ACS OMEGA

Open Access

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

Article

Physical Graph-Based Spatiotemporal Fusion Approach for Process Fault Diagnosis

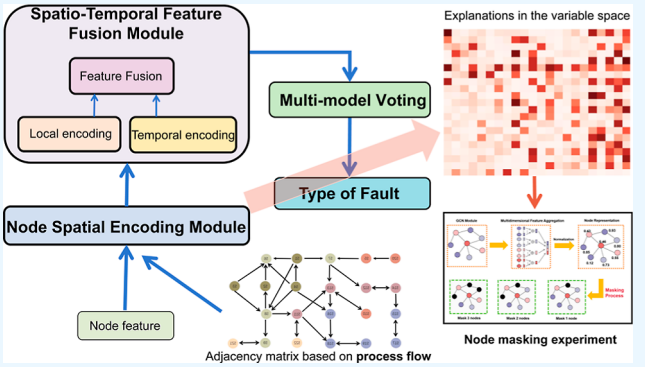
Fengzhen Zhang, Qibing Jin, Dazi Li,* Yang Zhang,* and Qian Zhu

Cite This: *ACS Omega* 2024, 9, 9486−9502

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The rapid development of big data technology and machine learning has increasingly focused attention on fault diagnosis in complex chemical processes. However, data-driven approaches often overlook the inherent physical correlations within the system and lack a robust mechanism for providing trusted explanations for fault diagnosis. To address this challenge, a graph-based fault diagnosis model framework is proposed along with a dependable fault node diagnosis analysis method. In order to enhance the extraction of chemical process features from a spatial perspective, a graph convolution network (GCN)-based node spatial encoding module is integrated. The construction of the adjacency matrix involves combining a priori knowledge of chemical processes with Pearson correlation, thereby incorporating the physical correlations between nodes. Simultaneously, to capture temporal dependencies in fault data, a spatiotemporal feature fusion module based on the long short-term memory network (LSTM) is employed. In terms of model training, a dual-supervision strategy is adopted to ensure stable convergence of the multiclass fault diagnosis model. For model inference, a multi-model voting strategy is designed to mitigate accuracy degradation resulting from model prediction bias. To tackle the interpretability challenge, a fault diagnosis analysis method based on node masking is designed, effectively identifying critical nodes contributing to system faults. Experimental validation on the Tennessee Eastman process demonstrates the effectiveness of the proposed model, achieving high accuracy in fault diagnosis. The average fault diagnosis rate for all fault types reaches 0.9844, showcasing state-of-the-art performance in fault diagnosis.



1. INTRODUCTION

Chemical process production plays a crucial role in the advancement of industrial production. It is of paramount importance to diagnose and analyze the anomalies in chemical processes for the safety of chemical systems.[1] Consequently, fault diagnosis in complex systems has garnered significant interest among researchers.[2] Fault detection aims to determine whether a system has malfunctioned, while fault diagnosis seeks to accurately identify and classify the types of faults that occur in a system, as well as determine the root cause (variable), magnitude, and location of the fault.[3] The fault diagnosis rate refers to the accuracy of fault classification.

In the past few years, the rapid emergence of artificial intelligence (AI) and big data technologies has garnered significant attention in the academic community for data-driven fault diagnosis methods.[4] Data-driven fault diagnosis methods can be mainly divided into several categories: data-based probability analysis, machine learning methods, time series analysis, multivariate statistical learning methods, and deep learning methods. Data-driven probabilistic analysis methods utilize probability and statistical theory to perform fault diagnosis through probabilistic analysis of the observed data. Examples of these methods include Bayesian networks,[5] hidden Markov models (HMMs),[6] and Markov chains.[7] Machine learning methods employ pattern recognition and learning from data to conduct fault diagnosis. Common machine learning algorithms used in this context include support vector machines (SVMs),[8] decision trees,[9] random forests,[10] and k-nearest neighbors (KNNs).[11] Time series analysis is particularly important for understanding and diagnosing time-dependent faults. Multivariate statistical learning methods include well-known classical techniques such as principal component analysis (PCA),[12] independent component analysis (ICA),[13] and partial least squares (PLS).[14] These methods aim to reduce the dimensionality of high-dimensional data by transforming it into a lower-dimensional

Received: November 16, 2023
Revised: January 8, 2024
Accepted: February 6, 2024
Published: February 16, 2024

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

space. Anomaly scores, typically based on statistics like Hotelling's T^2 or squared prediction error (SPE), are then computed among the new variables in the lower-dimensional space. These scores are compared to predefined thresholds to facilitate fault diagnosis.[15] Deep learning is a machine learning method based on artificial neural networks, which can automatically learn features and patterns from data through multilevel neural network structures and learning algorithms.

Deep learning methods have shown remarkable success in modeling complex chemical processes, particularly in the field of fault diagnosis. Various classical neural network structures, such as the convolutional neural network (CNN), deep confidence network (DCN), autoencoder (AE), long short-term memory (LSTM), generative adversarial network (GAN), gated recurrent units (GRUs), and attention mechanisms, have been employed for fault diagnosis tasks. For instance, Zhang and Zhao adopted a scalable deep belief network (DBN) and achieved an average fault diagnosis rate of 82.1% for all 20 fault types in the Tennessee Eastman (TE) process.[16] Wu and Zhao proposed a fault diagnosis method using a deep CNN (DCNN) model,[17] outperforming other fault diagnosis methods reported in the literature. Zhang et al. employed bidirectional recurrent neural networks (BiRNNs) to construct a fault diagnosis and detection model of complex RNN units.[18] Deng et al. introduced a dynamic CNN method based on a genetic algorithm, achieving an average diagnosis rate of 89.72% for 20 faults.[19] Chen et al. utilized an encoder-decoder network with a self-focusing mechanism for fault classification.[20]

In recent years, an increasing emphasis has been placed on the utilization of physics-inspired neural networks or AI methodologies to facilitate the effective integration of data-driven techniques with the dynamic characteristics inherent in physical systems.[21] This integration is designed to elevate the comprehension, modeling, and analysis of intricate physical phenomena, resulting in enhanced predictive capacities and profound insights into the underlying mechanisms. The incorporation of prior knowledge, such as physical space layout and reaction mechanisms, into graph structures within the chemical process industry has been explored.[22] This facilitates the explicit learning of inherent a priori knowledge by graph-based models, culminating in more precise representations and improved capabilities for capturing the intricacies of real-world processes. GCN is a class of neural network models designed specifically for processing graph-structured data. Relationships and topologies in graphs are effectively captured by GCN, making them well-suited for fault diagnosis tasks in the chemical industry. For instance, Zhang and Yu introduced a novel graph neural network called pruned graph convolutional network (PGCN) for feature learning on graph-structured data.[23] They transformed one-dimensional process data into graph data using graph construction methods and utilized GCN to extract features from the process data. Jia et al. developed a topology-guided graph learning framework for fault diagnosis, integrating graphs with process physics.[24] They constructed symbolic directed graphs (SDGs) to describe the process topology and employed GCN structures and convolutional gating mechanisms to propagate information based on the topological graph structure. Wu et al. integrated GCN, self-attention mechanisms, and process topology knowledge, incorporating information from both upstream and downstream processes, resulting in high-performance fault diagnosis models.[25] These studies underscore the importance of incorporating process knowledge into GCN-based models for fault diagnosis. By leveraging graph structures and considering underlying process relationships, enhanced accuracy in fault diagnosis tasks in the chemical industry can be achieved.

In the task of fault diagnosis, accurately extracting temporal features is crucial for comprehending the underlying patterns and dynamics of the system's behavior. LSTM networks are well-suited for capturing temporal dynamic features in complex systems. Equipped with specialized memory units and gating mechanisms that can selectively store and access relevant information over a long period of time,[26] LSTM networks have been leveraged by researchers in the field of fault diagnosis to develop effective models. For instance, Zhao et al. proposed a fault diagnosis network based on LSTM and conducted experiments on the TE process, achieving an average fault diagnosis rate of 80% for 20 different faults.[27] Zhang et al. proposed a three-layer stacked LSTM network that effectively models sequential data and detects anomalies by fully utilizing long-term dependency information in the raw data.[28] Additionally, Zhang and Qiu presented a semi-supervised approach called LSTM-LAE (long short-term memory ladder self-encoder) for fault diagnosis.[29]

Indeed, the interpretability of data-driven fault diagnosis models remains a significant challenge. While these models have achieved high diagnostic performance, transparency in terms of the specific process variables contributing to their fault prediction results is often lacking.[30] However, understanding the variables associated with faults is crucial for operators to make informed decisions and take corrective actions to restore normal operation.[31] To address this issue, current research is focused on the development of fault diagnosis models with both high diagnostic performance and interpretability. An approach to enhance interpretability by assessing the physical consistency of the model using a GNN interpreter was proposed by Jia et al.[24] This ensures transparency in the model prediction process, enabling operators to better understand how the model arrives at its results. Wu et al. explained the basis for fault decisions in a diagnostic model by visualizing the self-attention mechanism weights.[25] By examining the attention weights assigned to different variables, we facilitate understanding regarding which variables are crucial for fault classification. In addition, Gangopadhyay et al. proposed a spatiotemporal attention module that enhances understanding of the contributions of different features to the predictive outputs of time series.[32] This mechanism aids in identifying critical temporal dynamics and specific features during the fault process. These studies underscore the importance of interpretability in fault diagnosis models. Interpretive analysis of fault diagnosis provides operators with reliable and actionable troubleshooting guidance.

In this study, the synergies of GCN and LSTM are leveraged to propose an advanced fault diagnosis model, integrating prior knowledge. The model comprises a GCN-based node spatial encoding module and a LSTM-based spatiotemporal feature fusion module. This combined approach enables the extraction of features possessing both spatial and temporal characteristics, which are essential for precise fault diagnosis in complex industrial chemical process systems. Additionally, the study introduces a node-masking-based fault diagnosis analysis method, contributing to enhanced interpretability. This methodology effectively identifies pivotal variables, demon-

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

strating robust physical correlations with the origins of faults. The primary contributions of this work can be briefly summarized as follows:

1 A fault diagnosis model tailored for the chemical industry is presented. The model employs a GCN-based node spatial encoding module to extract spatial features from nodes. The construction of the adjacency matrix utilizes a methodology grounded in prior knowledge of the chemical process, complemented by Pearson correlation. This approach ensures a thorough consideration of the physical correlations between nodes. Furthermore, a LSTM-based spatiotemporal feature fusion module is introduced to effectively capture correlations within time series data, thereby enhancing the model's capacity for fault diagnosis.
2 A dual-supervision strategy is employed for model training to ensure stable convergence of the multiclass fault diagnosis model. Also, a multi-model voting strategy is devised for model inference to mitigate the accuracy degradation caused by prediction bias.
3 A fault diagnosis analysis method based on node masking is designed. This method identifies the crucial nodes associated with faults by analyzing the weight distribution of the model. The importance of these key nodes in the fault diagnosis inference process is verified through node masking experiments.

The rest of this article is organized as follows: Section 2 introduces the basic principles of graphs and LSTM networks. Section 3 presents the proposed model framework for fault diagnosis. Section 4 is a case study of the TE process, which shows the high performance of the proposed model. Section 4.5 describes the proposed fault diagnosis analysis method. Section 5 presents the results and discussion. Section 6 concludes the paper.

2. PRELIMINARIES

2.1. Graph Convolutional Network. A graph is generally defined as a collection of nodes (V) and edges (E), representing complex relationships between entities. It can be represented using an adjacency matrix A and node features X. Nodes are the fundamental units in a graph, while edges represent the relationships between nodes. Neighbor nodes refer to the nodes that are directly connected to a specific node. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes, and the elements in the matrix indicate whether there is an edge between nodes, $A = (a_{ij})_{n\times n}$. The expression of a_{ij} is shown in eq 1. Node features refer to the attributes or feature vectors associated with each node. Node features can be of any data type, such as numerical, textual, or image data

$$a_{ij} = \begin{cases} 1 & (v_i, v_j) \in E \\ 0 & (v_i, v_j) \notin E \end{cases} \tag{1}$$

where a_{ij} represents the element in the adjacency matrix corresponding to the connection between nodes v_i and v_j. $(v_i,v_j) \in E$: this condition checks whether an edge exists between nodes v_i and v_j. If an edge exists, a_{ij} is set to 1, indicating a connection. $(v_i,v_j) \notin E$: if there is no edge between nodes v_i and v_j, a_{ij} is set to 0, signifying the absence of a connection.

2.2. Long Short-Term Memory Network. LSTM is built upon the foundation of RNNs and addresses the issue of long-term dependencies in processing long sequential data.[26] It introduces gated structures that enable dynamic control over information retention and discard. The LSTM unit structure is depicted in Figure 1.

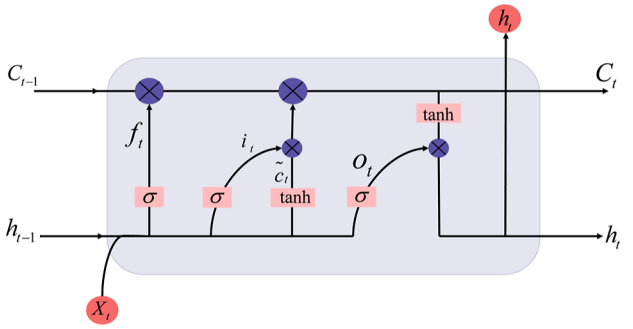


Figure 1. Architecture of LSTM cell.

LSTM utilizes a forget gate, an input gate, and an output gate to control the cell state. The forget gate is responsible for integrating the previous hidden state h_{t-1} and the current t time input vector X_t. It uses a sigmoid function to generate the output vector f_t. The previous cell state C_{t-1} is then multiplied element-wise with f_t to perform data forgetting, effectively controlling which information should be discarded from the cell state. The expression formula for the forgetting gate f_t is

$$f_t = \sigma(W_t[h_{t-1}, X_t] + b_c) \tag{2}$$

where $\sigma(\cdot)$ represents the sigmoid function. W_t refers to the weight matrix associated with the forget gate. b_c represents the bias term associated with the forget gate.

The input gate arithmetic expression is

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, X_t] + b_c) \tag{3}$$

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \tag{4}$$

where $\tilde{C}_t$ represents the candidate state. W_c and W_i are the weight matrices associated with the candidate state and the input gate, respectively. i_t represents the input gate weight vector. b_i represents the bias term associated with the input gate.

To convert the previous cell state C_{t-1} to the current cell state C_t, the conversion formula is

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_{t-1} \tag{5}$$

where "*" symbol represents the element-wise multiplication operator between matrices or vectors.

The output gate arithmetic expression is

$$O_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \tag{6}$$

$$h_t = O_t * \tanh C_t \tag{7}$$

where O_t represents the hidden layer state weight vector. W_o represents the output gate to be a trained parameter matrix. b_o represents the output gate to be trained for the bias term. h_t represents the current moment's hidden layer state.

The current hidden state h_t and the current input vector X_t interact at the input gate. Applying the sigmoid function to this interaction yields the hidden state weight vector O_t. The

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

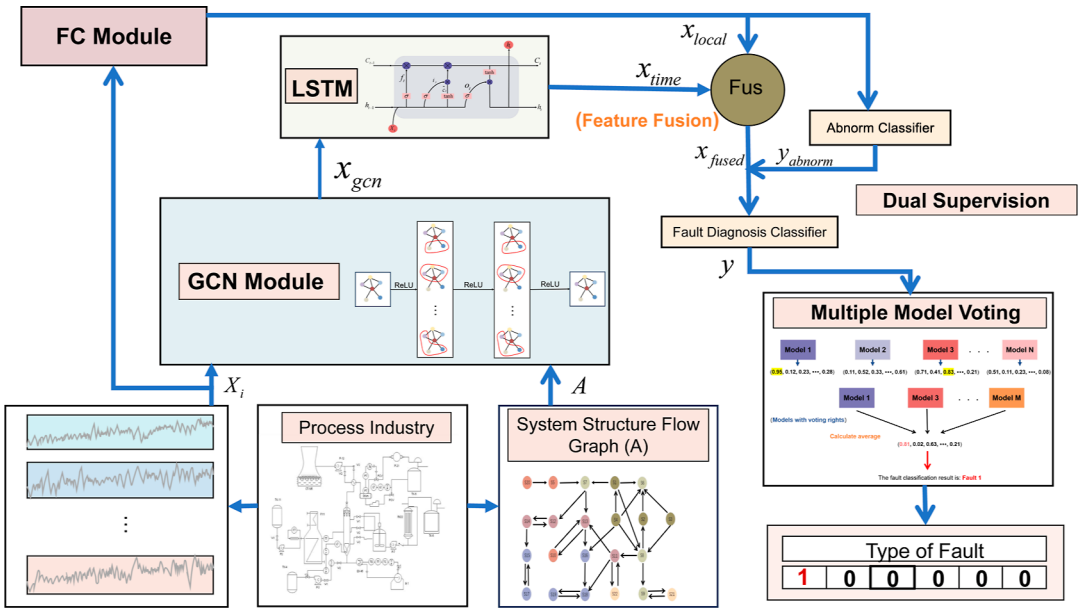


Figure 2. PG-STF model framework.

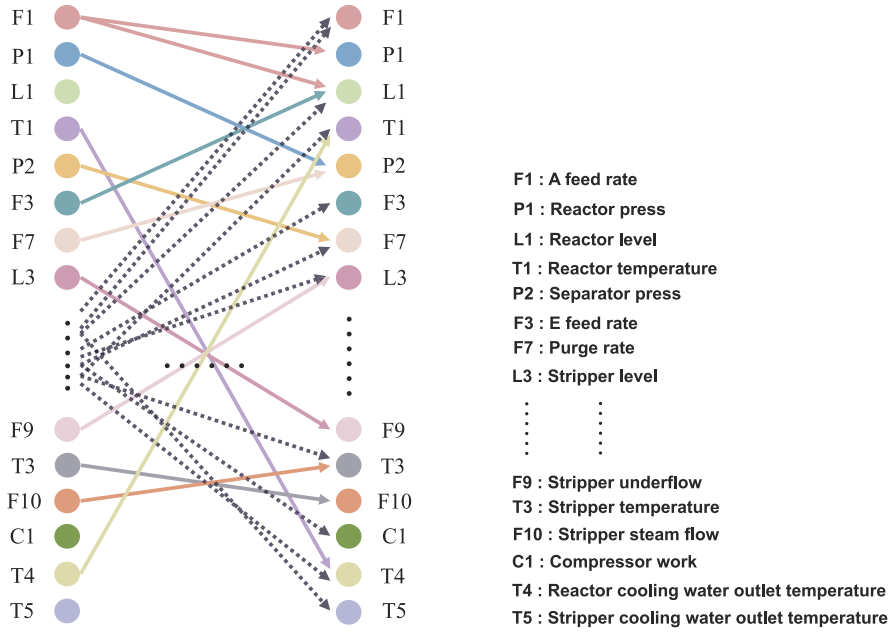


Figure 3. Visualization of 22 node connections.

current cell state C_t is then normalized using the hyperbolic tanh function, resulting in a vector with all elements ranging from −1 to 1. Multiplying this normalized cell state C_t with the hidden state weight vector, O_t selectively forgets certain information. This process yields the updated hidden state h_t at the current time step, which completes the updating of the short-term memory.

3. FAULT DIAGNOSIS FRAMEWORK

A novel fault diagnosis model framework, named PG-STF, is proposed, as illustrated in Figure 2. This framework comprises a node spatial encoding module and a spatiotemporal feature fusion module. Within the node spatial encoding module, GCN is employed to acquire node representations. The construction of the adjacency matrix incorporates prior knowledge of chemical processes, effectively depicting the physical correlations inherent in the system. Pertinent information and features crucial for fault diagnosis are derived through the utilization of spatial relationships between nodes. To integrate temporal information, the spatiotemporal feature fusion module amalgamates temporal encoding with local spatial encoding. This approach empowers the model to capture both temporal correlations and local spatial features, thereby elevating the overall fault diagnosis performance.

For model training, a dual-supervised training strategy is introduced to accelerate convergence and ensure a stable model performance. Additionally, an efficient multi-model voting inference strategy is proposed to improve the decision-making process. By aggregating predictions from multiple models, the model can make more accurate and reliable fault

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

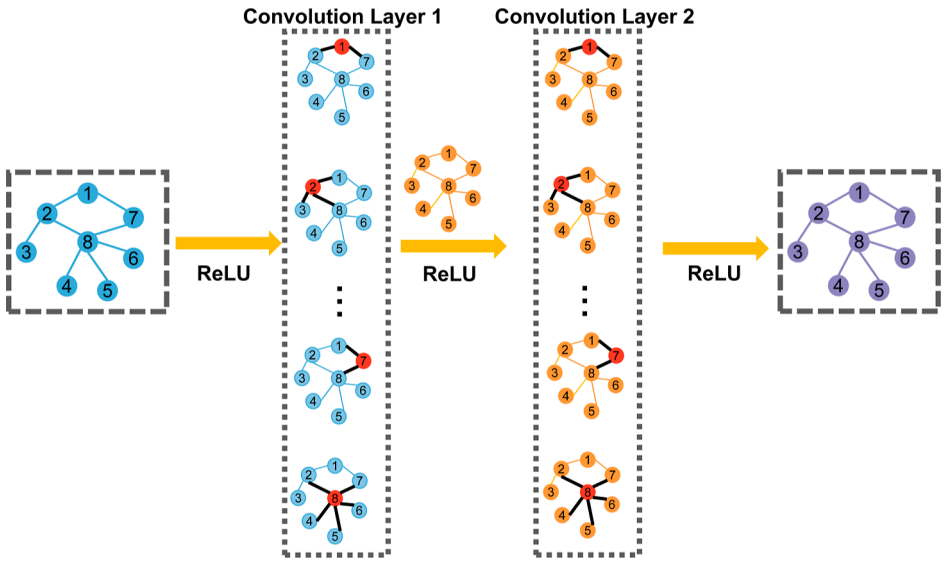


Figure 4. GCN with two graph convolution layers.

diagnosis decisions, mitigating the accuracy problem caused by prediction bias.

3.1. Node Spatial Encoding Module. The node spatial encoding module employs GCN to obtain the node representations. A priori knowledge of the chemical process is utilized to construct a topology that describes the physical associations within the system. The process consists of two steps: first, the graph data is transformed into an adjacency matrix representation and, second, spatial node encoding is conducted by defining feature vectors for each node.

3.1.1. Building the Adjacency Matrix. Constructing an adjacency matrix involves utilizing a priori knowledge of the structure of process flows in chemical systems. Analyzing the process flow allowed us to identify the nodes corresponding to process variables and their interconnections. This information is crucial for building an adjacency matrix that represents the relationships between nodes. The node set V consists of continuous variables in the chemical process system, while the edge set E delineates the connectivity between these variables. The construction of the adjacency matrix, guided by the chemical process flow, ensures that the GCN accurately captures the physical correlations within the system and effectively learn features from the graph data.

Figure 3 presents a visualization of the connectivity relationships among the 22 nodes (continuous variables) in the adjacency matrix. The matrix is based on a chemical simulation of the TE process. The 22 nodes are represented in the figure by different colored circles. The arrows indicate the connections between the nodes, illustrating how the variables in the TE process are interconnected and highlighting the effect that one variable may have on another. For example, F1 represents the feed flow rate of component A. In the constructed adjacency matrix, F1 is connected to two other nodes, namely, P1 (reactor pressure) and L1 (reactor level). This connection signifies that the component A feed flow rate (F1) has a relationship with both the reactor pressure (P1) and the reactor level (L1). Understanding these connections helps in comprehending the complex dynamics and dependencies between the continuous measurement variables in the TE process.

3.1.2. Spatial Node Encoding. The input feature vector $X_{N\times D}$ represents the node feature information, where N is the number of nodes and D is the dimension of the feature vector for each node. The coded representation of each node, $H^{(0)}$ is initialized randomly. Spatial node encoding is conducted using GCN, which leverages the graph structure to enhance the node representations by combining node features with the graph topology. The transfer function of GCN is defined as follows[33]

$$f(H^{(l)}, A) = \sigma(\hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}H^{(l)}W^{(l)}) \tag{8}$$

where A represents the adjacency matrix, $\hat{A} = A + I$, $\hat{A}$ represents the symmetric normalized matrix, and I is the identity matrix. $\hat{D}$ is the degree matrix of $\hat{A}$, which is a diagonal matrix with diagonal elements $\hat{D}_{ij} = \sum j\hat{A}_{ij}$. $H^{(l)}$ represents the node feature matrix at layer l, and the initial input layer $H^{(1)} = X$. $W^{(l)}$ is the weight matrix at layer l. σ is the activation function, typically ReLU.

The GCN structure conducts graph convolution operations using the adjacency matrix and node features. The architecture of the GCN is depicted in Figure 4. At each layer $I = 1, 2, ..., L$ (where L is the number of graph convolution layers), a graph convolution operation takes place. This operation aggregates the features of a node with those of its neighboring nodes, facilitating the learning of comprehensive node representations. To obtain richer node representations, a multilayer graph convolution operation is performed by stacking two graph convolution layers. Consequently, each node acquires an encoded representation x_{gcn}, serving as the output of the GCN and becoming the input for the subsequent spatiotemporal fusion model.

3.2. Spatiotemporal Feature Fusion Module. Modern chemical process data exhibit strong time correlation.[34] To extract more feature information from fault types, considering both the time-domain features and local spatial features of the data is essential. Therefore, this article introduces a temporal and spatial feature fusion module designed to effectively extract temporal and local spatial features from fault data. The module consists of two main parts: the temporal encoding module and the local encoding module.

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

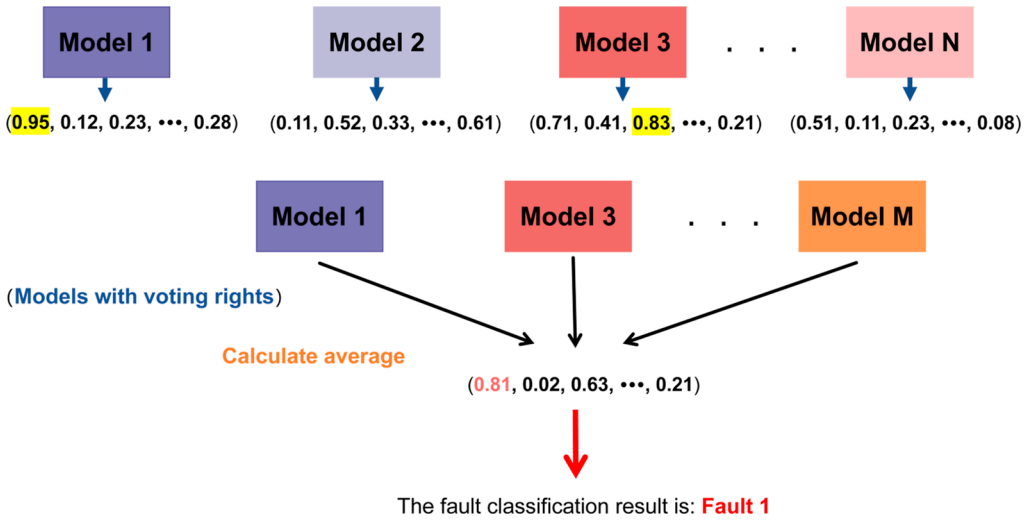


Figure 5. Multi-model voting mechanism.

The temporal encoding module is designed with a set of LSTM units to extract the temporal correlation from sequence features and encode the information. LSTM units are capable of handling data sequences of varying lengths and capturing long-term dependencies. Sequentially capturing time-dependent information, the module outputs the encoded temporal feature representation as x_{time}.

$$C_n = \sigma(x_{\text{gcn}}^{\text{f}})*C_{n-1} + \sigma(x_{\text{gcn}}^{\text{i}})*\tanh(x_{\text{gcn}}) \tag{9}$$

$$x_{\text{time}} = \sigma(x_{\text{gcn}}^{\text{o}})*\tanh(C_n) \tag{10}$$

where C_n is the memory unit, $x_{\text{gcn}}{}^{\text{f}}$ is the forgetting gate, $x_{\text{gcn}}{}^{\text{i}}$ is the input gate, and $x_{\text{gcn}}{}^{\text{o}}$ is the output gate. The tanh activation function is defined as

$$\tanh(x) = \frac{1 - \text{e}^{-2x}}{1 + \text{e}^{-2x}} \tag{11}$$

The local encoding module employs a fully connected layer for the extraction of local spatial features from input feature X. This layer extracts local features of the original signal to provide local spatial information about the data. Through linear mapping, the feature vector of each node is transformed into locally coded features. The local spatial features of a node at the current time step are denoted by x_{local}.

$$x_{\text{local}} = f_{\text{c}}(X) = WX + B \tag{12}$$

where W represents the weight matrix and B represents the bias term.

The fusion of time features and local spatial features results in the output feature x_{fused}. By incorporating both time-domain and spatial-domain features, the final extracted features possess rich spatiotemporal characteristics, enhancing the diagnostic performance of the model. The output features are then input to the fault classifiers to execute the fault classification task. x_{fused} is defined as follows

$$x_{\text{fused}} = \varphi([x_{\text{time}}, x_{\text{local}}]) \tag{13}$$

where φ denotes the feature fusion operation.

3.3. Dual-Supervised Training Strategy. To ensure stable convergence during model training, a dual-supervised training strategy is devised. This strategy involves joint training with two classifiers: an anomaly classifier and a fault diagnosis classifier. Both classifiers utilize multi-layer perceptron (MLP) networks.[35] The MLP classifier is described as a function denoted by f, which maps the input vector x to output vector y

$$y = f(x; \theta) \tag{14}$$

For an L-layer MLP, the function f consists of the following transformations

$$h_1 = \sigma(W_1 x + b_1) \tag{15}$$

$$h_l = \sigma(W_{l-1} h_{l-1} + b_{l-1}) \tag{16}$$

where θ represents all the parameters in the MLP, including the weights w and bias terms b. σ is the activation function such as ReLU.

The anomaly classifier, implemented using a MLP, predicts the abnormal state of the output node, treating it as a binary classification task. The input to the anomaly classifier is the locally encoded feature x_{local}. The predicted anomaly state is generated by the MLP model and denoted as y_{abnorm}, as defined in eq 17. The CrossEntropyLoss objective function is utilized to quantify prediction errors between y_{abnorm} and the truth label $\hat{y}_{\text{abnorm}}$ during the training of this anomaly classifier. Minimizing the CrossEntropyLoss enables the MLP to learn distinguishing patterns in x_{local} indicative of normal or abnormal states.

$$y_{\text{abnorm}} = \text{MLP}_{\text{abnorm}}(x_{\text{local}}) \tag{17}$$

$$\begin{aligned}\text{Loss}_{\text{abnorm}} &= \text{CrossEntropyLoss}(\hat{y}_{\text{abnorm}}, y_{\text{abnorm}}) \\ &= -y_{\text{abnorm}}^{T} \log(f(\hat{y}_{abnorm}; \theta))\end{aligned} \tag{18}$$

The fault diagnosis classifier employs another MLP to predict the fault type of the output nodes, representing it as a multiclass classification output. The input to the fault diagnosis classifier is the aggregated feature, x_{fused}, encompassing both the local encoding feature and the temporal encoding feature. Additionally, the output y_{abnorm} from the anomaly classifier is incorporated, enabling the fault diagnosis classifier to focus on classifying fault types within an abnormal range. The output of the fault diagnosis classifier is denoted as y, as defined in eq 19. To address the common issue of class imbalance in fault

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

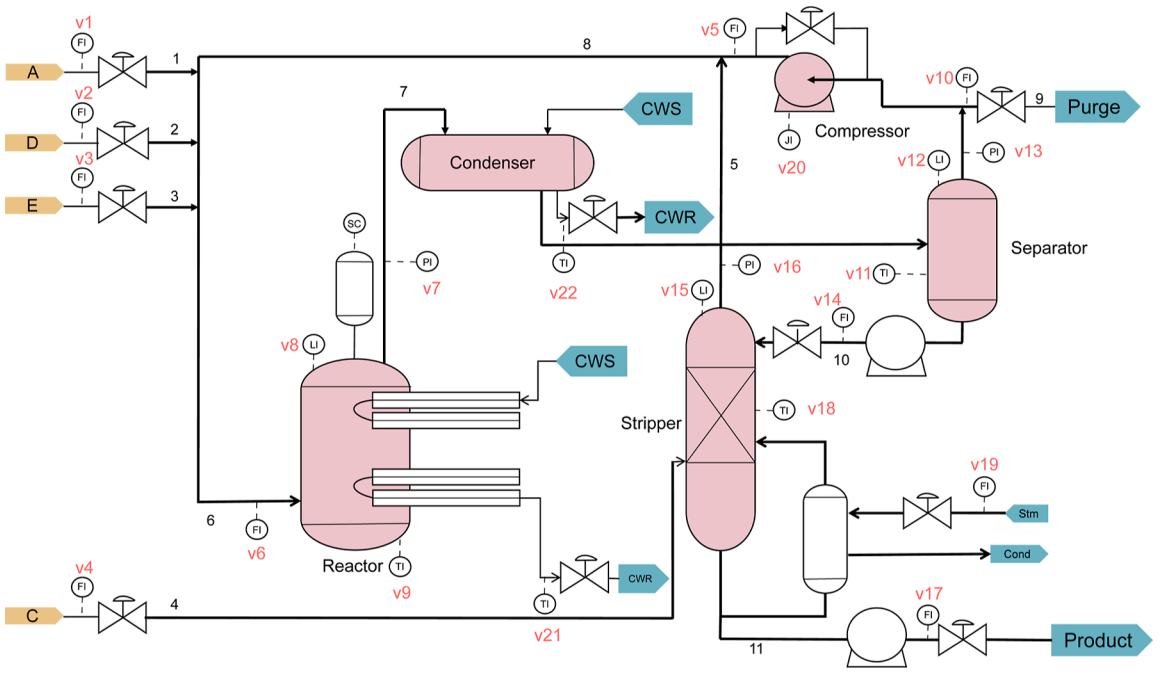


Figure 6. TE process flow.

diagnosis tasks, the FocalLoss function is employed for training the fault diagnosis classifier

$$y = \mathrm{MLP}(x_{\mathrm{fused}} \cdot y_{\mathrm{abnorm}}) \tag{19}$$

$$\mathrm{Loss} = \mathrm{FocalLoss}(\hat{y}, y) = -y^{T}[1 - f(\hat{y}; \theta)]^{\gamma} \log[f(\hat{y}; \theta)] \tag{20}$$

where γ is the focusing parameter that controls the degree of weight reduction for the well-classified samples. The higher the value of γ, the more attention is paid to samples that are difficult to classify.

Employing a dual-supervision strategy, where two classifiers are trained jointly, enhances the accuracy and reliability of model fault diagnosis. The anomaly classifier serves as an initial filter to identify potential anomalies, while the fault diagnosis classifier conducts a finer classification of detected anomalies, thereby achieving an accurate fault diagnosis.

3.4. Multi-Model Voting Inference Strategy. The multi-model voting inference strategy, an ensemble learning approach widely applied in integrated learning, combines predictions from multiple independent models to derive the final prediction. This approach aims to enhance accuracy and minimize errors by leveraging the strengths of models with comparable performance.[36] Employing a soft voting inference mechanism in this strategy utilizes prediction probabilities to improve accuracy.[37] Figure 5 illustrates the multi-model voting strategy, with specific steps outlined below.

- Step 1: integration of multiple models: select multiple independent models for training, considering their distinct features and strengths.
- Step 2: prediction: for a given input sample, input it into each model for prediction, obtaining the predicted probability for each fault.
- Step 3: set confidence threshold: based on a predetermined confidence threshold (set to 0.75 in this case), evaluate the prediction results of each model. Filter out model outputs with confidence levels below the threshold, retaining only valid votes.
- Step 4: determine the final prediction: weight and average the valid votes to obtain the final classification result.

The implementation of the multi-model voting inference strategy effectively leverages the strengths of multiple models, facilitating an accurate evaluation of each model's contribution. This approach ensures the acquisition of more reliable and accurate diagnostic results, overcoming the challenge of reduced diagnostic accuracy associated with prediction bias in individual models.

3.5. Interpretability Analysis. In our study, an investigation of the interpretability of the proposed model was undertaken. It was observed that GCN is highly effective in capturing complex relationships between nodes, rendering it particularly suitable for identifying potential anomaly patterns within a system. Valuable insights into the interactions among nodes are provided by the weight distribution of GCN, enabling an understanding of how the fault features are extracted from the graph structure. Through the weight analysis of GCN, insights into the fault diagnosis process of the model are gained, allowing the identification of nodes crucial for decision-making during model inference and the accurate pinpointing of potential fault sources. This process enhances the precision and credibility of the fault localization.

To objectively assess the relative contributions of nodes in the model, an innovative approach known as node masking experiments was introduced. Node importance is quantified by evaluating the impact of node masking on the accuracy of the model inference. Through this approach, the contribution of each node to fault prediction can be quantitatively measured, thereby identifying key nodes that play a critical role in the overall process.

The identification of key nodes essential for fault prediction can be accurately accomplished by combining the weight analysis of the model with the results of the node masking

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

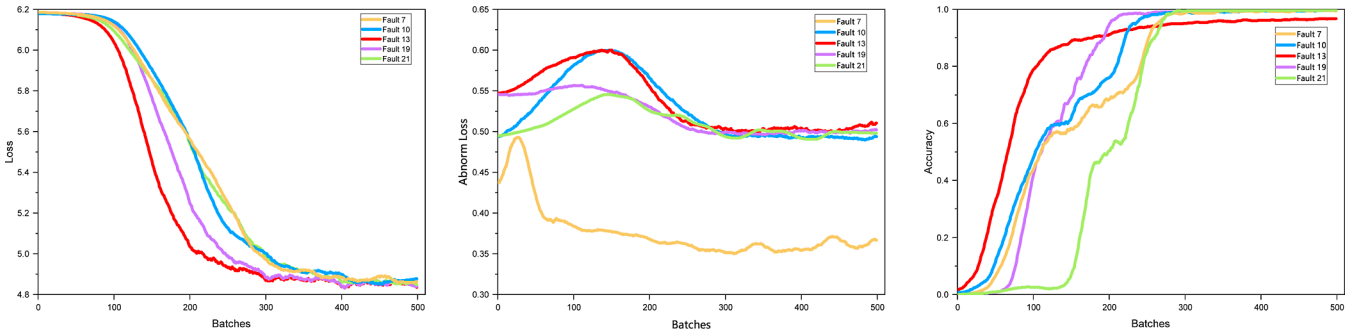


Figure 7. Model test training curves. (a) Change of loss with training batches. (b) Change of abnormal loss with training batches. (c) Change of accuracy rate with training batches.

experiments. This comprehensive analysis contributes to a deeper understanding of the specific causes of faults. Detailed experimental results related to this interpretability aspect can be found in the section Fault Diagnosis Analysis of the article.

4. CASE STUDY

4.1. Simulation Setup. The TE process is based on a chemical plant of Eastman Chemical Company in Tennessee, USA, and it closely resembles the actual production process.[20] The process (Figure 6) involves eight components, labeled A–H, where component B is an inert substance that does not participate in the reactions. The reactions in the process are all irreversible exothermic reactions,[38] and the reaction equations are as follows

$$\begin{aligned} &A(g) + C(g) + D(g) \rightarrow G(liq)\\ &A(g) + C(g) + E(g) \rightarrow H(liq)\\ &A(g) + E(g) \rightarrow F(liq)\\ &3D(g) \rightarrow 2F(liq) \end{aligned} \tag{21}$$

where (g) and (liq) represent the gas phase and liquid phase, respectively. Components A, C, D, and E are the main raw materials, while G and H are the main products. Component F is a byproduct.

The TE process dataset used in the study includes a total of 41 measurement variables, 12 control variables, and 21 predetermined faults. Among the measurement variables, XMEAS(1)-XMEAS(22) represent the continuous measurement variables of the process. IDV(1–15) and IDV21 correspond to faults with known types, while IDV(16–20) represents faults with unknown types. It is worth mentioning that except for fault 6, the training set samples with faults were obtained from a 25 h simulation, resulting in 480 observations. The test set samples with faults were obtained from a 48 h simulation, resulting in a total of 960 observations. Fault 6 caused the machine to shut down after 7 and 14 h, and its training and test sets produced 140 and 280 samples, respectively. Due to the long intermittent analysis variables in the TE process dataset, which have long sampling intervals and cannot effectively capture continuous changes in process states, the study only utilizes the 22 continuous measurement variables (XMEAS(1)-XMEAS(22)) as nodes for analysis.

4.2. Dataset Processing. In the dataset processing phase, a sliding window approach is employed for data collection. During the training phase, the sampling quantity is set to 100. Additionally, the input data undergoes normalization. The normalization process involves normalizing the data along the time dimension for each of the 100-time steps. Mean and variance calculations are performed on the input data for each dimension to achieve normalization across all dimensions.

4.3. Performance of the Model in Fault Diagnosis. To validate the effectiveness of the proposed PG-STF fault diagnosis model, it is trained on a training dataset and evaluated on a separate test dataset. Two evaluation metrics are employed to assess the model's performance: classification accuracy (ACC) and fault diagnosis rate (FDR). ACC measures the model's ability to correctly predict the type of faults in the samples, providing an overall measure of the model's effectiveness in terms of correctly classifying faults. FDR evaluates the model's ability to correctly identify specific types of faults, indicating how well the model can accurately diagnose and classify different fault types

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{22}$$

$$\mathrm{FDR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{23}$$

where TP represents true positives, which are the number of samples correctly classified as positive. TN represents true negatives, which are the number of samples correctly classified as negative. FP represents false positives, which are the number of samples incorrectly classified as positive. FN represents false negatives, which are the number of samples incorrectly classified as negative.

Figure 7 illustrates the trend of the loss value, abnormal loss value, and accuracy rate of the PG-STF model during the training process. The loss and abnormal loss values exhibit a gradual decrease and eventual stabilization as the number of training batches increases. This indicates that the model has effectively learned the relevant features in the data and improves its ability to minimize the discrepancy between predicted and actual values. Concurrently, the accuracy rate shows a gradual increase, approaching a value of 1 as the training progresses. This signifies that the model has become highly proficient in accurately classifying faults. The ascending accuracy rate further demonstrates the model's capacity to make correct predictions and classify fault types with a notable level of precision. These results affirm the effectiveness and performance of the model. Moreover, the introduction of the abnormal loss function contributes to the improved convergence speed of the model, which is crucial for the efficient fault diagnosis in practical applications.

To evaluate the efficacy of the proposed PG-STF model for fault diagnosis, the test dataset is input into the model, with the

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486–9502

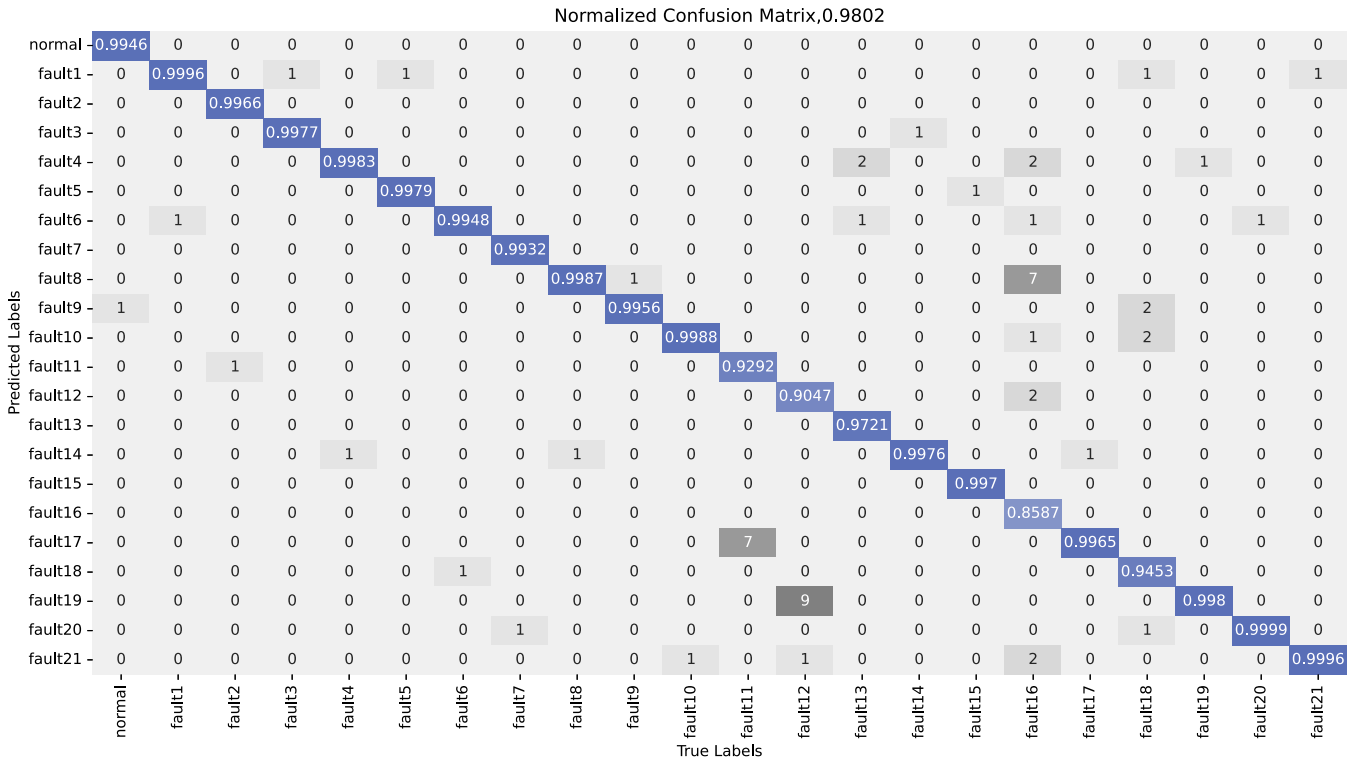


Figure 8. Confusion matrix for testing dataset.

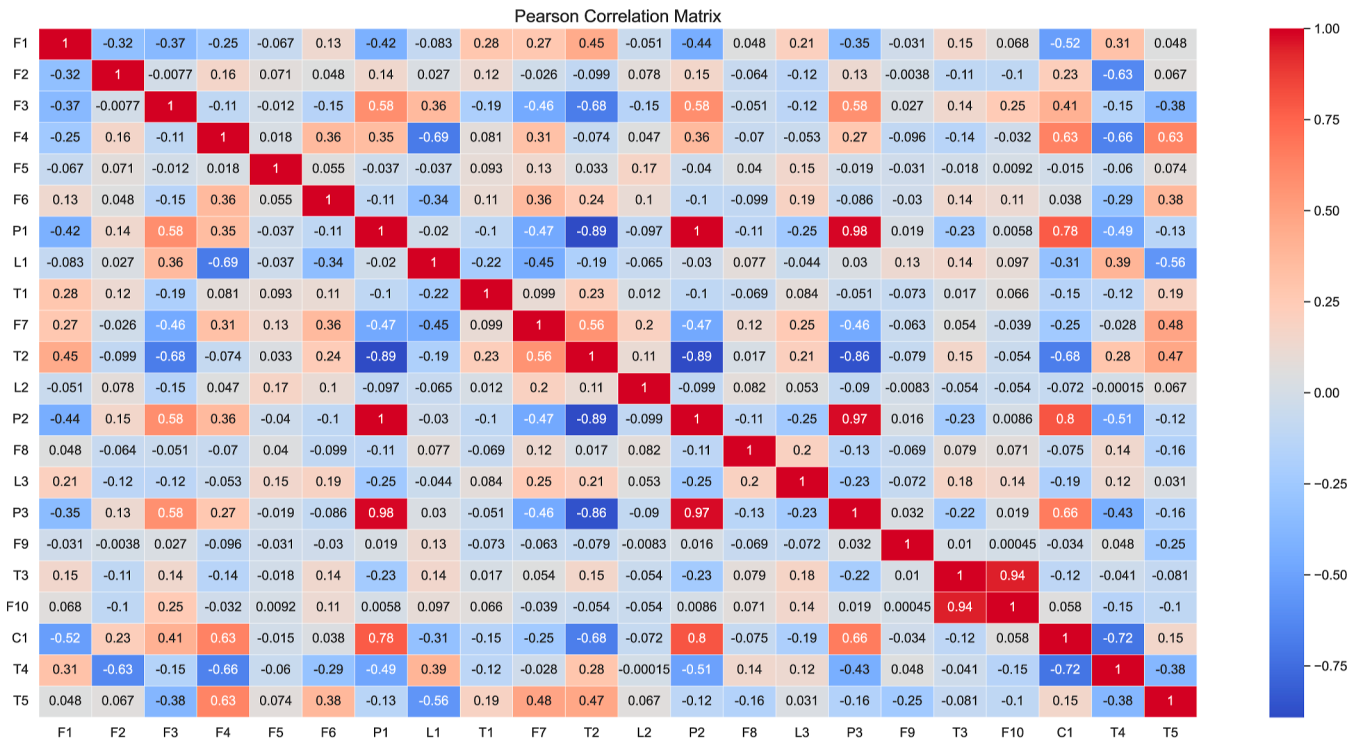


Figure 9. Fault 16 Pearson correlation coefficient.

confusion matrix employed as the assessment metric. As depicted in Figure 8, the confusion matrix provides a visual representation of the model's performance by showing the relationship between the actual fault labels and the predicted fault labels.[39] In the confusion matrix, the horizontal axis represents the actual labeling of the fault types, while the vertical axis represents the predicted labeling. The values on the diagonal represent the diagnostic accuracy of the PG-STF model for each fault type. The values off the diagonal represent the number of diagnostic errors in the PG-STF model for each fault type. Taking fault 12 as an example, in 100 diagnostic tasks, the model's correct diagnostic rate for fault 12 is 0.9047, with 9 times misdiagnosing fault 12 as fault 19 and 1 time misdiagnosing fault 12 as fault 21. Analyzing the confusion

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

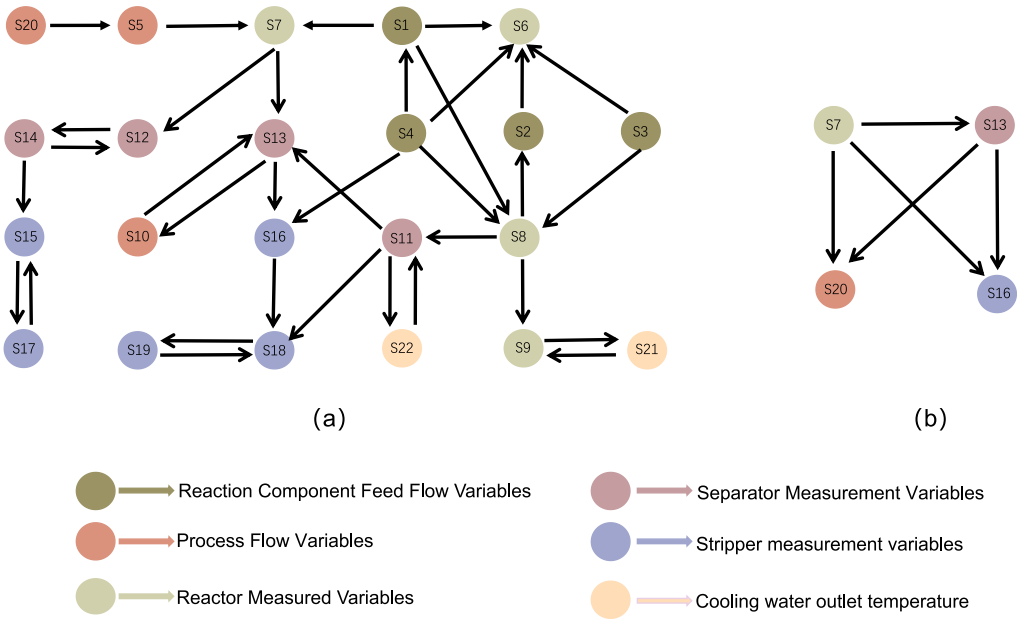


Figure 10. (a) TE process for continuous measurement of variables. (b) Process of calculating the Pearson correlation coefficient for fault 16.

Table 1. FDR Comparison of Fault Diagnosis Results

fault type	DCNN[17]	BiGRU[18]	CGN[41]	PTCN[42]	target transformer[43]	MEWMA-PCA-BM44	PG-STF
normal	0.978	0.969	0.985	0.9924		0.55	**0.9946**
IDV 1	0.986	0.986	0.975	0.9931	0.9975	0.90	**0.9996**
IDV 2	0.985	0.972	0.980	0.9819	0.9844	0.91	**0.9966**
IDV 3	0.917	0.935		0.935	0.9938	0.10	**0.9977**
IDV 4	0.976	0.974	0.824	0.9956	0.9962	0.89	**0.9983**
IDV 5	0.915	**0.998**	0.980	0.9786	0.9188	0.93	0.9979
IDV 6	0.975	**1**	**1**	**1**	0.9821	0.91	0.9948
IDV 7	0.999	**1**	**1**	**1**	0.9994	0.90	0.9932
IDV 8	0.922	0.753	0.966	0.9160	0.9556	0.88	**0.9987**
IDV 9	0.584	0.807		0.6601	0.6869	0.09	**0.9956**
IDV 10	0.964	**1**	0.881	0.9276	0.9769	0.45	0.9988
IDV 11	**0.984**	0.965	0.778	0.9798	0.9806	0.80	0.9292
IDV 12	0.956	0.961	**0.981**	0.9704	0.9706	0.89	0.9047
IDV 13	0.957	0.953	0.758	0.8969	0.9621	0.87	**0.9721**
IDV 14	0.987	0.996	0.986	0.9964	0.9875	0.51	**0.9976**
IDV 15	0.28	0.541		0.0035	0.3406	0.10	**0.9970**
IDV 16	0.442	0.788	0.814	**0.9685**	0.5269	0.63	0.9517
IDV 17	0.945	0.97	0.848	0.9254	0.9475	0.87	**0.9965**
IDV 18	0.939	0.923	0.685	0.9049	0.9425	0.82	**0.9453**
IDV 19	0.986	0.926	0.964	0.9650	0.9869	0.13	**0.9980**
IDV 20	0.933	0.981	0.871	0.8825	0.9425	0.49	**0.9999**
IDV 21							**0.9996**
avg	0.882	0.927		0.9392	0.9039	0.648	**0.9844**

matrix demonstrates that the PG-STF achieves an accuracy of 0.99 or more for 16 out of 21 fault types, 17 exceed an accuracy of 0.95, 20 achieve an accuracy of 0.90 or more, and only Fault 16 has an accuracy of less than 0.90. These results demonstrate the efficiency and reliability of the PG-STF model in fault diagnosis.

PG-STF initially achieves an average accuracy of 0.9795 on 21 fault types, but its performance on Fault 16 falls short of expectations, with an accuracy of only 0.8587. To address this issue and improve the model's overall performance, an optimization technique is implemented specifically for the adjacency matrix of the model inputs. The original adjacency matrix is constructed based on the knowledge of the chemical process structure. However, in order to enhance its effectiveness, a data-driven approach is employed. This approach involves calculating the Pearson correlation coefficient between different variables and using it as a weight to construct the adjacency matrix.

The Pearson correlation coefficient is a statistical measure that gauges the strength of the linear relationship between two variables.[40] It ranges from −1 to 1, where −1 signifies a completely negative correlation, 0 indicates no linear correlation, and 1 indicates a completely positive correlation. For a time series data input matrix $X = (x_1, x_2, ..., x_m)$, the

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

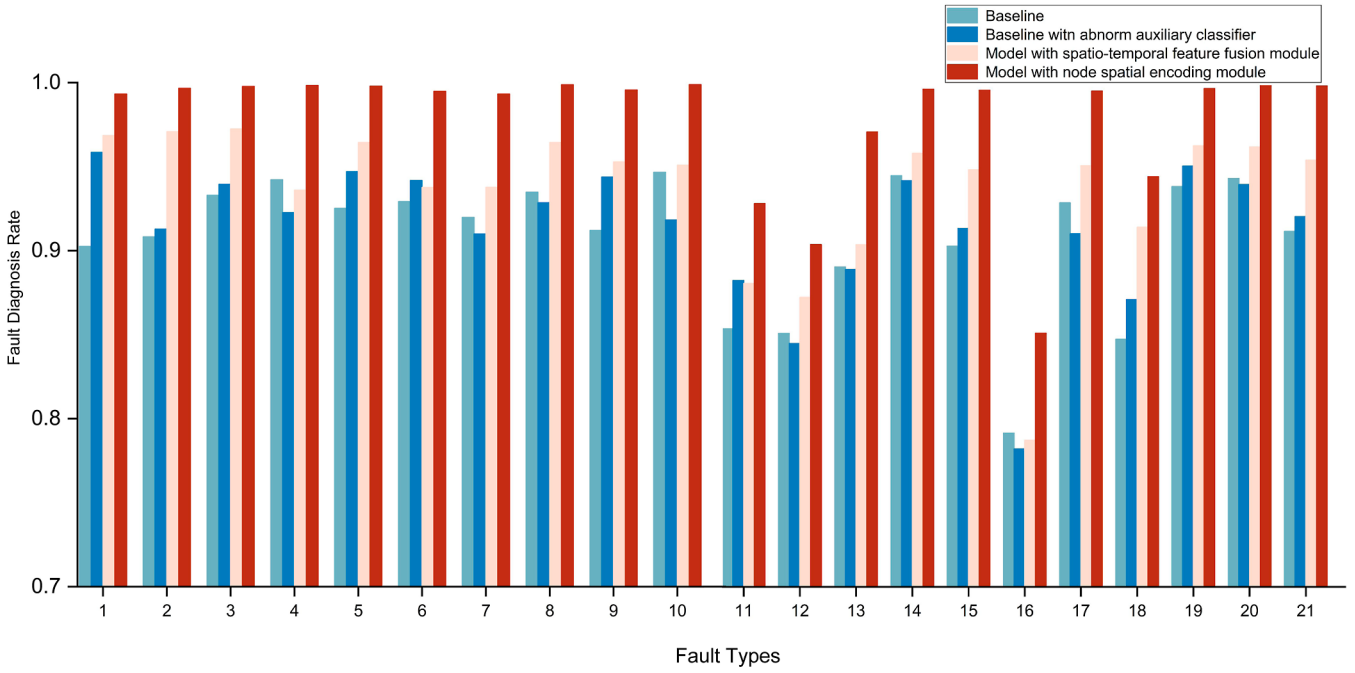


Figure 11. Comparison of ablation result.

Pearson correlation coefficient is the quotient of the covariance and standard deviation of the two features

$$\rho = \frac{E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]}{\sigma_{X_i}\sigma_{X_j}} = \frac{E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]}{\sqrt{\sum_{i=1}^{n}(X_i - \mu_{X_i})^2}\sqrt{\sum_{j=1}^{n}(X_j - \mu_{X_j})^2}} \quad (24)$$

where μ_{X_i} and μ_{X_j} represent the means of features X_i and X_j, respectively, and σ_{X_i} and σ_{X_j} represent their standard deviations.

The optimization process of the adjacency matrix involves calculating the Pearson correlation coefficients between 22 continuously measured variables using the time-series data of Fault 16. The resulting correlation coefficient matrix is depicted in Figure 9. In the figure, variable F is the flow rate, variable P is the pressure, variable L is the liquid level, variable T is the temperature, and variable C is the compressor power. By setting an appropriate threshold, it can be determined which variables exhibit sufficiently strong correlations and are included in the adjacency matrix. Figure 10 illustrates the difference between the two methods of adjacency matrices: the original matrix based on knowledge of the chemical process structure and the data-driven correlation matrix. To leverage both sources of information, the two matrices are fused together. This fusion process aims to retain the information from the original matrix while highlighting the similarities between the two matrices. The resulting optimized adjacency matrix is then utilized as an input for the PG-STF model. With the newly constructed adjacency matrix, the accuracy of the PG-STF model in predicting Fault 16 significantly improves from 0.8587 to 0.9517. Moreover, the average accuracy of the PG-STF model increases from 0.9802 to 0.9844. These results demonstrate the effectiveness and feasibility of optimizing the adjacency matrix to enhance the model's performance.

The experimental results presented in Table 1 demonstrate the performance of the PG-STF model for fault diagnosis and compare it with deep models based on CNN, RNN, and GCN, as well as the PTCN model that utilizes graph structures. In addition, comparisons were made with a modified transformer model called target transformer, and a comprehensive framework (MEWMA-PCA-BM) combining multivariate exponentially weighted moving average PCA and Bayesian methods. According to the results, the PG-STF model achieves a higher FDR compared to that of other models under normal conditions. A higher FDR indicates that the PG-STF model effectively reduces false alarms, which can improve operators' trust in the fault diagnosis system. Furthermore, the PG-STF model outperforms the baseline model in 14 types of fault diagnoses. This improvement can be attributed to the incorporation of prior knowledge about the TE process, which helps the model make sense of the learning process and enhances the diagnostic performance across various fault types. For example, for faults 9 and 15, which are two types of faults that are difficult to distinguish from the normal state, the PG-STF model achieves high accuracy rates of 0.9956 and 0.9970, respectively. However, the accuracy of the model in addressing faults 11 and 12 is slightly lower compared to that of the benchmark models. A detailed analysis of the confusion matrix reveals that this discrepancy originates from the tendency of the PG-STF model to misclassify fault 11 as unknown fault 17 and fault 12 as unknown fault 19. It is evident that in the pursuit of enhancing the accuracy for faults 17 and 19, a discernible trade-off exists, resulting in a modest reduction in accuracy for faults 11 and 12. Furthermore, it should be noted that the model has not yet achieved a perfect accuracy of 1, despite attaining notable accuracy on several faults. Overall, the PG-STF model achieves the highest average accuracy of 0.9844 among all of the compared models. These results highlight the significant advantages of the PG-STF model in fault diagnosis

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

tasks, showcasing its superior accuracy in identifying and classifying different types of faults.

4.4. Ablation Experiments. The key components of the PG-STF model include the addition of a node spatial encoding module, a spatiotemporal feature fusion module, and the introduction of an anomaly assisted classifier during the training process. These components play a crucial role in improving the model's performance for fault diagnosis tasks. To evaluate the effectiveness of these components, ablation experiments are conducted. These experiments involve comparing the FDR and ACC of 21 kinds of faults under different models. Figure 11 illustrates the FDR results for the four tested models, while Table 2 presents the corresponding ACC values.

Table 2. Results of the Ablation Experiment

ID	baseline	abnorm auxiliary classifier	spatiotemporal feature fusion module	node spatial encoding module	ACC
1	√				0.9079
2	√	√			0.9131
3	√	√	√		0.9362
4	√	√	√	√	0.9795

The whole experiment is as follows. Initially, a model relying only on basic temporal encoding is tested but shows poor performance, indicating that temporal encoding alone is insufficient for an accurate fault diagnosis. To overcome this limitation, an anomalous auxiliary classifier is introduced into the temporal encoding model. The addition of an anomaly classifier helps the model converge faster. The ACC value of the model is improved to 0.9131, which indicates a significant enhancement in the fault diagnosis capability. Then, the spatiotemporal feature fusion module is incorporated into the model. The module combines temporal encoding with local spatial encoding, enabling the model to capture both temporal correlations and spatial features. The results show that the integration of the spatiotemporal feature fusion module improves the ability of the model to diagnose faults accurately, and the ACC value of the model further increases to 0.9362. Finally, the model integrates the node spatial encoding module to learn node features and capture spatial relationships using GCN. This integration significantly improves the performance of the model, achieving the highest accuracy of 0.9795 among all of the tested models.

The results of ablation experiments show that each component plays an important role in improving the performance of the model, and these improvements contribute to improving the diagnostic accuracy of the model.

4.5. Fault Diagnosis Analysis. A fault diagnosis analysis method based on a node mask is designed, and the fault diagnosis results are analyzed and interpreted from the model's perspective. First, node representation vectors are used to identify variables contributing to the faults and analyze the root causes. Second, the importance of these variables in the fault diagnosis inference process is assessed through node masking experiments.

4.6. Fault Result Analysis. The lack of interpretability in deep learning-based fault diagnosis models is a common challenge due to the complexity and large number of network parameters. It often becomes difficult to accurately localize the cause of a fault involving specific variables, leading to a lack of trust in the model's results.

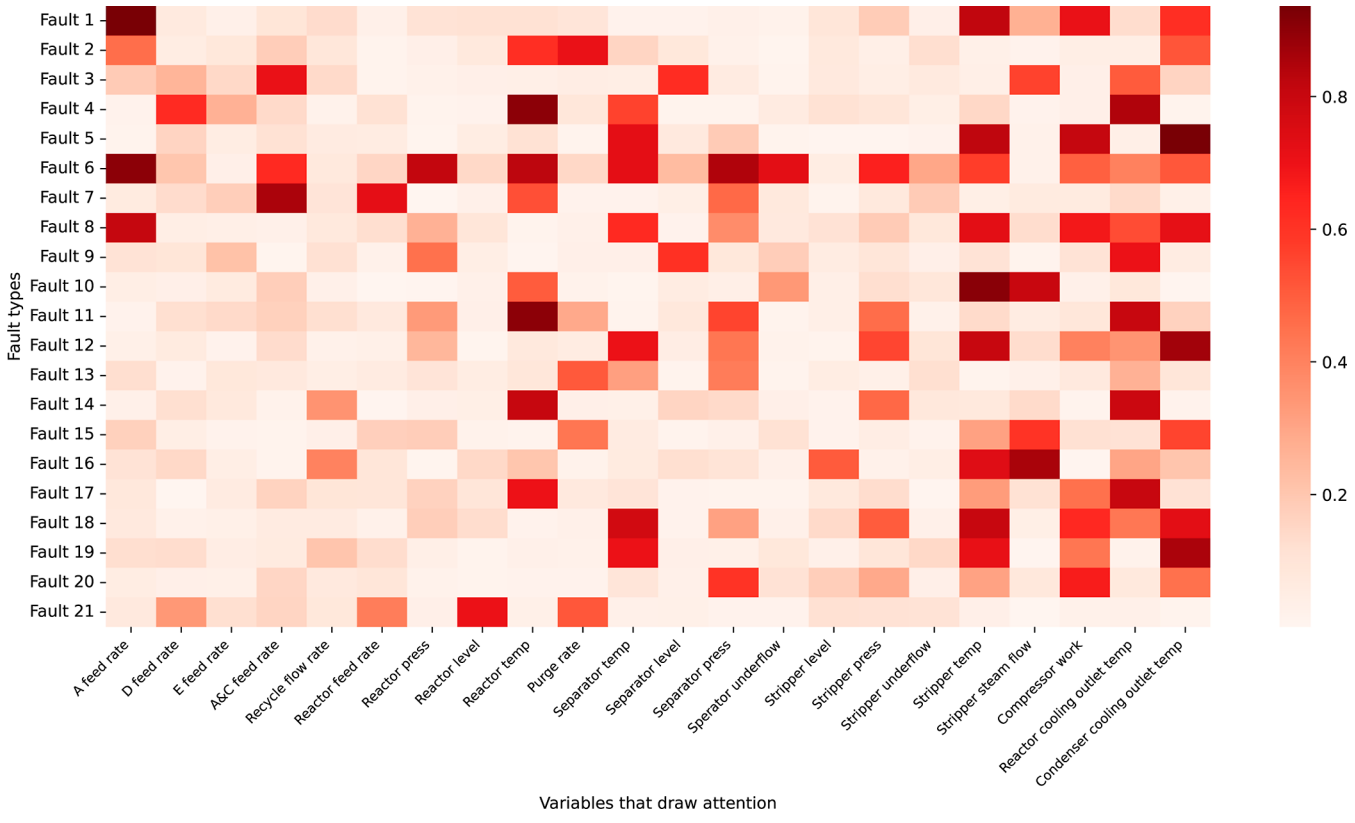


Figure 12. TE process 21 kinds of fault node representation weight heatmap.

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

To address this issue, an interpretable fault diagnosis analysis method is designed that focuses on locating critical nodes responsible for system faults. In this method, the importance of system nodes in fault diagnosis is estimated by utilizing the weights assigned to nodes in the shallow GCN weights in the spatial encoding module. Specifically, larger weights assigned to a node indicate a stronger influence on fault diagnosis and suggest a higher probability of that node being critical in the fault diagnosis process

$$H_1 = \sigma(AXW_1) \tag{25}$$

where $X \in R_{22}$ represents the input data and 22 denotes the number of nodes. $A \in R_{22\times22}$ represents the adjacency matrix, $W_1 \in R_{22\times64}$ represents the network weight of the first GCN layer, $H_1 \in R_{22\times64}$ represents the feature output after encoding the first GCN layer, and σ is the activation function.

Aggregation and normalization operations are performed on the GCN weight matrix, resulting in the generation of a node representation vector denoted as W_{norm}

$$W_{\text{norm}} = \sigma\left(\frac{1}{N}\sum_{j=1}^{N} W_1\right) \tag{26}$$

where N represents the total number of nodes and σ represents the activation function.

The visualization of node representation vector output values as a heatmap offers an intuitive understanding of the variables' features, which is crucial for fault diagnosis. By sorting the output values of all nodes, the top 4 nodes are selected as the most important ones for further analysis. This approach facilitates the explanation of fault causes and provides reliable and interpretable fault diagnosis results.

Figure 12 depicts a heatmap of the node representation weights for 21 different faults. Table 3 enumerates the top 4 variables with the highest weights from Figure 12. The first column delineates the fault type, the second column furnishes a fault description, and the third column presents the four variables with the highest extracted weights. Chemical process faults can originate from various underlying reasons, impacting different process variables in diverse ways. This implies that during the fault localization process, the model may extract multiple process variables as fault-related variables, exhibiting certain correlations. Through the analysis, the model successfully localizes the fault to the relevant continuous measurement variables in the TE process.

Fault 1 corresponds to a change in the A/C feed flow rate, which directly affects the feed flow rate of reactant A. The model successfully extracts this variable as a fault-related variable. Fault 4 is characterized by a change in the cooling water inlet temperature of the reactor. From Table 3, it can be observed that this fault directly impacts the reactor temperature and the outlet temperature of the cooling water. An increase in the cooling water inlet temperature leads to changes in the reactor outlet temperature, the product temperature of the separator, and the separator temperature. The concentration of reactant D and its closely related variables are significantly affected by this fault. Fault 12 has its root cause in random variations of the cooling water inlet temperature to the condenser. The model successfully extracts the cooling water inlet temperature from the condenser as a fault-related variable. This type of fault directly affects the cooling efficiency of the condenser. Fault 14 involves the sticking of the reactor cooling water valve, which is crucial for controlling the flow rate of cooling water to maintain the reactor temperature. When the valve sticks, the cooling water flow rate cannot be properly adjusted, resulting in an unstable reactor temperature and outlet temperature of the cooling water. Fault 16 is an unknown fault with an unknown root cause. Based on the top 4 variables with the highest weights listed in Table 3, it can be inferred that this fault primarily affects the operation state of the distillation column.

Table 3. TE Process 21 Kinds of Fault Node Representation Vectors[a]

fault	fault description	top 4 variables with largest weight
1	A/C feed flow ratio changes	$Q_{(\text{A feed})}$, $T_{(\text{stripper})}$, $W_{(\text{compressor})}$, $T_{(\text{condenser cooling})}$
2	B composition changes (A,C feed)	$Q_{(\text{purge})}$, $T_{(\text{reactor})}$, $T_{(\text{condenser cooling})}$, $Q_{(\text{A feed})}$
3	D feed temperature changes	$Q_{(\text{A,C feed})}$, $L_{(\text{separator})}$, $Q_{(\text{stripper steam})}$, $T_{(\text{reactor cooling})}$
4	reactor cooling inlet temperature changes	$T_{(\text{reactor})}$, $T_{(\text{reactor cooling})}$, $Q_{(\text{D feed})}$, $T_{(\text{separator})}$
5	D feed temperature changes randomly	$T_{(\text{condenser cooling})}$, $T_{(\text{stripper})}$, $W_{(\text{compressor})}$, $T_{(\text{separator})}$
6	A feed loss	$Q_{(\text{A feed})}$, $P_{(\text{separator})}$, $P_{(\text{reactor})}$, $T_{(\text{separator})}$
7	C feed header pressure loss-reduced availability	$Q_{(\text{A,C feed})}$, $Q_{(\text{reactor})}$, $T_{(\text{reactor})}$, $P_{(\text{separator})}$
8	A, B, C feed composition changes randomly	$Q_{(\text{A feed})}$, $T_{(\text{condenser cooling})}$, $W_{(\text{compressor})}$, $T_{(\text{stripper})}$
9	D feed temperature changes randomly	$T_{(\text{reactor cooling})}$, $L_{(\text{separator})}$, $P_{(\text{reactor})}$, $Q_{(\text{E feed})}$
10	C feed temperature changes randomly	$T_{(\text{stripper})}$, $Q_{(\text{stripper steam})}$, $T_{(\text{reactor})}$, $Q_{(\text{separator})}$
11	reactor cooling water inlet temperature changes randomly	$T_{(\text{reactor})}$, $T_{(\text{reactor cooling})}$, $P_{(\text{separator})}$, $P_{(\text{stripper})}$
12	condenser cooling inlet temperature changes randomly	$T_{(\text{condenser cooling})}$, $T_{(\text{stripper})}$, $T_{(\text{separator})}$, $P_{(\text{stripper})}$
13	reaction kinetics drift slowly	$Q_{(\text{purge})}$, $P_{(\text{separator})}$, $T_{(\text{separator})}$, $T_{(\text{reactor cooling})}$
14	reactor cooling water valve sticking	$T_{(\text{reactor})}$, $T_{(\text{reactor cooling})}$, $P_{(\text{stripper})}$, $Q_{(\text{recycle})}$
15	condenser cooling water valve sticking	$Q_{(\text{stripper steam})}$, $T_{(\text{condenser cooling})}$, $Q_{(\text{purge})}$, $T_{(\text{stripper})}$
16	unknown fault	$Q_{(\text{stripper steam})}$, $T_{(\text{stripper})}$, $L_{(\text{stripper})}$, $Q_{(\text{recycle})}$
17	unknown fault	$T_{(\text{reactor cooling})}$, $T_{(\text{reactor})}$, $W_{(\text{compressor})}$, $T_{(\text{stripper})}$
18	unknown fault	$T_{(\text{stripper})}$, $T_{(\text{separator})}$, $T_{(\text{condenser cooling})}$, $W_{(\text{compressor})}$
19	unknown fault	$T_{(\text{condenser cooling})}$, $T_{(\text{stripper})}$, $T_{(\text{separator})}$, $W_{(\text{compressor})}$
20	unknown fault	$W_{(\text{compressor})}$, $P_{(\text{separator})}$, $T_{(\text{condenser cooling})}$, $T_{(\text{stripper})}$
21	the valve for flow 4 is fixed in a steady-state position	$L_{(\text{reactor})}$, $Q_{(\text{purge})}$, $Q_{(\text{reactor})}$, $Q_{(\text{D feed})}$

[a] Q, T, W, L, and P are flow, temperature, power, liquid level, and pressure signals, respectively. Subscripts represent unit operations or streams.

Analyzing the top 4 variables with the highest weights for each fault enables the model to successfully identify the variables most affected by the faults. These variables offer valuable insights into the causes and impacts of the faults, contributing to a more profound understanding of the fault diagnosis process.

4.7. Fault Analysis Verification. The node masking experiments are designed to simulate the absence or unavailability of nodes during the inference process, aiming

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

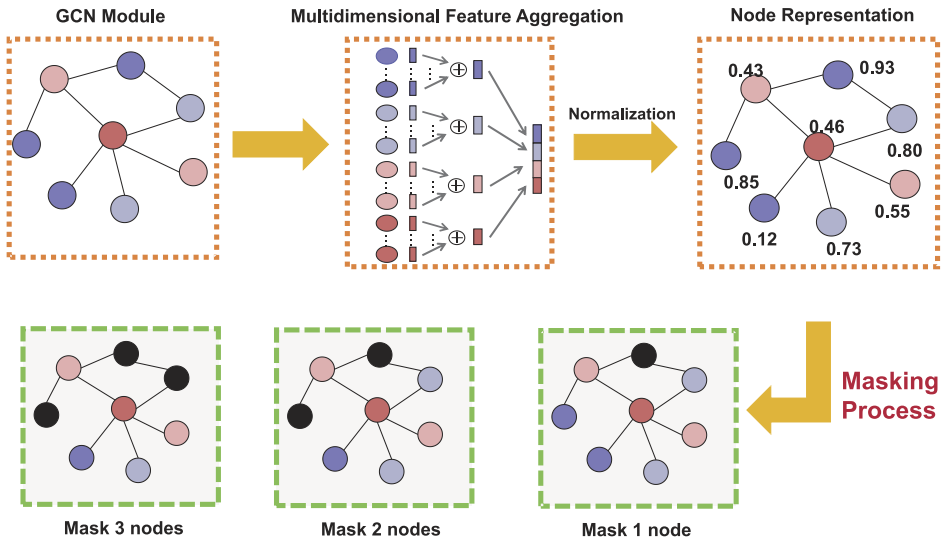


Figure 13. Node masking process.

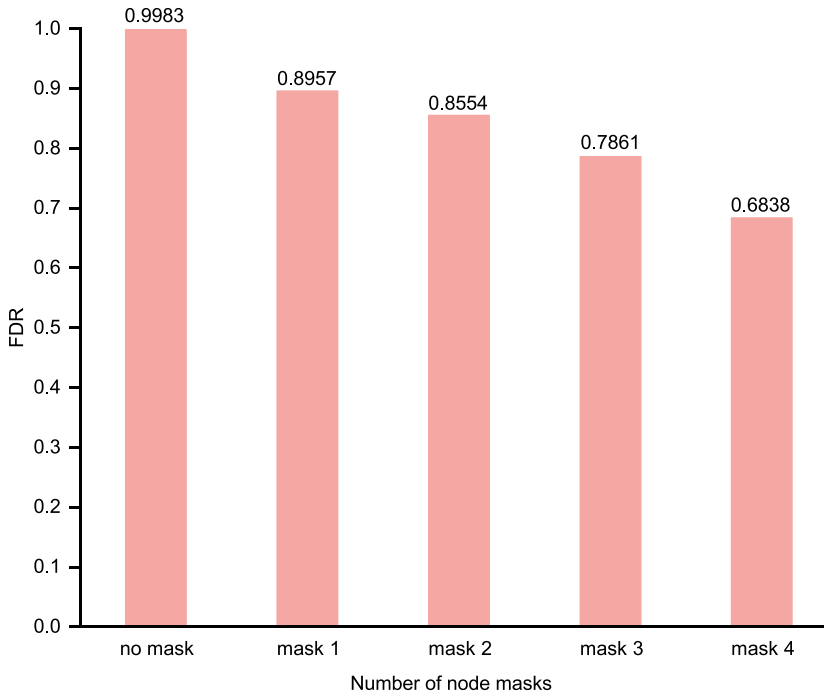


Figure 14. Fault 4 node masking FDR results.

to assess the impact of nodes on the model's inference. This is achieved by masking the nodes, setting their data value to 0, and observing the resulting performance changes in the model. Figure 13 provides an illustration of the node masking process. Table 3 presents an analysis of the top four variables with the highest weights for each fault in the TE process. To evaluate the importance of these variables, node masking experiments were conducted to observe the resulting changes in the model's diagnostic accuracy. In particular, fault 4, which involves changes in the reactor cooling water inlet temperature, was selected for detailed analysis. The model identified the following four variables as being most affected by this fault: reactor temperature (T1), reactor cooling water outlet temperature (T4), component D feed flow rate (F2), and product separator temperature (T2), as shown in Table 3. Subsequently, node masking experiments were carried out on these four variables individually to simulate the scenario where important information about them is lost during the model's inference process. The PG-STF diagnostic model was reexecuted after each masking to assess its diagnostic performance, and the outcomes are depicted in Figure 14. When masking the first variable T1, the diagnostic rate of the model decreased by 10.26%. Masking the first two variables T1 and T4 resulted in a 14.29% decrease in the model's diagnostic rate. Similarly, masking the first three variables T1, T4, and F2 led to a 21.22% decrease in the model's diagnostic rate. Finally, when all four variables were masked, the model's accuracy decreased by 31.45%. By comparing these results with the performance of the original model in diagnosing fault 4, the impact of node masking on the model's diagnostic performance can be observed. The significant decrease in diagnostic performance due to node masking implies that the masked

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

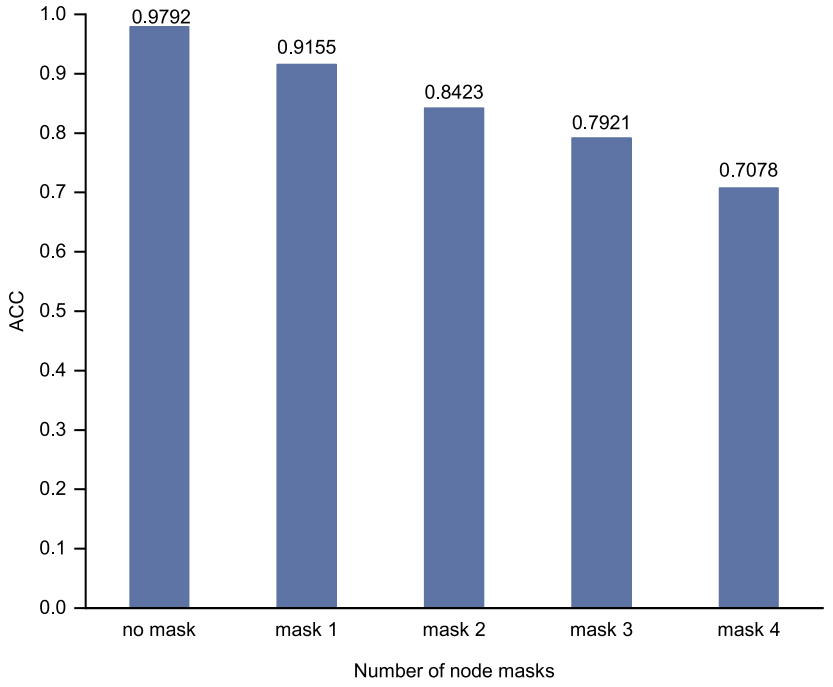


Figure 15. Twenty-one fault type node masking ACC results.

nodes play a crucial role in the fault diagnosis process, and the model effectively utilizes these nodes for accurate diagnosis.

The experimental findings depicted in Figure 15 reveal that as the first variable with the highest weight, the first two variables, the first three variables, and the first four variables are sequentially masked, the average accuracy of the model for diagnosing the 21 faults decreases by 6.37, 13.69, 18.71, and 27.16%, respectively. These experimental results emphasize a significant reduction in model accuracy as the number of masked nodes increases, further affirming the pivotal role of these nodes in the system.

Important nodes are obtained by visualizing the fault node representation vector. Subsequently, the important nodes undergo masking experiments to explore the contribution of key variables to fault diagnosis. The experimental results described above demonstrate that the fault diagnosis inference process can be analyzed from a model perspective using the proposed method, thereby better revealing the root cause of fault occurrence.

5. RESULTS AND DISCUSSION

In this section, a comprehensive evaluation of the performance of the proposed PG-STF model for chemical process fault diagnosis is provided, considering both its strengths and limitations. The selected inference mechanism for the model is the strategy of multi-model voting inference, offering the advantage of enhancing overall robustness and accuracy through the amalgamation of opinions from multiple models. However, it introduces challenges, such as the need to coordinate the training and cooperation of multiple models. Future research endeavors will focus on optimizing this inference framework, aiming to further enhance the model's effectiveness. An additional aspect to underscore is the deliberate focus on training and testing the proposed model specifically in TE chemical processes. This emphasis arises from the recognition that chemical systems in real-world applications may exhibit variations in process parameters, operating conditions, or equipment configurations. Future work will strive to broaden the applicability of the model, making it more generalizable and adaptable to diverse types of chemical systems. This expansion may involve incorporating a wider collection of datasets and implementing more sophisticated model adaptations.

It is important to note that the current research phase has already yielded encouraging results for the PG-STF model. It has demonstrated its capability to effectively capture anomalous patterns in chemical systems, serving as a robust tool for fault diagnosis. Nevertheless, additional efforts are imperative to address the challenges in practical applications, ensuring the reliability and usefulness of the model. Ongoing work will focus on refining the model and overcoming these challenges to broaden its applicability in real-world scenarios.

6. CONCLUSIONS

This paper has presented a new fault diagnosis method PG-STF for the chemical process, which integrates a node spatial encoding module with a spatiotemporal feature fusion module. The spatial encoding module, based on GCN, is utilized to extract features from the spatial perspective of chemical processes. The adjacency matrix is constructed by combining a priori knowledge of the chemical process with the Pearson correlation, taking into account the physical correlation between the nodes. The spatiotemporal feature fusion module based on the LSTM network extracts features from the time perspective to capture the time dependence of fault data. To ensure stable convergence of the multiclassification fault diagnosis model, a double-supervised training strategy is designed. During the model's inferring process, a multi-model voting inference strategy is employed to enhance the accuracy and robustness of the diagnosis by leveraging multiple models. Additionally, a fault diagnosis analysis method based on node masking is developed to identify the key variables that the model focuses on during the fault diagnosis process. Experimental results on the TE process demonstrate the

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502

effectiveness of the PG-STF model, achieving an average fault diagnosis rate of 0.9844 across all fault types, indicating a strong diagnostic performance. For each type of fault, the model successfully identifies several key variables, displaying a strong physical correlation with the underlying cause.

The integration of data-driven analyses with mechanism exploration remains pivotal for comprehending fault propagation relationships in chemical systems. Nevertheless, despite the theoretical advancements of the proposed approach, practical implementation encounters challenges. Specifically, further in-depth investigation is required for fault propagation paths and potential mechanisms. In our subsequent work, efforts will be directed toward addressing these challenges and delving deeper into the system's complexity to achieve a more comprehensive and profound understanding of fault propagation mechanisms in chemical systems.

AUTHOR INFORMATION

Corresponding Authors

Dazi Li − *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;* orcid.org/0000-0003-1610-6558; Email: lidz@mail.buct.edu.cn

Yang Zhang − *College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China;* orcid.org/0009-0002-4496-0460; Email: 2002500011@mail.buct.edu.cn

Authors

Fengzhen Zhang − *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*

Qibing Jin − *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*

Qian Zhu − *Jiangsu Academy of Chemical Inherent Safety, Jiangsu 210009, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c09122

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (no. 62273026) and the construction project of the Jiangsu Academy of Chemical Inherent Safety (BM 2021805).

REFERENCES

(1) Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S. N. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27*, 293−311.

(2) 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*, 513−553.

(3) Abid, A.; Khan, M. T.; Iqbal, J. A review on fault detection and diagnosis techniques: basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639−3664.

(4) Chen, H.; Jiang, B.; Ding, S. X.; Huang, B. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Trans. Intell. Transport. Syst.* **2022**, *23*, 1700−1716.

(5) Cai, B.; Huang, L.; Xie, M. Bayesian Networks in Fault Diagnosis. *IEEE Trans. Ind. Inf.* **2017**, *13*, 2227−2240.

(6) Zhao, W.; Shi, T.; Wang, L. Fault Diagnosis and Prognosis of Bearing Based on Hidden Markov Model with Multi-Features. *Appl. Math. Nonlinear Sci.* **2020**, *5*, 71−84.

(7) Kouadri, A.; Hajji, M.; Harkat, M.-F.; Abodayeh, K.; Mansouri, M.; Nounou, H.; Nounou, M. Hidden Markov model based principal component analysis for intelligent fault diagnosis of wind energy converter systems. *Renew. Energy* **2020**, *150*, 598−606.

(8) Widodo, A.; Yang, B.-S. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560−2574.

(9) Sun, W.; Chen, J.; Li, J. Decision tree and PCA-based fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2007**, *21*, 1300−1317.

(10) Hu, Q.; Si, X.-S.; Zhang, Q.-H.; Qin, A.-S. A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests. *Mech. Syst. Signal Process.* **2020**, *139*, 106609.

(11) Zhu, W.; Sun, W.; Romagnoli, J. Adaptive k-Nearest-Neighbor Method for Process Monitoring. *Ind. Eng. Chem. Res.* **2018**, *57*, 2574−2586.

(12) Alauddin, M.; Khan, F.; Imtiaz, S.; Ahmed, S. Bibliometric Review and Analysis of Data-Driven Fault Detection and Diagnosis Methods for Process Systems. *Ind. Eng. Chem. Process Des. Dev.* **2018**, *57*, 10719.

(13) Comon, P. Independent component analysis, a new concept? *Signal Process.* **1994**, *36*, 287−314.

(14) Kresta, J. V.; Macgregor, J. F.; Marlin, T. E. Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Eng.* **1991**, *69*, 35−47.

(15) Guo, Y.; Chen, H. Fault diagnosis of VRF air-conditioning system based on improved Gaussian mixture model with PCA approach. *Int. J. Refrig.* **2020**, *118*, 1−11.

(16) Zhang, Z.; Zhao, J. A deep belief network based fault diagnosis model for complex chemical processes. *Comput. Chem. Eng.* **2017**, *107*, 395.

(17) Wu, H.; Zhao, J. Deep convolutional neural network model based chemical process fault diagnosis. *Comput. Chem. Eng.* **2018**, *115*, 185−197.

(18) Zhang, S.; Bi, K.; Qiu, T. Bidirectional Recurrent Neural Network-Based Chemical Process Fault Diagnosis. *Ind. Eng. Chem. Res.* **2020**, *59*, 824−834.

(19) Deng, L.; Zhang, Y.; Dai, Y.; Ji, X.; Zhou, L.; Dang, Y. Integrating feature optimization using a dynamic convolutional neural network for chemical process supervised fault classification. *Process Saf. Environ. Prot.* **2021**, *155*, 473−485.

(20) Chen, S.; Luo, L.; Xia, Q.; Wang, L. Self-attention Mechanism based Dynamic Fault Diagnosis and Classification for Chemical Processes. *J. Phys.: Conf. Ser.* **2021**, *1914*, 012046.

(21) 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*, 513−553.

(22) Taqvi, S. A. A.; Zabiri, H.; Tufa, L. D.; Uddin, F.; Fatima, S. A.; Maulud, A. S. A review on data-driven learning approaches for fault detection and diagnosis in chemical processes. *ChemBioEng Rev.* **2021**, *8*, 239−259.

(23) Zhang, Y.; Yu, J. Pruning graph convolutional network-based feature learning for fault diagnosis of industrial processes. *J. Process Control* **2022**, *113*, 101.

(24) Jia, M.; Hu, J.; Liu, Y.; Gao, Z.; Yao, Y. Topology-Guided Graph Learning for Process Fault Diagnosis. *Ind. Eng. Chem. Res.* **2023**, *62*, 3238−3248.

(25) Wu, D.; Bi, X.; Zhao, J. ProTopormer: Toward Understandable Fault Diagnosis Combining Process Topology for Chemical Processes. *Ind. Eng. Chem. Res.* **2023**, *62*, 8350−8361.

(26) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735−1780.

(27) Zhao, H.; Sun, S.; Jin, B. Sequential Fault Diagnosis Based on LSTM Neural Network. *IEEE Access* **2018**, *6*, 12929−12939.

(28) Zhang, Q.; Zhang, J.; Zou, J.; Fan, S. A Novel Fault Diagnosis Method based on Stacked LSTM. *IFAC-PapersOnLine* **2020**, *53*, 790−795.
(29) Zhang, S.; Qiu, T. Semi-supervised LSTM ladder autoencoder for chemical process fault diagnosis and localization. *Chem. Eng. Sci.* **2022**, *251*, 117467.
(30) Zhang, Q. S.; Zhu, S. C.; California, U. O. Visual interpretability for deep learning:a survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27−39.
(31) Peng, D.; El-Farra, N. H.; Geng, Z.; Zhu, Q. Distributed data-based fault identification and accommodation in networked process systems. *Chem. Eng. Sci.* **2015**, *136*, 88−105.
(32) Gangopadhyay, T.; Tan, S. Y.; Jiang, Z.; Meng, R.; Sarkar, S. *Spatiotemporal Attention for Multivariate Time Series Prediction and Interpretation*, 2020.
(33) Wu, D.; Zhao, J. Process Topology Convolutional Network Model for Chemical Process Fault Diagnosis. *Process Saf. Environ. Prot.* **2021**, *150*, 93.
(34) Liu, K.; Lu, N.; Wu, F.; Zhang, R.; Gao, F. Model Fusion and Multiscale Feature Learning for Fault Diagnosis of Industrial Processes. *IEEE Trans. Cybern.* **2023**, *53*, 6465−6478.
(35) Amar, M. N. Modeling solubility of sulfur in pure hydrogen sulfide and sour gas mixtures using rigorous machine learning methods. *Int. J. Hydrogen Energy* **2020**, *45*, 33274.
(36) Peppes, N.; Daskalakis, E.; Alexakis, T.; Adamopoulou, E.; Demestichas, K. Performance of Machine Learning-Based Multi-Model Voting Ensemble Methods for Network Threat Detection in Agriculture 4.0. *Sensors* **2021**, *21*, 7475.
(37) Saqlain, M.; Jargalsaikhan, B.; Lee, J. Y. A Voting Ensemble Classifier for Wafer Map Defect Patterns Identification in Semiconductor Manufacturing. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 171.
(38) Downs, J.; Vogel, E. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, *17*, 245−255 Industrial challenge problems in process control.
(39) Krstinić, D.; Braović, M.; Šerić, L.; Božić-Štulić, D. Multi-label classifier performance evaluation with confusion matrix. *Comput. Sci. Inf. Technol.* **2020**, *10*, 1.
(40) Schober, P.; Boer, C.; Schwarte, L. A. Correlation coefficients: appropriate use and interpretation. *Anesth. Analg.* **2018**, *126*, 1763−1768.
(41) Lou, C.; Li, X.; Atoui, M. A. Bayesian Network Based on an Adaptive Threshold Scheme for Fault Detection and Classification. *Ind. Eng. Chem. Res.* **2020**, *59*, 15155−15164.
(42) Wu, D.; Zhao, J. Process topology convolutional network model for chemical process fault diagnosis. *Process Saf. Environ. Prot.* **2021**, *150*, 93−109.
(43) Wei, Z.; Ji, X.; Zhou, L.; Dang, Y.; Dai, Y. A novel deep learning model based on target transformer for fault diagnosis of chemical process. *Process Saf. Environ. Prot.* **2022**, *167*, 480−492.
(44) Amin, M. T.; Khan, F.; Imtiaz, S.; Ahmed, S. Robust Process Monitoring Methodology for Detection and Diagnosis of Unobservable Faults. *Ind. Eng. Chem. Res.* **2019**, *58*, 19149−19165.

https://doi.org/10.1021/acsomega.3c09122
ACS Omega 2024, 9, 9486−9502