluene binary mixture. *J. CO2 Util.* **2020**, *40*, 101206.

(6) Zhou, L.; Chuang, Y.-C.; Hsu, S.-H.; Yao, Y.; Chen, T. Prediction and Uncertainty Propagation for Completion Time of Batch Processes Based on Data-Driven Modeling. *Ind. Eng. Chem. Res.* **2020**, *59*, 14374−14384.

(7) Wen, H.; Khan, F.; Amin, M. T.; Halim, S. Z. Myths and misconceptions of data-driven methods: Applications to process safety analysis. *Comput. Chem. Eng.* **2022**, *158*, 107639.

(8) Ren, G.; Chuttar, A.; Banerjee, D. Exploring efficacy of machine learning (artificial neural networks) for enhancing reliability of thermal energy storage platforms utilizing phase change materials. *Int. J. Heat Mass Transfer* **2022**, *189*, 122628.

(9) Zadakbar, O.; Imtiaz, S.; Khan, F. Dynamic Risk Assessment and Fault Detection Using Principal Component Analysis. *Ind. Eng. Chem. Res.* **2013**, *52*, 809−816.

(10) Zadakbar, O.; Imtiaz, S.; Khan, F. Dynamic risk assessment and fault detection using a multivariate technique. *Process Saf. Prog.* **2013**, *32*, 365−375.

(11) Hu, J.; Khan, F.; Zhang, L.; Tian, S. Data-driven early warning model for screenout scenarios in shale gas fracturing operation. *Comput. Chem. Eng.* **2020**, *143*, 107116.

(12) Kumari, P.; Wang, Q.; Khan, F.; Kwon, J. S.-I. A unified causation prediction model for aboveground onshore oil and refined product pipeline incidents using artificial neural network. *Chem. Eng. Res. Des.* **2022**, *187*, 529−540.

(13) Mallick, M. R.; Imtiaz, S. A. A Hybrid Method for Process Fault Detection and Diagnosis. *IFAC Proceedings Volumes* **2013**, *46*, 827−832.

(14) Gharahbagheri, H.; Imtiaz, S. A.; Khan, F. Root Cause Diagnosis of Process Fault Using KPCA and Bayesian Network. *Ind. Eng. Chem. Res.* **2017**, *56*, 2054−2070.

(15) Amin, M. T. An integrated methodology for fault detection, root cause diagnosis, and propagation pathway analysis in chemical process systems. *Clean Eng. Technol* **2021**, *4*, 100187.

(16) Kumari, P.; Bhadriraju, B.; Wang, Q.; Kwon, J. S.-I. A modified Bayesian network to handle cyclic loops in root cause diagnosis of process faults in the chemical process industry. *J. Process Control* **2022**, *110*, 84−98.

(17) Chaudhari, P.; Ade, N.; Pérez, L. M.; Kolis, S.; Mashuga, C. V. Minimum Ignition Energy (MIE) prediction models for ignition sensitive fuels using machine learning methods. *J. Loss Prev. Process Ind.* **2021**, *69*, 104343.

(18) Arunthavanathan, R.; Khan, F.; Ahmed, S.; Imtiaz, S. A deep learning model for process fault prognosis. *Process Saf. Environ. Prot.* **2021**, *154*, 467−479.

(19) Lee, J. M.; Shin, S.; Ahn, S.; Chun, J. H.; Lee, K. B.; Mun, S.; Jeon, S. G.; Na, J. G.; Nho, N. S. Separation of solvent and deasphalted oil for solvent deasphalting process. *Fuel Process. Technol.* **2014**, *119*, 204−210.

(20) Shin, S.; Lee, J. M.; Hwang, J. W.; Jung, H. W.; Nho, N. S.; Lee, K. B. Physical and rheological properties of deasphalted oil produced from solvent deasphalting. *Chem. Eng. J.* **2014**, *257*, 242−247.

(21) Ma, Y.; Gao, Z.; Shi, P.; Chen, M.; Wu, S.; Yang, C.; Wang, J.; Cheng, J.; Gong, J. Machine learning-based solubility prediction and methodology evaluation of active pharmaceutical ingredients in industrial crystallization. *Front. Chem. Sci. Eng.* **2022**, *16*, 523−535.

(22) Voronov, S.; Jung, D.; Frisk, E. A forest-based algorithm for selecting informative variables using Variable Depth Distribution. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104073.

(23) Wang, L.; Huang, Z.; Wang, R. Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. *Infrared Phys. Technol.* **2021**, *115*, 103731.

(24) Gao, S.; Zhang, Q.; Tian, R.; Ma, Z.; Dang, X. Horizontal Data Augmentation Strategy for Industrial Quality Prediction. *ACS Omega* **2022**, *7*, 30782−30793.

(25) Liu, K.; Zheng, M.; Liu, Y.; Yang, J.; Yao, Y. Deep Autoencoder Thermography for Defect Detection of Carbon Fiber Composites. *IEEE Trans. Ind. Inf.* **2022**, 1.

(26) Wang, K.; Xu, L.; Huang, L.; Wang, C.-D.; Lai, J.-H. SDDRS: Stacked Discriminative Denoising Auto-Encoder based Recommender System. *Cogn. Syst. Res.* **2019**, *55*, 164−174.

(27) Yu, J.; Liu, G. Extracting and inserting knowledge into stacked denoising auto-encoders. *Neural Netw* **2021**, *137*, 31−42.

(28) Kumar, N.; Kumar, H. A fuzzy clustering technique for enhancing the convergence performance by using improved Fuzzy c-means and Particle Swarm Optimization algorithms. *Data Knowl Eng* **2022**, *140*, 102050.

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450

(29) Wang, J.; Yang, Z.; Liu, X.; Li, B.; Yi, J.; Nie, F. Projected fuzzy C-means with probabilistic neighbors. *Inf. Sci.* **2022**, *607*, 553−571.
(30) Zhang, M.; Zhu, L.; Sun, Y.; Niu, D.; Liu, J. Computed tomography of ground glass nodule image based on fuzzy C-means clustering algorithm to predict invasion of pulmonary adenocarcinoma. *J. Radiat. Res. Appl. Sci.* **2022**, *15*, 152−158.
(31) Moharm, K.; Eltahan, M.; Elsaadany, E.Wind Speed Forecast using LSTM and Bi-LSTM Algorithms over Gabal El-Zayt Wind Farm. *2020 International Conference on Smart Grids and Energy Systems (SGES), 23-26 Nov. 2020*, 2020; pp 922−927.
(32) Swathi, T.; Kasiviswanath, N.; Rao, A. A. An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Appl. Intell.* **2022**, *52*, 13675.
(33) Yang, S. D.; Ali, Z. A.; Kwon, H.; Wong, B. M. Predicting Complex Erosion Profiles in Steam Distribution Headers with Convolutional and Recurrent Neural Networks. *Ind. Eng. Chem. Res.* **2022**, *61*, 8520−8529.
(34) Selvaggio, A. Z.; Sousa, F. M. M.; Silva, F. V. d.; Vianna, S. S. V. Application of long short-term memory recurrent neural networks for localisation of leak source using 3D computational fluid dynamics. *Process Saf. Environ. Prot.* **2022**, *159*, 757−767.
(35) Liu, Q.; Jia, M.; Gao, Z.; Xu, L.; Liu, Y. Correntropy long short term memory soft sensor for quality prediction in industrial polyethylene process. *Chemom. Intell. Lab. Syst.* **2022**, *231*, 104678.
(36) Chen, J.; Wang, X.; Xu, X. GC.-LSTM. graph convolution embedded LSTM for dynamic network link prediction. *Appl. Intell.* **2022**, *52*, 7513−7528.
(37) Agarwal, P.; Tamer, M.; Sahraei, M. H.; Budman, H. Deep Learning for Classification of Profit-Based Operating Regions in Industrial Processes. *Ind. Eng. Chem. Res.* **2020**, *59*, 2378−2395.
(38) Zheng, J.; Ma, L.; Wu, Y.; Ye, L.; Shen, F. Nonlinear Dynamic Soft Sensor Development with a Supervised Hybrid CNN-LSTM Network for Industrial Processes. *ACS Omega* **2022**, *7*, 16653−16664.
(39) Rai, H. M.; Chatterjee, K. Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. *Appl. Intell.* **2022**, *52*, 5366−5384.
(40) Jing, Y.; Zhang, L.; Hao, W.; Huang, L. Numerical study of a CNN-based model for regional wave prediction. *Ocean Eng* **2022**, *255*, 111400.
(41) Yerkenov, T.; Tazikeh, S.; Tatar, A.; Shafiei, A. Asphaltene Precipitation Prediction during Bitumen Recovery: Experimental Approach versus Population Balance and Connectionist Models. *ACS Omega* **2022**, *7*, 33123−33137.
(42) Ke, K.-C.; Huang, M.-S. Enhancement of multilayer perceptron model training accuracy through the optimization of hyperparameters: a case study of the quality prediction of injection-molded parts. *Int J Adv Manuf Technol* **2022**, *118*, 2247−2263.
(43) Carneiro, T. C.; Rocha, P. A. C.; Carvalho, P. C. M.; Fernández-Ramírez, L. M. Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. *Appl. Energy* **2022**, *314*, 118936.

https://doi.org/10.1021/acsomega.2c06624
ACS Omega 2023, 8, 5437−5450