

Fault Diagnosis of the Dynamic Chemical Process Based on the Optimized CNN-LSTM Network

Honghua Chen, Jian Cen,* Zhuohong Yang, Weiwei Si, and Hongchao Cheng



Cite This: *ACS Omega* 2022, 7, 34389–34400



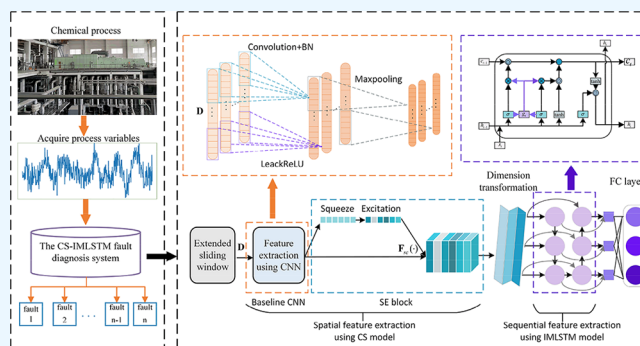
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Deep learning provides new ideas for chemical process fault diagnosis, reducing potential risks and ensuring safe process operation in recent years. To address the problem that existing methods have difficulty extracting the dynamic fault features of a chemical process, a fusion model (CS-IMLSTM) based on a convolutional neural network (CNN), squeeze-and-excitation (SE) attention mechanism, and improved long short-term memory network (IMLSTM) is developed for chemical process fault diagnosis in this paper. First, an extended sliding window is utilized to transform data into augmented dynamic data to enhance the dynamic features. Second, the SE is utilized to optimize the key fault features of augmented dynamic data extracted by CNN. Then, IMLSTM is used to balance fault information and further mine the dynamic features of time series data. Finally, the feasibility of the proposed method is verified in the Tennessee-Eastman process (TEP). The average accuracies of this method in two subdata sets of TEP are 98.29% and 97.74%, respectively. Compared with the traditional CNN-LSTM model, the proposed method improves the average accuracies by 5.18% and 2.10%, respectively. Experimental results confirm that the method developed in this paper is suitable for chemical process fault diagnosis.



1. INTRODUCTION

Chemical processes play a pivotal role in the development of the world economy and in the lives of people. New technologies, new equipment, and new materials are emerging, production scales are expanding, processes are becoming more complex, and operating environments are harsh, resulting in chemical process risks everywhere. Once a safety accident occurs, it will bring serious damage to people's lives and health, the ecological environment, social stability, and the enterprise economy.

Abnormal situation management (ASM) provides an early warning for abnormal situations, timely diagnoses the causes, and provides decision support for technicians to take measures and restore the process to normal, which has made a great contribution to improving process safety.¹ Proper risk assessment (RA) helps to control the risks before they occur. Fault detection and fault diagnosis (FDD) means detecting whether faults have occurred and, if so, classifying the fault. From a process safety perspective, FDD, RA, and ASM can form a closed loop. Among them, FDD is a key step to identify potential risks. RA evaluates the risk margin based on fault information provided by FDD and reports risk events to ASM. ASM makes decisions to ensure process security based on the feedback.² Khan et al.³ pointed out that process security could be improved by integrating dynamic FDD with RA. Dai et al.⁴ proposed that FDD is an effective way to control and mitigate process risks.

From the perspective of risk engineering, process safety usability is to effectively detect and diagnose faults.⁵ Therefore, the development of an intelligent and efficient FDD system is the key to maintain the ideal performance of digital industrial processes and safety in production.

The integration of FDD with process safety and risk assessment is an interesting research area. Amin et al.⁶ proposed a risk-based FDD method. This method carried out fault detection and diagnosis in the monitored risk profile. Experimental results showed that this method has better diagnostic performance than PCA and transfer entropy. Bao et al.⁷ proposed a risk-based process safety fault diagnosis technology. The advantage of this method is to identify and determine potential faults by risk index and realize the development of fault diagnosis technology from single variable to multivariable monitoring. Bhadriraju et al.⁸ designed an operational adaptive sparse system recognition to solve the

Received: June 27, 2022

Accepted: August 26, 2022

Published: September 12, 2022



problem that an offline training model has difficulty capturing the dynamic behavior of a process. The experimental result showed that the process behavior prediction based on this system can effectively predict faults and assess risks. From the perspective of process safety, it is again clear that FDD is an effective initiative to minimize risk and guarantee the safe operation of complex industrial processes.

Data-driven FDD methods can avoid the dependence on a complex process mechanism and mine the high-value information hidden in process data. Data driven methods can be further divided into multivariate statistical methods, machine learning methods, and deep learning methods. Multivariate statistical methods include principal component analysis (PCA), independent component analysis (ICA), partial least-squares (PLS), a Gaussian mixture model (GMM), and their variants.⁹ At present, multivariate statistical methods are mainly used for fault detection. Deng et al.¹⁰ proposed two local kernel PCA (KPCA) to solve the problem of missing local data information on KPCA in the case of an early fault. The method has been verified effectively in a continuous stirred tank reactor (CSTR). For fault detection of complex processes, a single method is usually not as superior as the detection results obtained by hybrid methods.⁴ Han et al.¹¹ adopted the hybrid fault detection method of adaptive kernel PCA and gray correlation analysis, which is superior to single kernel PCA and can provide a basis for ASM. Fault detection based on multivariate statistical learning methods usually depend on the threshold value of calculation to judge whether the fault exists. If the threshold value calculation slightly deviates, it may lead to the wrong result of fault diagnosis and increase the process risk.

Machine learning methods include locality preserving projection (LPP), naive bayes (NB), and support vector machine (SVM). He et al.¹² proposed a new discriminant LPP algorithm (DLPP) combined with Monte Carlo sampling, which not only solves the problem of high-dimensional process data but also solves the problem that DLPP is limited by a small sample size, thus effectively improving the fault diagnosis performance of industrial processes. Zhang et al.¹³ proposed an improved LPP and AdaBoost integration method. The improved LPP based on the heat-kernel and cosine weights can effectively extract the internal structure feature of data, so high fault diagnosis accuracy can be achieved in two chemical processes. Zhang et al.¹⁴ constructed a new farthest–nearest distance neighborhood and locality projections method and used it to reduce the dimension of high-dimensional process data to extract discriminant features. NB was adopted as a classifier for process fault diagnosis. Amin and Khan et al.¹⁵ proposed a hybrid diagnosis method of PCA and BN. This method achieved good diagnostic performance in a continuous stirred tank heater and binary distillation column because it used the correlation dimension to select the principal component and combined a vine copula and the BN theorem to capture the nonlinear dependence of high dimensional process data. Deng et al.¹⁶ proposed a fault detection method based on the integration of the spatial compression matrix and NB, which reduces the complexity of learning and helps to speed up the management of production risks. Machine learning can achieve a better effect of FDD when a small sample is used. However, the fault diagnosis ability of these methods depends on the quality of feature extraction, which has certain limitations on the dynamic feature representation of process data.

Deep learning is widely regarded as an effective tool for fault diagnosis in modern industrial applications. Diagnostic models

based on classical deep learning include convolutional neural network (CNN), deep belief network (DBN), stacked autoencoder (SAE), and long short-term memory network (LSTM). Among them, CNN has achieved a more advanced performance. In 2018, Wu et al.¹⁷ used deep CNN (DCNN) for fault diagnosis of TEP. The time-frequency domain features of process variables are converted into two-dimensional (2D) matrices, which are input into DCNN to extract spatial features of variables, and then fault classification is carried out. This method has achieved 88.2% classification accuracy. Song et al.¹⁸ used matrix maps and multiscale CNN for chemical process fault diagnosis, and the classification accuracy is 88.54%. For the above models, process variables need to be converted into 2D matrices or complex images as inputs of CNN, which leads to a large consumption of computing resources. Yu et al.¹⁹ designed a multichannel 1D CNN model (MC1-DCNN) on the basis of a wavelet transform and applied it to batch-fed fermentation of penicillin and TEP. The result shows that MC1-DCNN has the ability to learn high-dimensional process signal characteristics and a good performance of fault diagnosis. Yu et al.²⁰ designed a broad CNN with incremental learning ability, which is characterized by self-renewal in the face of new faults.

In addition, LSTM has also attracted scholars' attention in industrial process fault diagnosis as a result of its stronger adaptability in time series data analysis. Zhao et al.²¹ came up with an end-to-end sequential fault diagnosis method based on LSTM to address the problem that most conventional fault diagnosis techniques cannot learn dynamic features from raw data. Han et al.²² presented an optimized LSTM, which improves the accuracy of diagnosis of single and multiple faults in TEP by optimizing the number of hidden layer nodes of the LSTM. Park et al.²³ proposed an integration method of convolutional LSTM and autoencoder to detect rare faults in industrial processes. Gravanis et al.²⁴ combined the feature reduction method with LSTM and a time delay network, respectively, to conduct FDD of nonlinear processes. The fault diagnosis model based on the combination of CNN and LSTM has become a research hotspot because of its ability to extract spatial and temporal features of industrial data and improve diagnosis performance. Shao et al.²⁵ used a multichannel LSTM-CNN (MCLSTM-CNN) fault diagnosis model. This method inputs a data set into LSTM and then uses multiple parallel convolution layers to mine the output features of the hidden layer at the same time. The research indicates that the fault diagnosis accuracy of applying MCLSTM-CNN to TEP is as high as 92.06%. Wang et al.²⁶ designed the feature extraction method of the LSTM-CNN parallel structure and then fused and compressed the features by MLP. This method can extract the temporal and spatial features of process variables, so as to improve the diagnostic performance. Yuan et al.²⁷ used a chemical process monitoring and fault diagnosis scheme based on multiscale CNN-LSTM, with the purpose of mining high-dimensional fault features in a multiscale and hierarchical manner. Huang et al.²⁸ transformed the process data into two-dimensional data and input it into CNN-LSTM to extract the spatial and delay characteristics of the data. This method improved the diagnostic accuracy and noise sensitivity. However, the following problems still exist in industrial process fault diagnosis based on CNN, LSTM, or CNN-LSTM:

- (1) CNN in the above literature is only a series of convolution layers, connecting features in the channel dimension. However, fault data is usually composed of many variables

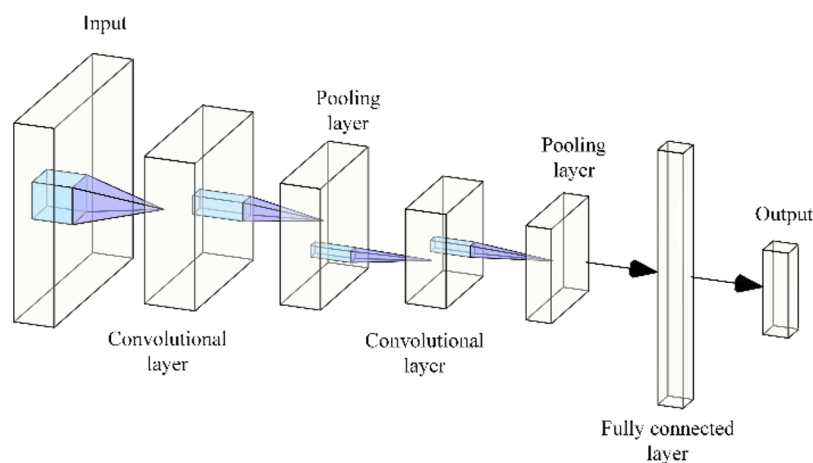


Figure 1. Basic structure of the CNN.

collected by many sensors in an industrial process. Each variable provides a different degree of importance of distinguishing features for fault diagnosis. Therefore, the above research methods lack the proper mechanism to reflect the correlation and importance of fault dynamic characteristics between different channels.

- (2) The above methods used LSTM to extract time series characteristics of chemical process data. However, there will be the problem of an unbalanced distribution of fault dynamic information because of the special gating mechanism. Therefore, the fault information on time series data cannot be extracted efficiently.

The problems mentioned above can make it difficult for the traditional CNN-LSTM network and other forms of this network to extract dynamic fault features of chemical process dynamic data. Therefore, this paper designs a model which combines CNN, the SE attention mechanism, and improved LSTM (CS-IMLSTM) for the fault diagnosis of TEP. First, the time series of industrial process variables contains the dynamic evolution process of the faults. Therefore, to enhance the dynamic characteristics of sequential fault data, extended sliding window preprocessing technology is proposed to obtain the augmented dynamic data, which provide sufficient fault diagnosis information for the proposed model. Second, aiming at the problem that a single network CNN cannot automatically select important channel features, a network architecture combining the CNN and SE attention mechanism (CS) is proposed, which makes the proposed model give more weight to critical channel fault features and reduce attention to redundant features. Finally, an improved LSTM is proposed to optimize the gating mechanism of original LSTM and balance the characteristic information on the industrial process in the time dimension. It is helpful for the proposed model to further mine dynamic information on industrial process fault data. The proposed method can not only adaptively extract dynamic fault features, weighting the features of different channels, but also balance fault information. Cascaded CS-IMLSTM can simultaneously extract the spatial and temporal dynamic features of process data, so as to enhance the capabilities of industrial process fault diagnosis. In terms of process safety, the proposed method can minimize the risk of industrial process operation and improve the safety of chemical process production.

2. RELATED THEORIES

2.1. Convolutional Neural Network. CNN has received extensive attention in the field of industrial fault diagnosis. CNN adaptively learns the spatial features of data by back-propagating using multiple blocks such as the convolution layer and pooling layer.²⁹ The basic structure is shown in Figure 1.

The convolution layer is mainly used to mine the local features of input. The mathematical expression for the convolution layer is

$$x^i = f\left(\sum w^i \times x^{(i-1)} + b^i\right) \quad (1)$$

where i represents layer i of the network and x^i denotes the output of feature data at the layer i . Similarly, x^{i-1} is the input data of layer i . b^i denotes the bias of layer i , and ω^i is the convolution kernel. $f(\cdot)$ denotes activation function. LeakyReLU is used as the activation function, and its mathematical model can be represented as

$$x^j = \begin{cases} x^j, & x^j > 0 \\ \alpha x^j, & x^j \leq 0 \end{cases} \quad (2)$$

Here α is the fixed parameter, and $\alpha = 0.01$.

The pooling layer can reduce data redundancy, preserving the key elements of the feature map and controlling overfitting. The mathematical model of pool operation can be expressed as

$$x^{m',n'} = pool(x^{m,n}) \quad (3)$$

where $x^{m,n}$ and $x^{m',n'}$ respectively represent the values before and after the pooling operation of the point (m, n) in the output feature graph of the convolution layer.

The fully connected (FC) layer map features are extracted from the convolutional layer and down-sampled by the pooling layer to the sample label space. For specific information about CNN, please refer to the literature.³⁰

2.2. Squeeze-and-Excitation Attention Mechanism. Recently, the benefits of attention mechanisms have been demonstrated in a variety of tasks. However, the advantage of the attention mechanism in chemical process fault diagnosis has not been fully exploited. Therefore, this paper uses an attention mechanism to mine important features of fault data. The emergence of the SE block is to work out a loss problem caused by different proportions of feature map channels in convolution operation and improve the depth representation ability of the

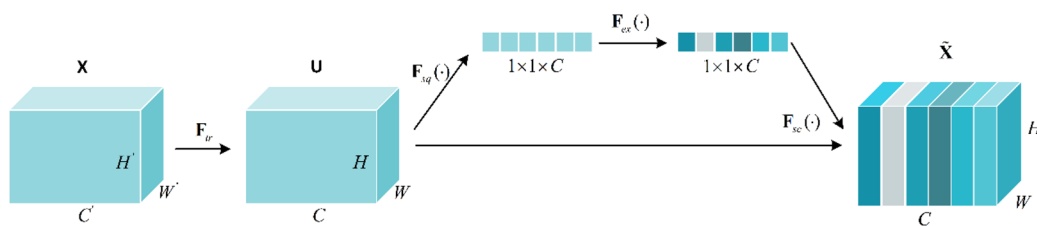


Figure 2. Basic structure of the SE block.

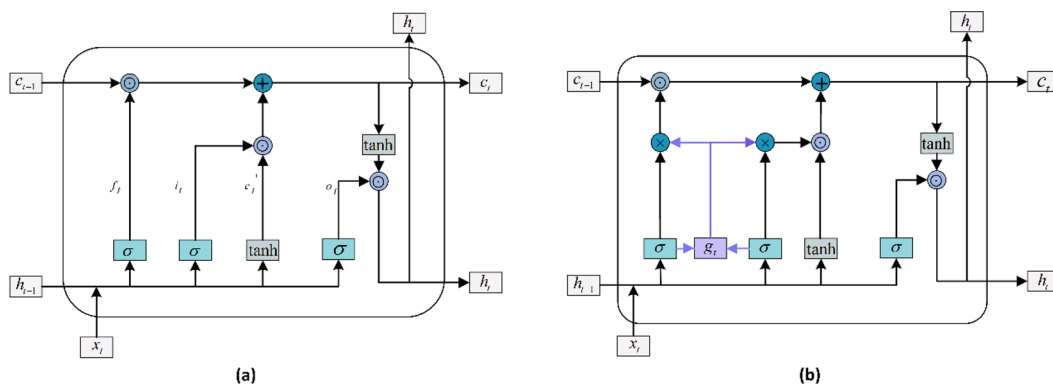


Figure 3. Structures of the (a) LSTM and (b) IMLSTM.

CNN. The SE block can model dynamic nonlinear dependencies between channels using global information learned by the CNN. Thus, it can enhance feature information that is effective for fault classification and suppress the ineffective feature information. The structure of the SE block is presented in Figure 2. For the detailed process of the SE block, the reader is referred to ref 31, and its brief process is as follows:

Given transformation F_{tr} , let $\mathbf{U} = F_{tr}(\mathbf{X})$, $\mathbf{X} \in \mathbb{R}^{H' \times W' \times C'}$, $\mathbf{U} \in \mathbb{R}^{H \times W \times C}$. Assuming that F_{tr} is a convolution operator, the feature map of \mathbf{X} can be expressed as $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_C]$, and

$$\mathbf{u}_c = \mathbf{k}_c * \mathbf{X} \quad (4)$$

where $\mathbf{u}_c \in \mathbb{R}^{H \times W}$ and $\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_C]$ represents the learned filter kernels. * refers to the convolution operation.

In the squeeze operation, the spatial dimension $H \times W$ of \mathbf{U} is compressed by global average pooling to obtain the channel statistic $\mathbf{z} \in \mathbb{R}^C$. The c th element in \mathbf{z} can be expressed as

$$z_c = F_{sq}(\mathbf{u}_c) \quad (5)$$

In the excitation operation, a gating mechanism with sigmoid function is utilized to obtain the dependencies between channels. The operation can be expressed as

$$\mathbf{s} = F_{ex}(\mathbf{z}, \mathbf{W}) \quad (6)$$

Here, $\mathbf{s} \in \mathbb{R}^C$ is the channel weight. \mathbf{W} is a parameter that needs to be learned. The final output $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_C]$ of the SE block is generated by the scaling operation

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{sc}(\mathbf{u}_c, s_c) = s_c \mathbf{u}_c \quad (7)$$

Here, $\tilde{\mathbf{X}} \in \mathbb{R}^{H \times W \times C}$, $\mathbf{F}_{sc}(\mathbf{u}_c, s_c)$ is channel-wise between scalar s_c and the feature map $\mathbf{u}_c \in \mathbb{R}^{H \times W}$.

2.3. Improved Long Short-Term Memory Network. Hochreiter et al.³² proposed LSTM, which can maintain the nondispersion of a data gradient over a long time span. LSTM

has recently been successful in various areas of sequence modeling, including but not limited to speech recognition and machine translation. The basic structure of LSTM is presented in Figure 3a. Key elements in the LSTM layer include input gate i_t , forget gate f_t , output gate o_t , and internal memory cell c_t . Moreover, each logic gate has its own parameters (U, W, b), so that information can be filtered at the corresponding position, the weight of useful information can be enhanced, and redundant information can be effectively filtered.

- (1) The forget gate f_t is expressed by the following equation:

$$f_t = \delta(W_f x_t + U_f h_{t-1} + b_f) \quad (8)$$

where $\delta(\cdot)$ is a sigmoid function and $0 < f_t < 1$.

- (2) The input gate i_t can be expressed by the following equation:

$$i_t = \delta(W_i x_t + U_i h_{t-1} + b_i) \quad (9)$$

$$c'_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (10)$$

where $\tanh(\cdot)$ denotes the hyperbolic tangent activation function and $0 < i_t < 1$.

- (3) The internal memory cell c_t is expressed by the following equation:

$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t \quad (11)$$

- (4) The output gate o_t can be expressed by the following equation:

$$o_t = \delta(W_o x_t + U_o h_{t-1} + b_o) \quad (12)$$

$$h_t = o_t \odot \tanh(c_t) \quad (13)$$

The expression formula of the improved internal memory cell c_t is as follows:

$$g_t = 1/(f_t + i_t) \quad (14)$$

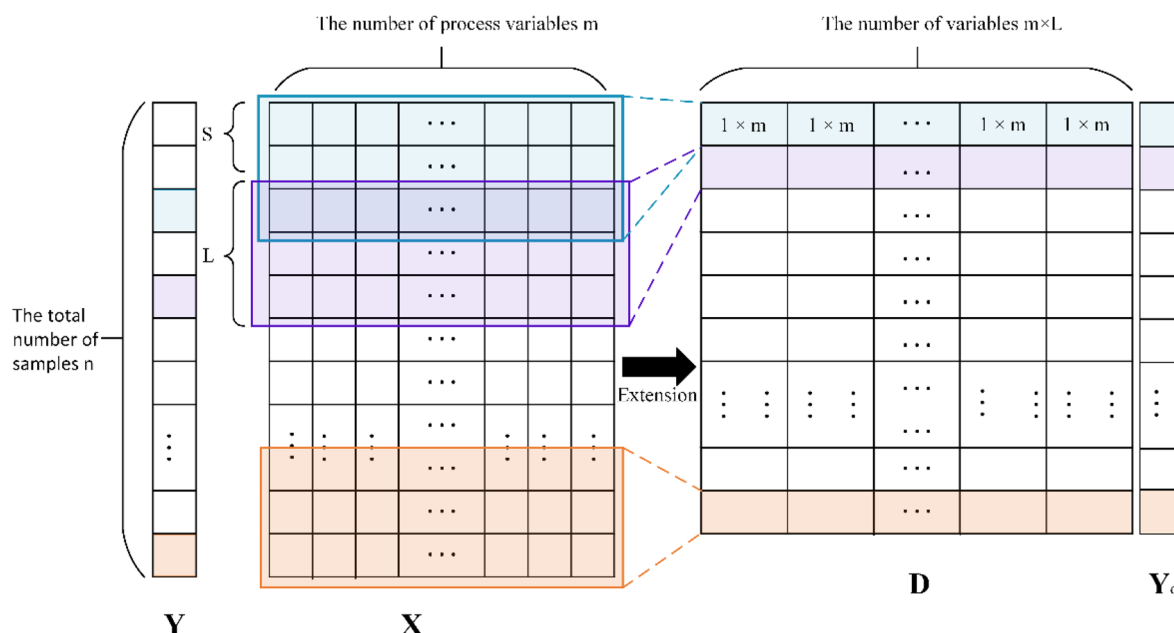


Figure 4. Extended sliding window mechanism schematic.

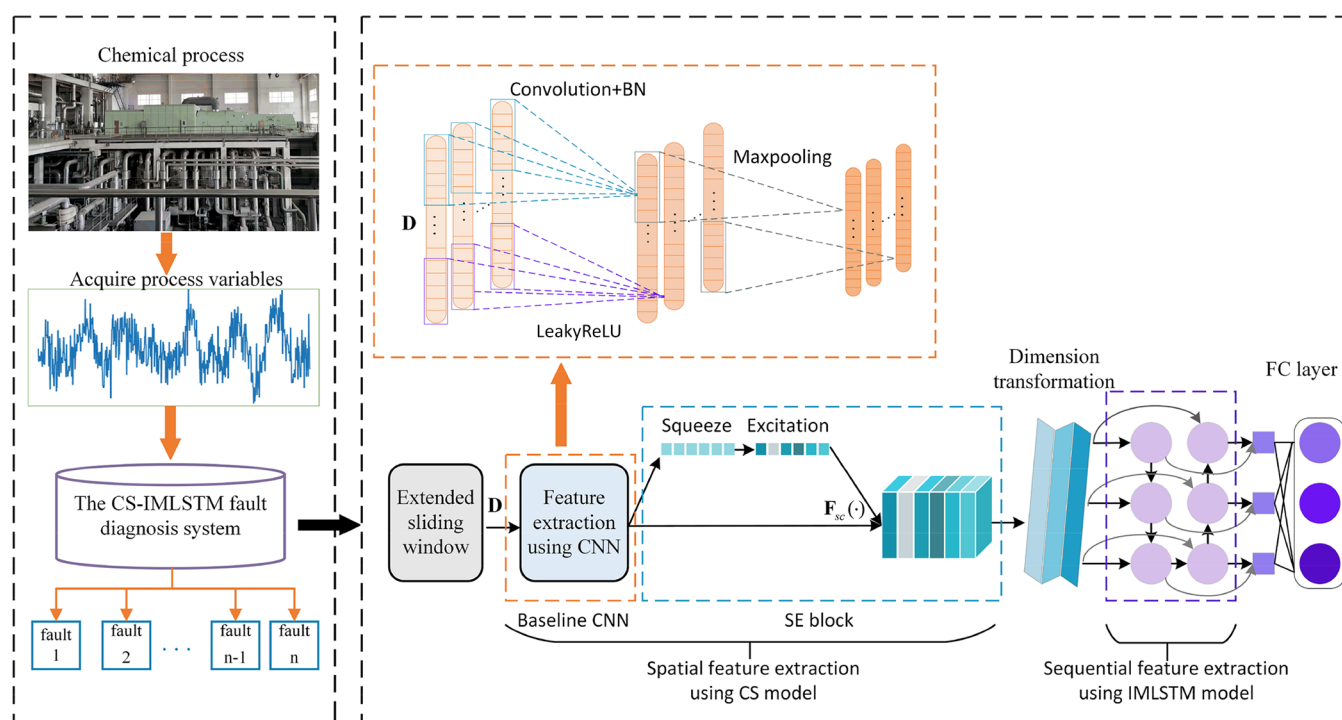


Figure 5. Fault diagnosis flow based on the CS-IMLSTM model.

$$c_t = f_t g_t \odot c_{t-1} + i_t g_t \odot c'_t \quad (15)$$

It can be seen from eqs 8–13 that the forgetting gate and input gate of LSTM are independent of each other. However, the values of f_t and i_t respectively determine the degree of retention for the previous moment internal memory cell c_{t-1} and current moment memory cell in eq 11.²² This also means that when it is applied to complex chemical process fault diagnosis, the internal memory cell c_t at the current moment will excessively rely on c_{t-1} or c'_t if f_t or i_t approaches 1, which will lead to the problem of unbalanced fault features of the chemical process.³³

The internal memory cell c_t of LSTM is improved to solve the above problem. The structure of IMLSTM is shown in Figure 3b, where improved c_t is shown in eqs 14 and 15. The introduction of g_t in c_t makes the degree of information retention in c_{t-1} dependent on $f_t/(f_t + i_t)$ and not only on f_t . Similarly, the degree of information retention in c'_t depends on $i_t/(f_t + i_t)$. By balancing the information on forgetting and input gates, IMLSTM can process the dynamic features of temporal data more efficiently.

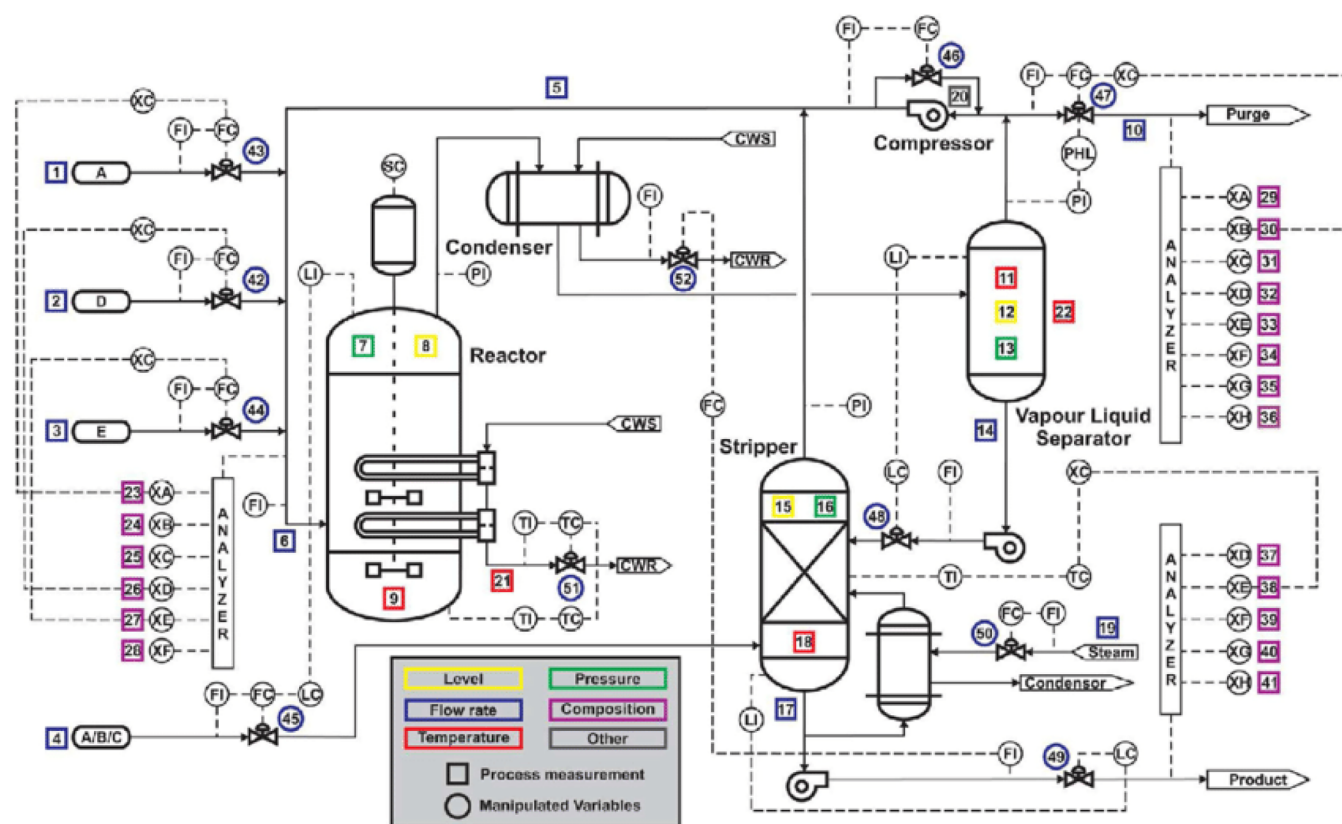


Figure 6. Flowchart of TEP.³⁶ Reprinted with permission from ref 36. Copyright 2019 Elsevier.

3. PROPOSED METHOD

3.1. Data Preprocessing. Data collected by industrial processes are usually dynamic; that is, faults occurring at the current moment may depend on changes in system state at the previous moment.²⁸ It is difficult to describe the change characteristics of industrial processes accurately by establishing a single global diagnostic model. In this paper, an extended sliding window mechanism is introduced to transform raw data into augmented dynamic data. The whole process is transformed into a time-varying dynamic process, and a local model is established. With the continuous change of the process, the model needs to be constantly updated to adapt to this change, which can be more accurate in the analysis of new samples and is more conducive to the proposed model to mine the dynamic feature information on time series data. The principle of the extended sliding window mechanism is shown in Figure 4. Formula 16 represents raw data set X and its corresponding labels Y , where n and m respectively refer to the number of observed samples and variables and $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tm})$ denotes observed variables collected by industrial process at moment t . Let the sliding step of the sliding window be S ($S \in \mathbb{N}^*$ and $S \leq L$), and L is the length of the sliding window. As shown in Formula 17, dynamic data D and corresponding label Y_d can be obtained by extended sliding window operation on the raw data set, which is the input of the proposed model.

After the extended sliding window operation, the raw data set is transformed into an augmented dynamic data set, which allows the proposed model to be able to capture the features of small changes in observed variables and learn dynamic information. Thus, the performance of fault diagnosis in an industrial process can be greatly improved.

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ \vdots & \vdots & \dots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tm} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & x_{n2} & \dots & a_{nm} \end{pmatrix} = (\mathbf{x}_1 \dots \mathbf{x}_t \dots \mathbf{x}_n)^T \quad \text{and} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_t \\ \vdots \\ y_n \end{pmatrix} \quad (16)$$

$$D = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_L \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{x}_t & \mathbf{x}_{t+1} & \dots & \mathbf{x}_{t+L-1} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{x}_{n-L+1} & \mathbf{x}_{n-L+2} & \dots & \mathbf{x}_n \end{pmatrix} \quad \text{and} \quad Y_d = \begin{pmatrix} y_L \\ \vdots \\ y_{t+L-1} \\ \vdots \\ y_n \end{pmatrix} \quad (17)$$

3.2. Diagnostic Process of the Proposed Method. In recent years, CNN, LSTM, CNN-LSTM, and their variants have been widely used in the field of fault diagnosis, but these deep learning methods have difficulty capturing the dynamic characteristics of dynamic data in the process industry. In this paper, we aim to build a diagnostic model of dynamic chemical processes based on an optimized CNN-LSTM (CS-IMLSTM) network. The fault diagnosis flowchart based on CS-IMLSTM is shown in Figure 5. The convolution layer extracts spatial features of data. The batch normalization (BN) layer improves the training speed and mitigates the risk of overfitting. LeakyReLU increases the network sparsity. The pooling layer reduces the number of model parameters and optimizes the workload. The SE block weights important channel features. The IMLSTM balances the fault information and extracts the temporal dynamic features of the data. The FC layer bridges all features

Table 1. Fault Modes of Case 1 and Case 2

case	fault	fault cause	fault type
Case 1	1	A/C feed ratio fluctuates, B feed is stable	Step
	2	B feed fluctuates, A/C feed ratio is stable	Step
	6	A material leak	Step
	7	Feed C inlet pressure loss: availability reduction	Step
	8	A, B, C feed composition fluctuation	Random variable
Case 2	4	Temperature disturbance at reactor cooling water inlet	Step
	5	Temperature disturbance at reactor cooling water inlet	Step
	10	C feed temperature disturbance	Random variable
	11	Inlet temperature fluctuation of reactor cooling water	Random variable
	12	Inlet temperature fluctuation of condenser cooling water	Random variable

and feeds the output values into classifiers for classification. CS-IMLSTM is an effective improvement of CNN-LSTM. It is worth noting that the proposed method uses CS-IMLSTM combined with the extended sliding window mechanism, which can not only automatically extract spatial and temporal features from the original industrial data but also perceive the deep dynamic information, so as to realize the identification of different fault types, optimize decision-making for risk assessment and ASM, and help the process run safely and steadily for a long period. The proposed method consists of the following five core steps:

- (1) Industrial process fault data and corresponding labels are collected.
- (2) The extended sliding window mechanism is used to generate augmented dynamic data by setting the sliding step S and sliding window length L .
- (3) The training set and corresponding label are fed into the CS-IMLSTM network. CS is used to extract spatial features of data and enhance critical fault features. The spatial feature vector of the data is transformed and input to IMLSTM. IMLSTM is used to balance the fault information and further extract dynamic features of augmented industrial data.
- (4) The extracted features are input into the classifier for fault classification, and the trained model is saved.
- (5) After extended sliding window processing, the test data and corresponding label are input to the trained model to prove the efficiency of the model.

4. EXPERIMENTAL VERIFICATION

4.1. Introduction to the Tennessee-Eastman Process.

TEP is a simulation process developed by Eastman Chemical Company,³⁴ which is basically the same as the actual production process. Therefore, the TEP is often taken as a simulation example to assess the feasibility of fault detection and diagnosis methods for industrial processes. The flowchart of the TEP is displayed in Figure 6. The TEP mainly consists of five operation units: reactor, condenser, gas–liquid separator, vapor extraction tower, and circulating compressor. The chemical reactions occurring in the TEP involve a total of eight components, where the reactants include gaseous substances A, C, D, and E and inert catalyst B and the products include liquid products G and H and byproduct F.

There are 52 variables in the overall process, including 11 control variables, 19 component measurement variables, and 22 continuous process variables. TEP can generate a data set of 1 normal state and 21 different fault states. Referring to the literature,²¹ we select 10 representative faults and divide them

into two cases with the aim of verifying the generalization ability and robustness of the proposed method. Generally, the selected 10 fault data have large overlap and are difficult to classify.^{14,35} The 10 fault types and descriptions are shown in Table 1. The fault type in case 1 is affected by feed and flow, and case 2 is affected by temperature. The faults in both cases occur under different operating conditions. Therefore, industrial process faults under different working conditions are diagnosed to verify the feasibility of our method. Each fault state includes 480 raw training samples and 800 raw test samples, respectively. Each sample is sampled at a frequency of 3 min.

4.2. Application Research of the Proposed Method in TEP Fault Diagnosis. With the goal of verifying the feasibility of the proposed method, we tested two subdata sets of TEP, and the accuracy of the test sets is taken as the effective performance of the industrial process fault diagnosis. All experiments are performed in Python 3.8 and Pytorch, running on Ubuntu 18.04 with 64GB RAM and an NVIDIA Quadro P4000 GPU.

4.2.1. Experimental Setup. The extended sliding window mechanism is adopted to convert the raw data set X into the augmented dynamic data, and the sliding step $S = 1$ and the sliding window length $L = 20$ are set to ensure that the augmented data D has enough dynamic information for neural network learning. Table 2 clearly reflects the number of samples

Table 2. Sample Size of Raw Data and Augmented Dynamic Data

fault	data set	sample size of raw data	sample size of augmented dynamic data
Each fault	Train	480	461
	Test	800	781

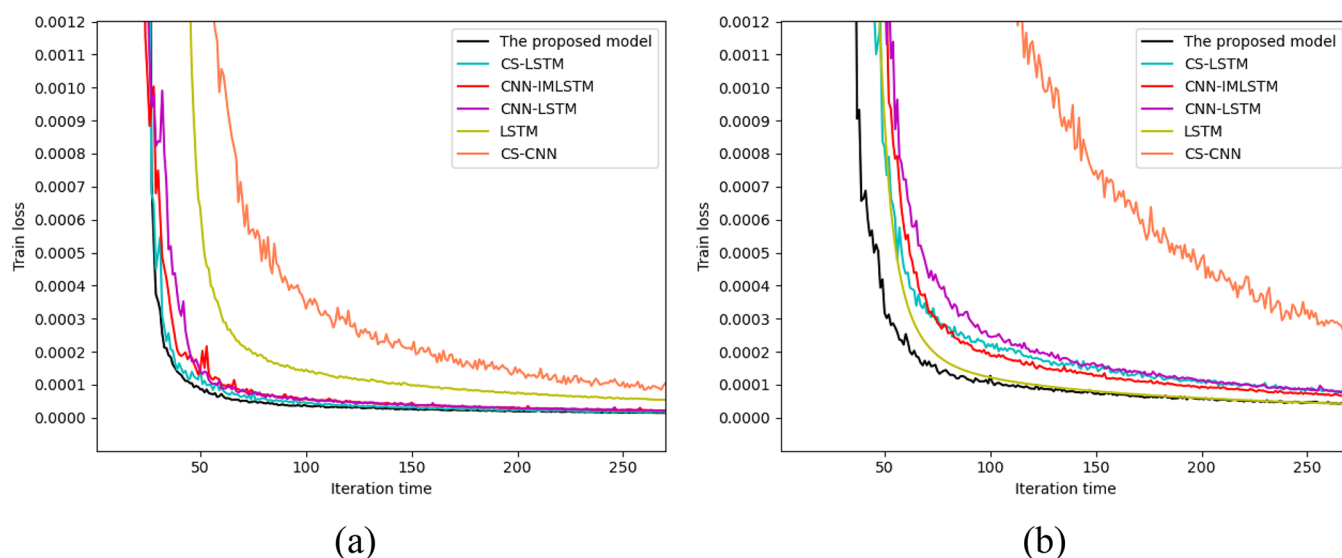
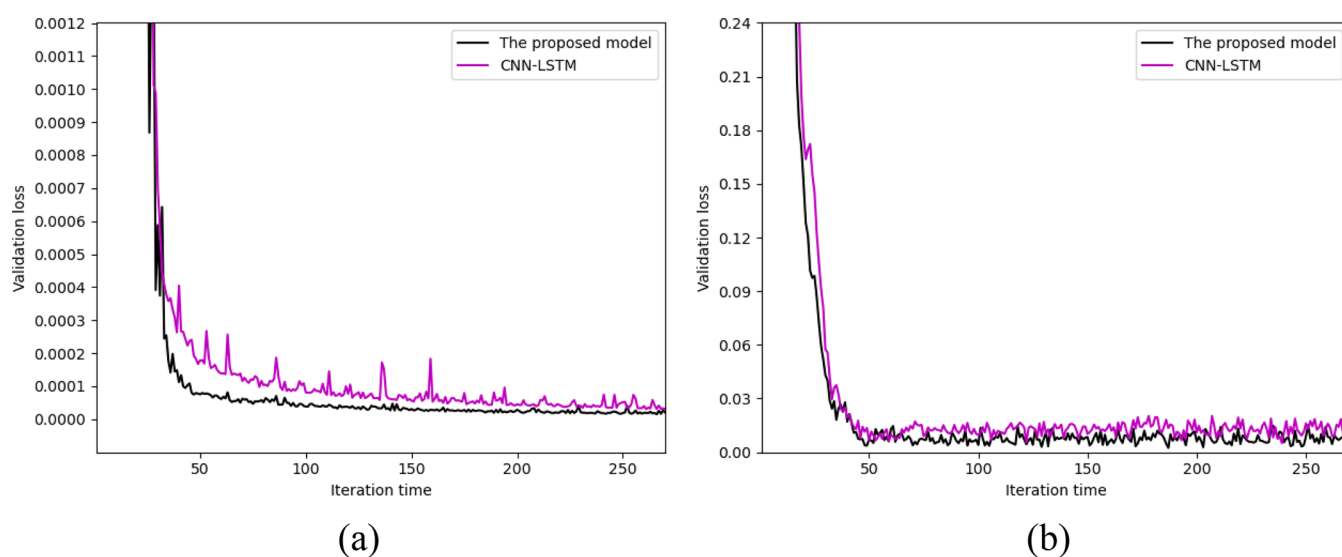
in the raw data set as well as the number of samples processed by the extended sliding window. Thus, the total number of train samples in each case is $461 \times 5 = 2305$, and the total number of test samples is $781 \times 5 = 3905$. It is worth noting that here each sample has $52 \times 20 = 1040$. In addition, we will draw 25% of data from the training sets of each case as the validation set during training.

In the training of the proposed model, the batch size is 32, learning rate is 0.001, convolution kernel is 1×3 , and max-pooling kernel is set to 1×2 . We choose Adam as the optimizer, use a cross-entropy loss function to evaluate the performance of the network, and use back-propagation to update the weights. For the sake of verifying the superiority of the proposed method, we set up five ablation comparison experiments. The hyperparameters of the models are approximately the same, and the complexity is approximately equal. The structure and other parameters of the different models are set as shown in Table 3.

Table 3. Model Structure and Parameter Settings

model	structure ^a
CS-IMLSTM	CONV(32)-SE(32)-CONV(64)-SE(64)-CONV(64)-SE(64)-FC*(512)-IMLSTM(1024)-FC(5)
CNN-IMLSTM	CONV(32)-CONV(64)-CONV(64)-FC*(512)-IMLSTM(1024)-FC(5)
CS-LSTM	CONV(32)-SE(32)-CONV(64)-SE(64)-CONV(64)-SE(64)-FC*(512)-LSTM(1024)-FC(5)
CNN-LSTM	CONV(32)-CONV(64)-CONV(64)-FC*(512)-LSTM(1024)-FC(5)
LSTM	Lstm1(1024)-lstm2(1024)-lstm3(1024)-lstm4(1024)-FC(5)
CS-CNN	CONV(32)-SE(32)-CONV(64)-SE(64)-CONV(64)-SE(64)-FC*(512)-CNN(512)-FC(5)

^aFor convenience, the CONV(@) module is used to denote Conv1d(@)-BN(@)-LeakyReLU-maxpooling(@), where @ denotes the output channel. * indicates FC layer with dropout rate of $p = 0.5$.

**Figure 7.** The 10 times average training loss curves of (a) case 1 and (b) case 2 on different models.**Figure 8.** The 10 times average validation loss curves of (a) case 1 and (b) case 2 on different models.

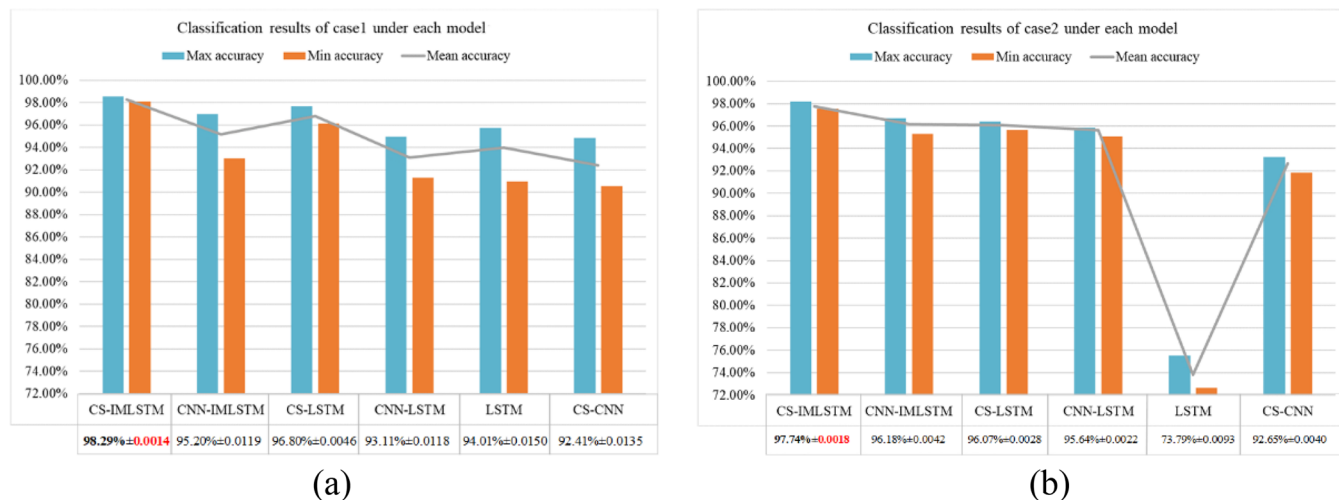
All experiments are repeated 10 times with the same terms. Finally, we use the accuracy of the test sets to evaluate the fault diagnosis capabilities of the different models.

4.2.2. Results and Discussion. The case 1 and case 2 training sets after extended sliding window processing are input to different models for training. After five epochs, the average training loss curve of 10 times obtained by each model are presented in Figure 7. From Figure 7a,b, it can be seen that the

models with LSTM structure or improved LSTM structure have a stronger convergence ability compared with the CS-CNN models. It shows that LSTM or improved LSTM can handle the time series data problem of TEP very well. From Figure 7, it can be observed that the proposed model has the most stable training loss value and the strongest convergence ability in both case 1 and case 2. Besides, from Figure 7 and Figure 8, the convergence ability of the proposed model is significantly better

Table 4. Classification Accuracy of Each Fault in Each Model

case	fault	proposed model (%)	CS-LSTM (%)	CNN-IMLSTM (%)	CNN-LSTM (%)	LSTM (%)	CS-CNN (%)
Case 1	1	99.36	99.35	98.85	99.36	99.74	97.57
	2	100.0	99.74	99.87	99.62	99.62	99.49
	6	99.23	99.87	99.87	100.0	98.98	99.74
	7	100.0	90.39	97.95	86.94	94.88	84.76
	8	94.24	94.37	88.35	88.99	85.53	92.70
Case 2	4	100.0	99.74	100.0	99.87	95.26	97.95
	5	99.23	97.70	98.98	98.72	91.93	97.18
	10	99.49	97.95	96.41	92.70	73.24	83.99
	11	93.85	90.01	91.17	90.78	47.25	89.88
	12	98.21	96.67	96.93	97.18	70.04	97.06

**Figure 9.** Classification results of (a) case 1 and (b) case 2 under each model.

than the traditional CNN-LSTM in terms of training loss and validation loss. Therefore, the proposed model has the strongest convergence and generalization ability compared to other models such as CNN-LSTM. This is mainly because SE can give more weight to the key channel features from CNN, and IMLSTM can balance historical fault information and adaptively capture the dynamic features of fault data through the updated gating mechanism.

The trained model is utilized to classify test sets and obtain classification accuracy. Table 4 shows the classification accuracy of each fault in the best results of each model. The best results are highlighted in bold in the table. From Table 4, the accuracy of the proposed method in case 1 and case 2 is more than 93.85%. The recognition accuracies of CS-LSTM, CNN-IMLSTM, CNN-LSTM, LSTM, and CS-CNN are more than 90.01%, 88.35%, 86.94%, 47.25%, and 83.99%, respectively. In the proposed model, fault 2, fault 4, and fault 7 can achieve 100% prediction accuracy. Compared with the other five models, fault 5, fault 10, fault 11, and fault 12 can get the best prediction accuracies, which are 99.23%, 99.49%, 93.85%, and 98.21%, respectively. The performance of LSTM in case 2 is not as good as that in case 1, which shows that the generalization performance of the LSTM model for chemical process fault diagnosis is poor. It is difficult for LSTM to mine the spatial information on industrial data without the assistance of CNN. Therefore, it is shown again that the fusion model CS-IMLSTM can pay attention to the important characteristics of industrial process fault data and adaptively process the dynamic information on data. From Table 4, the fault identification

results of other models are not as stable as those of the proposed model, indicating that the proposed model can learn more advanced features from extended dynamic data and improve the level of risk perception.

The work is repeated 10 times with the same terms, and the max accuracy, min accuracy, average accuracy, and standard deviation (std) are calculated. The diagnostic results are presented in Figure 9. The bold black text represents the average accuracy, and the bold red text represents the std. CS-IMLSTM achieved the highest average accuracy in both case 1 and case 2 test data sets with $98.29\% \pm 0.0014$ and $97.74\% \pm 0.0018$, respectively. The results demonstrate that the proposed model has high prediction accuracy and an excellent generalization performance. Specifically, the minimum accuracy obtained by the proposed model in case 1 is 1.15% higher than the maximum accuracy of CNN-IMLSTM, while the minimum accuracy obtained by the proposed model in case 2 is 0.85% higher than the maximum accuracy of CNN-IMLSTM. This indicates that the SE attention mechanism can focus on important channel features and boost the fault diagnosis performance of the model. The minimum accuracy obtained by the proposed model in case 1 is 0.46% higher than that of CS-LSTM, while the minimum accuracy obtained by the proposed model in case 2 is 1.13% higher than that of CS-LSTM. This indicates that IMLSTM can balance the fault information on industrial process data and capture the dynamic features of the temporal data more adequately than LSTM. The minimum accuracy obtained by the proposed model in case 1 is 3.15% higher than the maximum accuracy of CNN-LSTM, while the

minimum accuracy obtained by the proposed model in case 2 is 1.69% higher than the maximum accuracy of CNN-LSTM. This indicates that the organic combination of the SE attention mechanism, IMLSTM, and CNN can more fully exploit the feature information on augmented dynamic industrial data, achieve efficient flow of information, and improve the security of the process.

In the confusion matrix, the row stands for predicted fault labels, the column stands for actual fault labels, and the diagonal line indicates that predicted results are consistent with the real labels. Figure 10 provides the confusion matrix of the worst

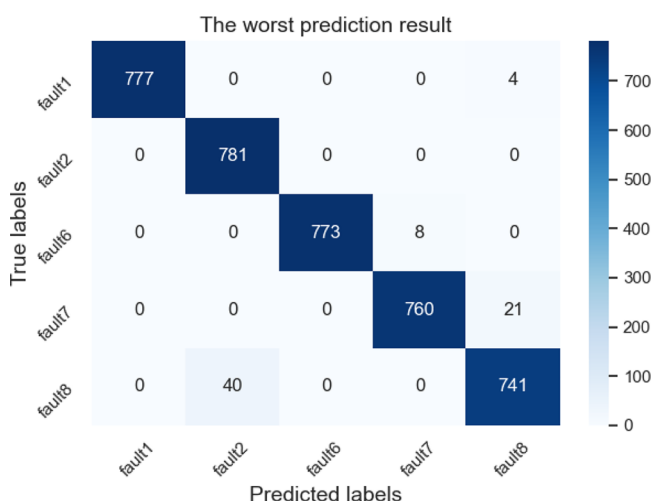


Figure 10. Confusion matrix for worst case prediction in case 1.

result of the proposed method in case 1, and its prediction accuracy is 98.13%. From Figure 10, fault 1, fault 2, and fault 6 are correctly predicted with 773 samples and above, while 760 samples are correctly predicted and 21 samples are incorrectly predicted as fault 8 in fault 7. Only 741 samples are correctly predicted, and 40 samples are misclassified as fault 2 in fault 8. In addition, we analyze the positive predictive value (PPV), true positive rate (TPR), and F1_Score³⁷ of this confusion matrix. It is worth noting that F1_Score here returns the score for each fault category. MacroF1_Score is the simple arithmetic mean of F1_Score. The results are presented in Table 5. The proposed

Table 5. Analytical Results of the Worst Confusion Matrix in Case 1

indicator	fault 1	fault 2	fault 6	fault 7	fault 8
PPV (%)	100	95.13	100	98.96	96.74
TPR (%)	99.49	100	98.98	97.31	94.88
F1_Score (%)	99.74	97.50	99.49	98.13	95.80
MacroF1_Score (%)			98.13		

method has high PPV, TPR, and F1_Score, and MacroF1_Score is 98.13%, which indicates that CS-IMLSTM can adequately extract the dynamic features of the data, thus enhancing the effectiveness of the fault diagnosis, improving the safety risk status of process industrial processes, guaranteeing process safety production, and increasing the economic efficiency of enterprises.

Figure 11 shows the confusion matrix for the worst result of the proposed method in case 2, and its prediction accuracy is 97.54%. All samples of fault 10 were correctly predicted. A total of 778 and 777 samples were correctly predicted for faults 4 and

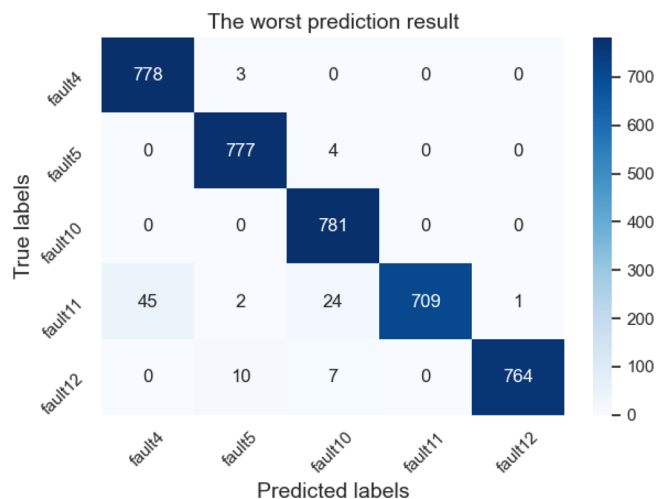


Figure 11. Confusion matrix for worst case prediction in case 2.

5, respectively. Fault 11 and fault 12 are inlet temperature fluctuations of the reactor and condenser cooling water, respectively, and are consistent with the fault descriptions of fault 4 and fault 5, respectively. Fault 11 and fault 12 are random variable types, and fault 4 and fault 5 are step variable types. Therefore, faults 11 and 12 are easily confused with faults 4 and 5, respectively. As can be seen from the figure, 45 samples of fault 11 are misclassified as fault 4, and 10 samples of fault 12 are misclassified as fault 5. Similarly, PPV, TPR, and F1_Score of the confusion matrix are analyzed, respectively. The analysis results are shown in Table 6. The method has high indicated PPV, TPR, and F1_Score, and MacroF1_Score is 97.52%, which indicates the effectiveness of using CS-IMLSTM for fault diagnosis.

Table 6. Analytical Results of the Worst Confusion Matrix in Case 2

indicator	fault 4	fault 5	fault 10	fault 11	fault 12
PPV (%)	94.53	98.10	95.71	100	99.87
TPR (%)	99.62	99.49	100	90.78	97.82
F1_Score (%)	97.01	98.79	97.81	95.17	98.84
MacroF1_Score (%)			98.13		

4.2.3. Comparison with Existing Advanced Methods. To further verify the superiority of the proposed method in extracting dynamic features of the chemical process industry, this paper compares it with dynamic PCA-SVM (DPCA-SVM) and transformer neural network. In DPCA-SVM, DPCA is a classical method for extracting dynamic features of data, and SVM is used for fault identification. The dynamic order h of DPCA is 2, and the contribution rate of the principal component is 0.99. The kernel function of SVM is RBF. Transformer is a neural network based on a pure attention mechanism to reflect the global dependence between input and output and has good identification performance in chemical process fault diagnosis.³⁸ Transformer's network architecture and hyperparametric references³⁹ take the enhanced dynamic data as the input, and the size of the input subsequence is 20. Similarly, all experiments are repeated 10 times under the same conditions, and the average accuracy is taken as the experimental result.

The results are shown in Table 7. The proposed method achieves the best performance in both case 1 and case 2. Compared with DPCA-SVM, the average accuracies of the

Table 7. Compare with Existing Advanced Methods

method	case 1	case 2	average accuracy
DPCA-SVM	89.73%	60.83%	75.28%
transformer	97.23%	78.51%	87.87%
proposed method	98.29%	97.74%	98.02%

proposed method in case 1 and case 2 are improved by 8.56% and 36.91%, respectively. Compared with the results of Transformer, the average accuracies of the proposed method in case 1 and case 2 are improved by 1.06% and 19.23%, respectively. Experimental results show that, compared with these advanced fault diagnosis methods, the proposed method can extract the dynamic features of process data better and has the highest fault diagnosis results and the best generalization performance.

From all the above results and analysis, it can be concluded that the proposed method has the most desirable fault diagnosis performance compared to all comparison experiments. This is mainly because, before classification, the extended sliding window is used to generate expanded dynamic data. The CS-IMLSTM model is used to learn spatial, channel, and temporal information on industrial process data and deeply excavates the dynamic information on the data. Therefore, the classification performance is improved.

5. CONCLUSION

In this paper, the CS-IMLSTM model is designed for chemical industrial process fault diagnosis, which solves the problem that the traditional CNN-LSTM model and other forms of this model have difficulty extracting dynamic fault features of chemical processes. The contributions are specified as follows:

- (1) In terms of data preprocessing, an extended sliding window mechanism is proposed. The mechanism provides raw data with strong dynamic information for the proposed model and lays a foundation for the highest accuracy of the proposed model on the TEP data set.
- (2) In terms of feature extraction, the CS-IMLSTM model is proposed. We introduce the SE attention mechanism into the CNN, which can adaptively assign more weight to key fault features to optimize fault features. In addition, we propose an IMLSTM, which alleviates the excessive dependence of LSTM on the current or previous fault information, so that LSTM can pay more attention to the features of industrial data in the time dimension, balance the fault information, and adaptively extract the dynamic information of the data. Finally, CS-IMLSTM is constructed by integrating the CS network and IMLSTM, which can extract the spatial and temporal dynamic characteristics of process industry data simultaneously.
- (3) The effectiveness of the proposed method is verified in TEP. Compared with five comparison experiments including CNN-LSTM, CS-IMLSTM obtain the highest average accuracies of $98.29\% \pm 0.0014$ and $97.74\% \pm 0.0018$ in both subdata of the TEP. The simulation results verify the feasibility of the proposed method

The proposed method can better capture the dynamic fault information of a chemical process and enhance the performance of fault diagnosis. Therefore, CS-IMLSTM can provide RA and ASM with a more favorable decision-making basis based on the dynamic fault information of chemical processes and deploy

remedial actions and implement safety measures in time to minimize process risks and avoid safety accidents.

The extended sliding window mechanism and deep learning network need to occupy a large amount of computer memory resources. Therefore, in future research, from the perspective of data preprocessing, it is an effective approach to improve the data quality by variable screening of multivariable industrial process data. In terms of network architecture design, network quantization, network decomposition, and lightweight network design are worthy of future research.

AUTHOR INFORMATION

Corresponding Author

Jian Cen – School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China; Guangzhou Intelligent Building Equipment Information Integration and Control Key Laboratory, Guangzhou 510665, China; orcid.org/0000-0002-1714-7397; Email: mmcjian@163.com

Authors

Honghua Chen – School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China; Guangzhou Intelligent Building Equipment Information Integration and Control Key Laboratory, Guangzhou 510665, China

Zhuohong Yang – School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China; Guangzhou Intelligent Building Equipment Information Integration and Control Key Laboratory, Guangzhou 510665, China

Weiwei Si – School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China; Guangzhou Intelligent Building Equipment Information Integration and Control Key Laboratory, Guangzhou 510665, China

Hongchao Cheng – School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665, China; Guangzhou Intelligent Building Equipment Information Integration and Control Key Laboratory, Guangzhou 510665, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c04017>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Innovative Team Project of the Ordinary University of Guangdong Province [Grant Number 2020KCXTD017]; the Guangdong Special Project in Key Field of Artificial Intelligence for the Ordinary University [Grant Number 2019KZDZX1004]; the Guangzhou Key Laboratory Construction Project [Grant Number 202002010003]; and the Guangzhou Key Research and Development Project [Grant Number 202206010022].

REFERENCES

- (1) Shu, Y.; Ming, L.; Cheng, F.; Zhang, Z.; Zhao, J. Abnormal situation management: challenges and opportunities in the big data era. *Comput. Chem. Eng.* **2016**, *91*, 104–113.
- (2) Arunthavanathan, R.; Khan, F.; Ahmed, S.; Imtiaz, S. An analysis of process fault diagnosis methods from safety perspectives. *Comput. Chem. Eng.* **2021**, *145*, 107197.

- (3) Khan, F.; Rathnayaka, S.; Ahmed, S. Methods and models in process safety and risk management: past, present and future. *Process Saf. Environ. Prot.* **2015**, *98*, 116–147.
- (4) Dai, Y.; Wang, H.; Khan, F.; Zhao, J. Abnormal situation management for smart chemical process operation. *Curr. Opin. Chem. Eng.* **2016**, *14*, 49–55.
- (5) Benson, C.; Argyropoulos, C. D.; Dimopoulos, C.; Mikellidou, C. V.; Boustras, G. Safety and risk analysis in digitalized process operations warning of possible deviating conditions in the process environment. *Process Saf. Environ. Prot.* **2021**, *149*, 750–757.
- (6) Amin, M. T.; Khan, F.; Ahmed, S.; Imtiaz, S. Risk-based fault detection and diagnosis for nonlinear and non-Gaussian process systems using R-vine copula. *Process Saf. Environ. Prot.* **2021**, *150*, 123–136.
- (7) Bao, H.; Khan, F.; Iqbal, T.; Chang, Y. Risk-based fault diagnosis and safety management for process systems. *Process Saf. Prog.* **2011**, *30* (1), 6–17.
- (8) Bhadriraju, B.; Kwon, J. S.-I.; Khan, F. Risk-based fault prediction of chemical processes using operable adaptive sparse identification of systems (OASIS). *Comput. Chem. Eng.* **2021**, *152*, 107378.
- (9) Md Nor, N.; Che Hassan, C. R.; Hussain, M. A. A review of data-driven fault detection and diagnosis methods: applications in chemical process systems. *Rev. Chem. Eng.* **2020**, *36* (4), 513–553.
- (10) Deng, X.; Cai, P.; Cao, Y.; Wang, P. Two-step localized kernel principal component analysis based incipient fault diagnosis for nonlinear industrial processes. *Ind. Eng. Chem. Res.* **2020**, *59* (13), 5956–5968.
- (11) Han, Y.; Song, G.; Liu, F.; Geng, Z.; Ma, B.; Xu, W. Fault monitoring using novel adaptive kernel principal component analysis integrating grey relational analysis. *Process Saf. Environ. Prot.* **2022**, *157*, 397–410.
- (12) He, Y.-L.; Li, K.; Liang, L.-L.; Xu, Y.; Zhu, Q.-X. Novel discriminant locality preserving projection integrated with monte carlo sampling for fault diagnosis. *IEEE T Reliab.* **2021**, 1–11.
- (13) Zhang, N.; Xu, Y.; Zhu, Q.-X.; He, Y.-L. Improved locality-preserving projections based on heat-kernel and cosine weights for fault classification in complex industrial processes. *IEEE T Reliab.* **2022**, 1.
- (14) Zhang, N.; Xu, Y.; Zhu, Q.-X.; He, Y.-L. Farthest-nearest distance neighborhood and locality projections integrated with bootstrap for industrial process fault diagnosis. *IEEE T Ind. Inform.* **2022**, DOI: 10.1109/TII.2022.3182774.
- (15) Amin, M. T.; Khan, F.; Ahmed, S.; Imtiaz, S. A data-driven Bayesian network learning method for process fault diagnosis. *Process Saf. Environ. Prot.* **2021**, *150*, 110–122.
- (16) Deng, Z.; Han, T.; Cheng, Z.; Jiang, J.; Duan, F. Fault detection of petrochemical process based on space-time compressed matrix and Naive Bayes. *Process Saf. Environ. Prot.* **2022**, *160*, 327–340.
- (17) Wu, H.; Zhao, J. Deep convolutional neural network model based chemical process fault diagnosis. *Comput. Chem. Eng.* **2018**, *115*, 185–197.
- (18) Song, Q.; Jiang, P. A multi-scale convolutional neural network based fault diagnosis model for complex chemical processes. *Process Saf. Environ. Prot.* **2022**, *159*, 575–584.
- (19) Yu, J.; Zhang, C.; Wang, S. Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis of industrial processes. *Neural Comput. Appl.* **2021**, *33* (8), 3085–3104.
- (20) Yu, W.; Zhao, C. Broad Convolutional Neural Network based Industrial Process Fault Diagnosis with Incremental Learning Capability. *IEEE Trans Ind. Electron.* **2020**, *67* (6), 5081–5091.
- (21) Zhao, H.; Sun, S.; Jin, B. Sequential Fault Diagnosis based on LSTM Neural Network. *IEEE Access.* **2018**, *6*, 12929–12939.
- (22) Han, Y.; Ding, N.; Geng, Z.; Wang, Z.; Chu, C. An optimized long short-term memory network based fault diagnosis model for chemical processes. *J. Process Control.* **2020**, *92*, 161–168.
- (23) Park, P.; Marco, P. D.; Shin, H.; Bang, J. Fault detection and diagnosis using combined autoencoder and long short-term memory network. *Sensors.* **2019**, *19* (21), 4612.
- (24) Gravanis, G.; Dragogias, I.; Papakiriakos, K.; Ziogou, C.; Diamantaras, K. Fault detection and diagnosis for non-linear processes empowered by dynamic neural networks. *Comput. Chem. Eng.* **2022**, *156*, 107531.
- (25) Shao, B.; Hu, X.; Bian, G.; Zhao, Y. A multichannel LSTM-CNN method for fault diagnosis of chemical process. *Math Probl Eng.* **2019**, *2019* (3), 1032480.
- (26) Wang, N.; Yang, F.; Zhang, R.; Gao, F. Intelligent fault diagnosis for chemical processes using deep learning multimodel fusion. *IEEE Trans Cybern.* **2020**, 7121–7135.
- (27) Yuan, J.; Tian, Y. A multiscale feature learning scheme based on deep learning for industrial process monitoring and fault diagnosis. *IEEE Access.* **2019**, *7*, 151189–151202.
- (28) Huang, T.; Zhang, Q.; Tang, X.; Zhao, S.; Lu, X. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artif Intell. Rev.* **2022**, *55* (2), 1289–1315.
- (29) Chen, Z.; Cen, J.; Xiong, J. Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network. *IEEE Access.* **2020**, *8*, 150248–150261.
- (30) Albawi, S.; Mohammed, T. A.; Al-Zawi, S. Understanding of a convolutional neural network. *Proceedings of the 2017 International Conference on Engineering and Technology (ICET)*; IEEE: 2017; pp 1–6.
- (31) Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell.* **2020**, *42* (8), 2011–2023.
- (32) Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9* (8), 1735–1780.
- (33) Han, Y.; Qi, W.; Ding, N.; Geng, Z. Short-time wavelet entropy integrating Improved LSTM for fault diagnosis of modular multilevel converter. *IEEE Trans Cybern.* **2021**, 7504–7512.
- (34) Downs, J. J.; Vogel, E. F. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, *17* (3), 245–255.
- (35) Arunthavanathan, R.; Khan, F.; Ahmed, S.; Imtiaz, S.; Rusli, R. Fault detection and diagnosis in process system using artificial intelligence-based cognitive technique. *Comput. Chem. Eng.* **2020**, *134*, 106697.
- (36) Ma, L.; Dong, J.; Peng, K. A novel key performance indicator oriented hierarchical monitoring and propagation path identification framework for complex industrial processes. *ISA Trans.* **2020**, *96*, 1–13.
- (37) Azamfar, M.; Li, X.; Lee, J. Deep Learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing. *IEEE Trans Semicond Manuf.* **2020**, *33* (3), 445–453.
- (38) Zhang, L.; Song, Z.; Zhang, Q.; Peng, Z. Generalized transformer in fault diagnosis of Tennessee Eastman process. *Neural Comput. Appl.* **2022**, *34* (11), 8575–8585.
- (39) Yang, Z.; Cen, J.; Liu, X.; Xiong, J.; Chen, H. Research on bearing fault diagnosis method based on transformer neural network. *Meas. Sci. Technol.* **2022**, *33* (8), 085111.