A New Mechanism Model-Assisted Spatiotemporal Information Fusion Quality-Related Fault Diagnosis Method for Large-Scale Industrial Processes

Dongjie Hua, Jie Dong, Kaixiang Peng, *Member, IEEE*, and Qichun Zhang, *Senior Member, IEEE*

***Abstract*— Modern industrial processes with long production lines and multiple operating units exhibit complex spatiotemporal characteristics under the constraints of information flow, material flow, and energy flow. The quality spatial heritability and dynamic-coupled variables of long-term processes bring serious challenges to conventional quality-related fault diagnosis. Against this background, a new mechanism model-assisted spatiotemporal information fusion quality-related fault diagnosis method is proposed. First, the coupling relationships among variables in the mechanism models are intuitively represented as causal correlation networks (CCNs), and large-scale industrial processes are decomposed into multiple subsystems interpretably by CCNs. Second, a novel 1-D convolutional autoencoder with neighborhood preserving embedding in Riemannian space (1DCAE-RNPE) is designed, where 1DCAE is adopted to capture temporal features stably and RNPE mines spatial features in higher dimensions. On this basis, the correlations among spatiotemporal features and quality variables are explored by distributed canonical correlation analysis (DCCA). Bayesian inference is enforced to establish global monitoring statistics. In addition, after a fault occurs, the contributions of the variables are obtained to identify the fault variables and combined with CCNs to recognize the root cause. Finally, the effectiveness and advantages of the proposed method are demonstrated through a practical large-scale industrial process, the hot strip mill process (HSMP). Superior fault diagnosis performance has been attained compared to other competing methods. Meanwhile, root cause recognition is accurately realized.**

***Index Terms*— Causal correlation networks (CCNs), distributed fault diagnosis, large-scale industrial process, quality-related, spatiotemporal characteristics.**

I. Introduction

With the expansion of industries, plant-wide processes are developing toward large-scale, multiunit cooperation and complex mechanisms, which put forward high requirements for safe production and product quality [1]. Thanks to the plant-wide deployment of advanced sensors, data-driven fault diagnosis methods are widely used in complex industries. However, in actual industrial processes, multisensor data exhibit complex spatiotemporal characteristics, that is, temporal characteristics influenced by process inertia and feedback, and so on, and spatial characteristics with different spatial distribution but highly coupled [2]. Accurately capturing the dynamics of the variables and analyzing the spatial correlation that exists among them bring a great challenge to fault diagnosis [3]. Therefore, the research on fault diagnosis methods that integrate spatiotemporal characteristics and quality information has profound significance.

Data-driven fault diagnosis methods can be divided into multivariate statistical process monitoring (MSPM) methods, machine learning, and deep neural networks (DNNs). In detail, traditional MSPM methods [4], [5], [6] exhibit limitations in addressing the nonlinear and dynamic complexity inherent in industrial data. To well accommodate the complexities of modern large-scale industrial processes, approaches such as dynamic modeling [7], slow feature analysis [8], and the application of kernel functions [9] have been sequentially introduced to tackle the challenges posed by the dynamics of industrial data, slow feature, nonlinearities, and so on. An increasing number of machine learning methods [10], [11], [12] are merged, leveraging the strengths of each to enhance online monitoring and fault diagnosis capabilities. However, the shallow network architecture inherent in these methods hinders the efficient extraction of intricate deep spatiotemporal features, falling short of adequately representing the complexity of industrial process data.

To further mine spatiotemporal features, the combination of different DNNs is extensively employed [13], [14]. Liu et al. [7] designed stacked autoencoders (SAEs) with different structures to extract spatiotemporal features and combined them with a manifold learning method to achieve quality prediction by taking into account the complex nonlinear features. Both Qiao et al. [15] and Xu et al. [16] proposed spatiotemporal feature extraction and fusion strategies for machine condition monitoring based on cascaded DNNs. Young et al. [17] proposed a deep learning architecture based on a recursive clustering algorithm to learn how to represent spatiotemporal patterns in data, where each node was trained independently in parallel. Ma et al. [18] proposed a bidirectional minimal gated unit structure for nonlinear dynamic soft

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant U21A20483, Grant 62273031, and Grant 62373040. *(Corresponding author: Jie Dong.)*

Dongjie Hua, Jie Dong, and Kaixiang Peng are with the Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China (e-mail: dongjie_hua@163.com; dongjie@ies.ustb.edu.cn; kaixiang@ustb.edu.cn).

Qichun Zhang is with the School of Creative and Digital Industries, Buckinghamshire New University, HP11 2JZ High Wycombe, U.K. (e-mail: kit.zhang@bucks.ac.uk).

sensor modeling, of which spatial features are represented by different subsystems. Although the above networks enhance modeling efficiency by accounting for spatiotemporal characteristics, the extraction of spatial features only stays in two dimensions, and recurrent neural network's (RNNs) variants have the risk of gradient explosion when processing industrial time-series big data. Meanwhile, the above methods ignore the interconnections and couplings between signals, resulting in modeling that fails to reflect the intricate dynamics of complex systems, which, in turn, can diminish detection accuracy.

Recently, interconnections and coupling between signals have been paid more and more attention in modeling. Zhao et al. [19] treated each univariate time series as a distinct feature and incorporated two parallel graph attention layers to capture the complex dependencies in multivariate time series across both temporal and feature dimensions. Deng and Hooi [20] combined a structure learning approach with graph neural networks and learned relationships between sensors as a graph and then identified and explained deviations from the learned patterns. The above methods only considered the correlation information between nodes. In practical large-scale industries, the relevant correlations between different subsystems jointly affect the quality of output. To further explore the coupling between subsystems, Chen and Jiang [21] employed stacked denoising autoencoders (AEs) to extract robust features for each block, capturing the variable correlations and constructed a DNN to model the correlations between a local unit and its neighboring units. Zhang et al. [22] deeply mined the hidden features in different subsystems by the SAE. Xiang et al. [23] proposed convolutional neural networks (CNNs) and bidirectional gated recurrent unit (BiGRU) cascade to extract spatiotemporal features, which improves the efficiency of feature extraction through a simplified model structure. He et al. [24] proposed a multiscale deep echo state network module and a multiscale residual network module to extract spatiotemporal features. Although the above methods account for the coupling of features between different subsystems, they did not involve the correlations among spatiotemporal features with quality variables. Moreover, the reliance on a priori knowledge for process decomposition does not guarantee the rationality of the outcomes, and as the complexity of the network increases, the interpretability decreases.

Strongly inspired by the above observations, a new mechanism model-assisted spatiotemporal information fusion quality-related fault diagnosis method is proposed, which deeply explores the correlation between spatiotemporal features and quality variables, and improves the interpretability of the overall fault diagnosis method through the mechanism models. We first transform the mechanism model into causal correlation networks (CCNs) and use CCNs to implement process decomposition. On this basis, a novel spatiotemporal analytic approach, 1-D convolutional autoencoder with neighborhood preserving embedding in Riemannian space (1DCAE-RNPE), is proposed to mine the spatial–temporal features. The maximum correlations among spatiotemporal features and quality variables are then mined through canonical correlation analysis (CCA). Finally, contribution plots are combined with CCNs to recognize the root cause. The main contributions of this article can be summarized as follows.

1) A new mechanism model-assisted spatiotemporal information fusion quality-related fault diagnosis method is proposed, where the mechanism model is used to model the deep correlations between subsystems, and the correlation of higher-order spatial–temporal features with quality variables is explored.
2) A novel spatiotemporal analytic approach is designed for the stable mining of temporal features and higher-dimensional spatial features, which can reveal the intrinsic correlation among spatiotemporal features and quality variables.
3) A method for constructing CCNs based on mechanism models is developed, which enhances the interpretability of fault diagnosis methods by integrating mechanism models into process decomposition and root cause recognition with the help of CCNs.

The organization of the remaining parts of this article is as follows. Section II gives the basic ideas of AEs, neighborhood preserving embedding (NPE), and CCA and the motivation of this work. Section III explains the fault diagnosis process of the proposed method in detail. The performance of the proposed method is demonstrated and discussed in a real hot strip mill process (HSMP) in Section IV. Finally, Section V summarizes the article.

II. Background and Motivation

In this section, the preliminary theoretical knowledge of AEs, NPE, and CCA is described and the motivation of this work is highlighted.

A. AE-Based Feature Extraction

AEs consist of an encoder and a decoder. The encoder is responsible for mapping the input data into a hidden space, which is usually realized using a nonlinear activation function. The high-dimensional variable matrix $\mathbf{X}$ of the input layer is transformed into a low-dimensional hidden variable $\mathbf{Z}$ using a nonlinear activation function

$$\mathbf{Z} = \sigma_1\left(\mathbf{W'}_1\mathbf{X} + b'_1\right) \tag{1}$$

where $\mathbf{W'}_1$ is the encoding weight matrix, b'_1 is the encoding bias vector, and σ_1 is the coded activation function.

The decoder reproduces the original data and maps this hidden representation back to the initial input space. Using a decoder to reconstruct the original variable $\mathbf{X}$ from the hidden layer variable $\mathbf{Z}$

$$\mathbf{X'} = \sigma_2\left(\mathbf{W'}_2\mathbf{Z} + b'_2\right) \tag{2}$$

where $\mathbf{W'}_2$ is the decoding weight matrix, b'_2 is the decoding bias vector, and σ_2 is the decoding activation function. The whole process can be viewed as a data compression and decompression process.

The goal of AEs is to make the reconstructed output data $\mathbf{X'}$ as close as possible to the original input data $\mathbf{X}$, which is usually achieved by minimizing the reconstruction error or loss function.

B. NPE-Based Feature Extraction

NPE is a dimensionality reduction algorithm based on local manifold structures [25]. Suppose a sample set $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_p]^{\mathrm{T}} \in \mathbb{R}^{p\times n}$ of high-dimensional data is collected, where p denotes the dataset dimension and n denotes the number of samples. The adjacency graph of each sample is first constructed by the k-nearest neighbor method [26]. Second, in this adjacency graph, the weight coefficient is denoted as ω_{ij} when there is a connection between samples $\mathbf{x}_i$ and $\mathbf{x}_j$. If there is no connection, the weight coefficient is zero.

In the original high-dimensional space $\mathbb{R}^p$, all samples can be reconstructed based on linear combinations of the nearest neighbor samples in their neighborhood. The goal of the NPE algorithm is to find an optimal weight matrix $\mathbf{W}$ that minimizes the error in the reconstruction. The problem is described as follows:

$$\Phi(\mathbf{W}) = \min \sum_{i=1}^{p} \left\| \mathbf{x}_i - \sum_{j=1}^{p} \omega_{ij}\mathbf{x}_j \right\|^2, \quad \text{s.t.} \sum_{j=1}^{n} \omega_{ij} = 1. \tag{3}$$

Then, it is assumed that all samples in the low-dimensional space can be reconstructed by linear combinations of the nearest neighbor samples. Therefore, the weights of the reconstructed data can be used to map the high-dimensional data to the low-dimensional data space [27]. The solution of the projection matrix $\mathbf{A}$ is achieved by minimizing the reconstruction error of (4) so that each sample can be reconstructed as accurately as possible by samples in its neighborhood, thus preserving the local structure of the original data in the low-dimensional space

$$\begin{aligned}\mathbf{A} &= \arg\min_{\mathbf{h}} \sum_{i=1}^{p} \left\| \mathbf{h}_i - \sum_{j=1}^{p} \omega_{ij}\mathbf{h}_j \right\|^2 \\ &= \arg\min_{\mathbf{h}} \mathbf{h}^{\mathrm{T}}(\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})\mathbf{h} \\ &= \arg\min_{\mathbf{h}} \mathbf{a}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}(\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})\mathbf{X}\mathbf{a}. \end{aligned} \tag{4}$$

Supposing $\mathbf{h}^{\mathrm{T}}\mathbf{h} = \mathbf{a}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{a} = 1$, the above optimization problem can be transformed into

$$\mathbf{X}^{\mathrm{T}}\mathbf{M}\mathbf{X}\mathbf{a} = \lambda \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{a} \tag{5}$$

where $\mathbf{M} = (\mathbf{I}-\mathbf{W})^{\mathrm{T}}(\mathbf{I}-\mathbf{W})$. The low-dimensional features can be obtained by solving (5) for the generalized eigenvalues $\mathbf{A}$. The eigenvectors $\mathbf{a}$ corresponding to the smallest f eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_f$ are solved to form the optimal projection matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_f]$. Then, projection matrix $\mathbf{A}$ is used to project the high-dimensional dataset into the low-dimensional space to obtain the low-dimensional dataset $\mathbf{H} = [\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_f^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{f\times n}$

$$\mathbf{H} = \mathbf{X}\mathbf{A}. \tag{6}$$

C. Canonical Correlation Analysis

CCA is a statistical method for studying the correlation of two sets of variables. To study the correlation between the two sets of variables $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)^{\mathrm{T}}$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_q)^{\mathrm{T}}$, the correlation matrices $\mathbf{J}$ and $\mathbf{L}$ are computed separately for the random variables.

First, the two sets of variables $\mathbf{U} = \mathbf{J}^{\mathrm{T}}\mathbf{X} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_p\mathbf{x}_p$ and $\mathbf{V} = \mathbf{L}^{\mathrm{T}}\mathbf{Y} = \beta_1\mathbf{y}_1 + \beta_2\mathbf{y}_2 + \cdots + \beta_q\mathbf{y}_q$ are defined as a linear combination of the two original sets of variables. Then, the Pearson correlation coefficient [28] can be used to represent the optimization problem as follows:

$$\max_{\mathbf{J},\mathbf{L}} \rho(\mathbf{U},\mathbf{V}) = \frac{\mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xy}}\mathbf{L}}{\sqrt{(\mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}\mathbf{J})(\mathbf{L}^{\mathrm{T}}\mathbf{S}_{\mathbf{yy}}\mathbf{L})}} \tag{7}$$

where $\mathbf{S}_{\mathbf{xx}} = \mathrm{cov}(\mathbf{X},\mathbf{X})$, $\mathbf{S}_{\mathbf{yy}} = \mathrm{cov}(\mathbf{Y},\mathbf{Y})$, and $\mathbf{S}_{\mathbf{xy}} = \mathbf{S}_{\mathbf{yx}}^{\mathrm{T}} = \mathrm{cov}(\mathbf{X},\mathbf{Y})$. Suppose that $\mathbf{J} = \mathbf{S}_{\mathbf{xx}}^{-(1/2)}\mathbf{U}$ and $\mathbf{L} = \mathbf{S}_{\mathbf{yy}}^{-(1/2)}\mathbf{V}$. The key problem is to compute $\mathbf{J}$ and $\mathbf{L}$. It is assumed that $\mathbf{U}$ and $\mathbf{V}$ have a mean of 0 and variances of $\mathrm{var}(\mathbf{U})$=$\mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}\mathbf{J}$ and $\mathrm{var}(\mathbf{V}) = \mathbf{L}^{\mathrm{T}}\mathbf{S}_{\mathbf{yy}}\mathbf{L}$ both are 1. The covariance between typical variables can be expressed as $\mathrm{cov}(\mathbf{U},\mathbf{V}) = \mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xy}}\mathbf{L}$. Then, subject to the constraints on $\mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}\mathbf{J} = \mathbf{L}^{\mathrm{T}}\mathbf{S}_{\mathbf{yy}}\mathbf{L} = 1$, the problem is further transformed into

$$\max_{\mathbf{J},\mathbf{L}} \rho(\mathbf{U},\mathbf{V}) = \mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xy}}\mathbf{L}. \tag{8}$$

Singular value decomposition (SVD) is applied to solve the eigenvalue and eigenvector optimization functions of the correlation matrix. The constraint of the optimization function is

$$\begin{cases} \mathbf{J}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}\mathbf{J} = \mathbf{U}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}^{-\frac{1}{2}}\mathbf{S}_{\mathbf{xx}}\mathbf{S}_{\mathbf{xx}}^{-\frac{1}{2}}\mathbf{U} = 1 \\ \mathbf{L}^{\mathrm{T}}\mathbf{S}_{\mathbf{yy}}\mathbf{L} = \mathbf{V}^{\mathrm{T}}\mathbf{S}_{\mathbf{yy}}^{-\frac{1}{2}}\mathbf{S}_{\mathbf{yy}}\mathbf{S}_{\mathbf{yy}}^{-\frac{1}{2}}\mathbf{V} = 1. \end{cases} \tag{9}$$

It can be obtained that $\mathbf{U}^{\mathrm{T}}\mathbf{U} = 1$, $\mathbf{V}^{\mathrm{T}}\mathbf{V} = 1$, and the optimization objective under this constraint is transformed into $\max_{\mathbf{u},\mathbf{v}} \mathbf{U}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}^{-(1/2)}\mathbf{S}_{\mathbf{xy}}\mathbf{S}_{\mathbf{yy}}^{-(1/2)}\mathbf{V}$.

Define the relevance assessment matrix γ

$$\gamma = \mathbf{S}_{\mathbf{xx}}^{-\frac{1}{2}}\mathbf{S}_{\mathbf{xy}}\mathbf{S}_{\mathbf{yy}}^{-\frac{1}{2}}. \tag{10}$$

For γ, SVD is performed according to the following equation:

$$\gamma = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Delta}^{\mathrm{T}} \tag{11}$$

where

$$\mathbf{\Gamma} = (\gamma_1, \ldots, \gamma_s), \quad \mathbf{\Delta} = (\delta_1, \ldots, \delta_m), \quad \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_\kappa & 0 \\ 0 & 0 \end{pmatrix}$$

$$\mathbf{\Lambda}_\kappa = \mathrm{diag}\{\lambda_1, \ldots, \lambda_\kappa\}, \quad \mathrm{rank}(\gamma) = \kappa.$$

It is obviously that $\max_{\mathbf{u},\mathbf{v}} \mathbf{U}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}^{-(1/2)}\mathbf{S}_{\mathbf{xy}}\mathbf{S}_{\mathbf{yy}}^{-(1/2)}\mathbf{V} = \max_{\mathbf{u},\mathbf{v}} (\mathbf{U}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Delta}^{\mathrm{T}}\mathbf{V})$. Since $\mathbf{\Gamma}$ and $\mathbf{\Delta}$ are orthogonal arrays, the maximum value of $\mathbf{U}^{\mathrm{T}}\mathbf{S}_{\mathbf{xx}}^{-(1/2)}\mathbf{S}_{\mathbf{xy}}\mathbf{S}_{\mathbf{yy}}^{-(1/2)}\mathbf{V}$ is the maximum value of the singular value. The coefficients can be derived as follows:

$$\mathbf{J} = \mathbf{S}_{\mathbf{xx}}^{-\frac{1}{2}}\mathbf{\Gamma}(:, 1:\kappa) \tag{12}$$

$$\mathbf{L} = \mathbf{S}_{\mathbf{yy}}^{-\frac{1}{2}}\mathbf{\Delta}(:, 1:\kappa). \tag{13}$$

Process detection can be realized by constructing residual vectors using the feature matrix, and then constructing statistics and setting the corresponding control limits.

D. Motivation of This Work

In real large-scale industrial cases, data-driven fault diagnosis-based methods are widely used. However, the data from multiple operating units of long-term industrial processes exhibit complex spatiotemporal characteristics under the influence of material, information, and energy flows, and the spatiotemporal characteristic is closely related to product quality. Meanwhile, when constructing data-driven models, the interpretability is limited [29]. Based on the above analysis, the motivations of this article can be concluded as follows.

1) Increased model accuracy and interpretability. Because of the inherent black-box characteristic of data-driven models constructed based on deep networks, the results generated are often difficult to understand intuitively in practical applications in industrial sites, which may pose potential security risks.
2) Deep extraction of spatiotemporal features to mine the association with product quality. Existing temporal feature extraction methods based on the RNN and its variants have the risk of gradient explosion, and spatial feature extraction methods based on convolution have limited extraction dimensions.
3) Improvement of the rationality and interpretability of root cause recognition. Existing fault root cause recognition methods tend to focus on the predictive function of variables while failing to profoundly reflect strict causation when discerning causal relationships between variables, resulting in the results obtained being deficient in terms of interpretability.

Inspired by the above issues, this article aims to propose a mechanism model-assisted quality-related fault diagnosis method considering spatiotemporal features and quality correlation, to ensure that highly interpretable detection and root cause recognition results are obtained. The proposed method is presented in Section III.

III. Methodology

In this section, the fault diagnosis method as shown in Fig. 1 is described in detail. First, the screening of key core variables based on mechanism models and a preliminary decomposition of large-scale industrial processes is performed based on the locations of the process variables in the offline modeling process. Then, further process decomposition is performed based on CCNs. Second, the spatiotemporal features of each unit are extracted. Through distributed CCA (DCCA), a distributed quality-related monitoring model is established. In the online process, the online data are input into the trained model to get the monitoring results, and then based on the contribution plots and CCNs, the root causes of faults are analyzed.

A. Process Decomposition Based on CCNs

Since modern large-scale industrial processes are generally characterized by long production lines at the plant level and sensors distributed throughout the plant collect data along the production lines, it is difficult to construct a global monitoring model to meet the actual production requirements.

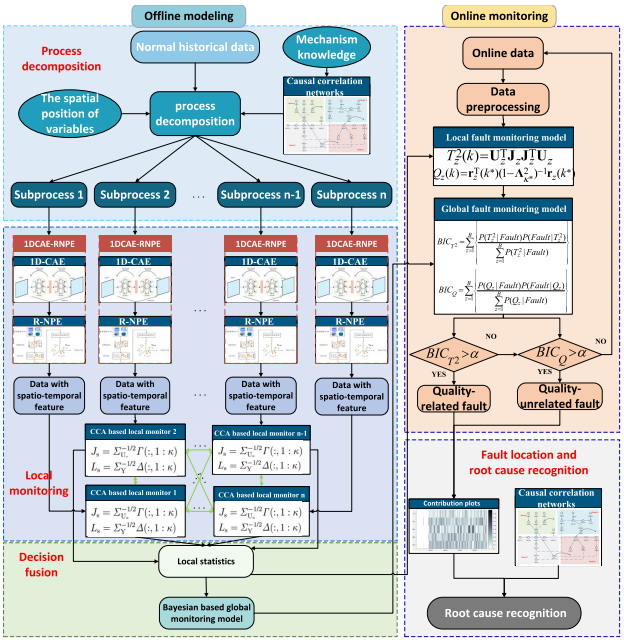

Fig. 1. Diagram of the mechanism model-assisted spatiotemporal information fusion quality-related fault diagnosis method.

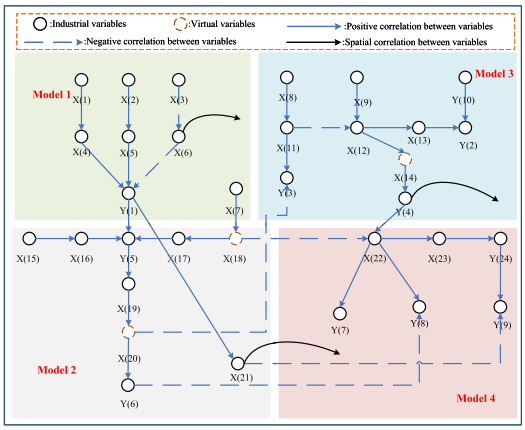

Fig. 2. Schematic of CCNs for industrial processes.

Decomposing the huge system into multiple subsystems for distributed monitoring interpretatively is necessary.

Key process and quality variables are first filtered out based on mechanism models. Then, the large-scale industrial process is decomposed into submodels and the variables are categorized into different submodels. Variables are connected according to the mechanism models, and coupling between models can be represented by connections between cross-model variables. To better illustrate the causal relationship between variables, some process variables are added in conjunction with the mechanism models, which are corrected to maintain a strong causal relationship connecting the variables to obtain the CCNs of the industrial process. Fig. 2 shows a schematic of an industrial process CCNs.

It is worth stating that virtual variables indicated by orange dashed circles are used as a medium to express potential relationships between variables, further refining the structure of the CCNs.

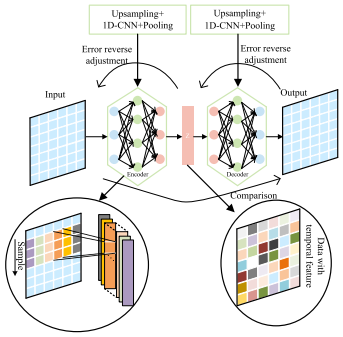

Fig. 3. 1DCAE network structure diagram.

The industrial process decomposition is divided into two steps. First, since the spatial characteristics of the process variables cannot be ignored, the variables are initially divided according to their spatial location. Second, the decomposition results are tuned according to the CCNs, and the process decomposition is realized by grouping the variables with strong causal correlation into the same subsystem and deleting the variables with no direct causal relationship but only in the same spatial location. The CCNs constructed from the mechanism models can visually observe the correlations between different spatial location variables, thus assisting the process decomposition and making the results more interpretable.

According to the above scheme, the industrial process is decomposed into B subsystems. For the bth subsystem, its internal variable set is $\mathbf{x}_b$ and the corresponding sample matrix, the input of the bth submodel, is $\mathbf{X}_b \in \mathbb{R}^{p_b \times n}$, $b = 1, 2, \ldots, B$, where p_b is the number of variables in the bth subsystem and n is the number of samples.

B. Spatiotemporal Feature Extraction

Large amounts of industrial process data are sampled sequentially from different locations in the production line. As a result, spatiotemporal characteristic is one of the inherent properties of industrial data. One-dimensional CNN performs well in temporal feature extraction, but many of the data features lack ground truth, which leads to poor interpretability of the network. In contrast, the AE is an unsupervised learning method [30]. Therefore, a novel 1DCAE network is proposed which can unsupervised extract temporal features of data. Each subsystem process data from the process decomposition in Section III-A are input into the temporal feature extraction network 1DCAE separately. The schematic of the 1DCAE network designed in this article is shown in Fig. 3.

1DCAE network starts with the encoder part. After process decomposition, the dimension of some subsystem process data may be too low for further learning of the deep network. Therefore, the dimension needs to be increased by an upsampling layer. Next, the data are convolved by 1-D CNN. The convolution part consists of two 1-D CNN layers and two pooling layers. Setting the sampling of the convolutional kernels along the direction of the time axis ensures that the 1-D CNN efficiently extracts temporal features from the data. Assuming that the internal temporal data of the bth subsystem is $\mathbf{X}_b = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{p_b}]$, which is input to the bth encoder for encoding, then the implicit layer $\mathbf{Z}_b$ is input to the decoder, and the output is obtained by decoding through the symmetric network structure. The loss function [31] for unsupervised training 1DCAE is defined as follows:

$$\text{Loss}_b = \frac{1}{p_b}\sum_{i=1}^{p_b} \left\| \mathbf{x}'_i - \mathbf{x}_i \right\|^2. \tag{14}$$

By minimizing the loss function between the input and the output, the 1DCAE network can efficiently and unsupervised learn the temporal features compared to the CNN method. It is worth noting that during the model validation phase, the use of the Monte Carlo Dropout method to measure the model's uncertainty has enhanced its reliability [32], [33]. Then, we can obtain the temporal feature $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_B]$.

The metric and geodesics based on Riemannian manifolds can essentially reflect the geometric positional relationship and similarity of each sample, so we propose the R-NPE algorithm for extracting spatial features in temporal features. It is worth noting that the spatiotemporal features extract key information from two perspectives of the process data, respectively. The serial structure does not affect the separate feature extraction.

First, the delayed reconstruction method [34] is used to project the original 1-D time-domain data into a high-dimensional space and construct a symmetric positive definite (SPD) matrix to describe the data. Given the temporal features of $\mathbf{Z}_b = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{p_b}]^{\mathrm{T}}$, the reconstructed variables of the time series $\mathbf{P} = [\mathbf{P}_1, \ldots, \mathbf{P}_{p_b}]$ are obtained using the delayed reconstruction method as

$$\mathbf{P}_b = \begin{bmatrix} z_b^1(t), z_b^{1+\tau}(t), \ldots, z_b^{1+(m-1)\tau}(t) \\ z_b^2(t), z_b^{2+\tau}(t), \ldots, z_b^{2+(m-1)\tau}(t) \\ \ldots, \ldots, \ldots, \ldots, \ldots \\ z_b^{N_z}(t), z_b^{N_z+\tau}(t), \ldots, z_b^{N_z+(m-1)\tau}(t) \end{bmatrix}, \quad b = 1, \ldots, p_b \tag{15}$$

where t denotes the observation time, N_z is the number of phase spaces, m denotes the dimension of the embedding space, which can be determined by Cao method [27], and τ denotes the delay time, which can be determined by mutual information method [27]. In addition, the SPD matrix of the bth subsystem is obtained by the following equation:

$$\mathbf{S}_{\text{SPD},b} = \mathbf{P}_b \mathbf{P}_b^{\mathrm{T}}. \tag{16}$$

To consider the internal structural characteristics of the manifold in more detail, logarithmic Euclidean measurements are used. First, through logarithmic operations

$$\mathbf{T}_b = \log(\mathbf{S}_{\text{SPD},b}), \quad b = 1, 2, \ldots, p_b \tag{17}$$

mapping the SPD matrix to the tangent space. Then, the SPD vectorization operation is performed

$$\mathbf{V}_b = \text{vec}(\mathbf{T}_b) \tag{18}$$

where vec denotes vectorization and $\mathbf{V}_b = [\mathbf{v}_1, \ldots, \mathbf{v}_{p_b}]$. This metric maintains various properties of the input data. Finally, logarithmic Euclidean measurements $\mathbf{D_b}$ is calculated

$$\mathbf{D}_b = \begin{bmatrix} d(\mathbf{S}_{\text{SPD},1}, \mathbf{S}_{\text{SPD},1}) & \cdots & d(\mathbf{S}_{\text{SPD},1}, \mathbf{S}_{\text{SPD},n}) \\ \vdots & \ddots & \vdots \\ d(\mathbf{S}_{\text{SPD},n}, \mathbf{S}_{\text{SPD},1}) & \cdots & d(\mathbf{S}_{\text{SPD},n}, \mathbf{S}_{\text{SPD},n}) \end{bmatrix} \tag{19}$$

where $d(\mathbf{S}_{\text{SPD},i}, \mathbf{S}_{\text{SPD},j}) = \|\mathbf{v}_i - \mathbf{v}_j\|_F^2$.

Algorithm 1 R-NPE Algorithm

Input: Temporal feature $\mathbf{Z}$
Output: Low-dimensional embedding matrix $\mathbf{H}$

1: **for all** temporal feature $\mathbf{Z}_b$ of each subsystem b **do**
2: Calculate the delayed reconfiguration timing matrix $\mathbf{P}_b$ by Eq. (15)
3: Calculate the SPD matrix $\mathbf{S}_{SPD,b}$ by Eq. (16)
4: The SPD matrix is logarithmically operated by Eq. (17), Eq. (18) and vectorized to obtain $\mathbf{V_b}$
5: Calculate logarithmic Euclidean measurements by Eq. (19)
6: Calculate the projection matrix $\mathbf{A}_b$ by Eq. (4)
7: The low-dimensional embedding matrix $\mathbf{H}_b$ is obtained through Eq. (6)
8: **end for**
9: Splicing the low-dimensional embedding matrices of the subsystems yields $\mathbf{H}$
10: **return H**

After obtaining the geodesic distances between points on the manifold, the spatial features can be extracted using the NPE algorithm introduced in Section II-B. The specific process of the R-NPE algorithm is shown as Algorithm 1.

C. Quality-Related Fault Detection

The DCCA algorithm is adopted to maximize the correlation among spatiotemporal features and quality variables and to model the relationship between them. Given that the quality variables are $\mathbf{Y}^{\mathrm{T}} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_q)$ and spatiotemporal features of the bth subsystem $\mathbf{H}_b$. According to Section II-C, transforming local spatiotemporal features and quality variables into $\mathbf{U}_b = {\mathbf{J}_b}^{\mathrm{T}}\mathbf{H}_b$ and $\mathbf{V} = \mathbf{L}^{\mathrm{T}}\mathbf{Y}$. Then, the correlation matrices $\mathbf{J}_b$ and $\mathbf{L}$ that describe the correlation between $\mathbf{H}_b$ and $\mathbf{Y}$ can be obtained. Reconstruct the stable relationship between input and output

$$\mathbf{\Phi}_v(k^*) = \mathbf{\Phi}_u(k^*) + \varepsilon(k^*) \tag{20}$$

where $\mathbf{\Phi}_u = \mathbf{J}_b^{\mathrm{T}}\mathbf{U}_b$ and $\mathbf{\Phi}_v = \mathbf{L}^{\mathrm{T}}\mathbf{V}$ denote new typical correlation variables, $\varepsilon(k^*)$ denotes $\mathbf{\Phi}_v(k^*)$ correlated noise term. Define the residual vector as follows:

$$\mathbf{r}_b(k^*) = \mathbf{\Phi}_v(k^*) - \mathbf{\Lambda}_{\kappa^*}\mathbf{\Phi}_u(k^*). \tag{21}$$

Constructing two statistics T_b^2 and Q_b for process monitoring. T_b^2 statistic is used to detect the portion of the fault that occurs in the input subspace and is correlated with the output

$$T_b^2(k) = \mathbf{U}_b^{\mathrm{T}}\mathbf{J}_b\mathbf{J}_b^{\mathrm{T}}\mathbf{U}_b. \tag{22}$$

Let $T_{b,\mathrm{lim}}^2$ denote the control limit of T_b^2 statistic which can be calculated by kernel density estimation (KDE) [9].

Q_b statistic is used to detect the portion of the fault that occurs in the output subspace and is related to the input

$$Q_b(k) = \mathbf{r}_b^{\mathrm{T}}(k^*)\left(1 - \mathbf{\Lambda}_{\kappa^*}^2\right)^{-1}\mathbf{r}_b(k^*). \tag{23}$$

In the same way, the control limit of Q_b statistic $Q_{b,\mathrm{lim}}$ can also be calculated by KDE [9].

The Bayesian inference is adopted to form the final decision [28]. Plant-wide process monitoring statistics are obtained as

$$\mathrm{BIC}_{T^2} = \sum_{b=1}^{B}\left\{\frac{P\left(T_b^2|\mathrm{Fault}\right)P\left(\mathrm{Fault}|T_b^2\right)}{\sum_{b=1}^{B}P\left(T_b^2|Fault\right)}\right\} \tag{24}$$

$$\mathrm{BIC}_{Q} = \sum_{b=1}^{B}\left\{\frac{P(Q_b|\mathrm{Fault})P(\mathrm{Fault}|Q_b)}{\sum_{b=1}^{B}P(Q_b|\mathrm{Fault})}\right\} \tag{25}$$

where $P(\mathrm{Fault}|T_b^2)$ and $P(\mathrm{Fault}|Q_b)$ are the fault probabilities in each subspace, and $P(T_b^2|\mathrm{Fault})$ and $P(Q_b|\mathrm{Fault})$ denote the conditional probability of a fault occurring.

Based on BIC_{T^2}, BIC_Q, and the confidence level α, the monitoring logic of the whole process is

$$\begin{cases}\mathrm{BIC}_{T^2} > \alpha, & \mathrm{quality-related\ fault}\\ \mathrm{BIC}_{T^2} \le \alpha \text{ and } \mathrm{BIC}_Q > \alpha, & \mathrm{quality-unrelated\ fault}\\ \mathrm{others}, & \mathrm{fault\ free}.\end{cases} \tag{26}$$

D. Root Cause Recognition Based on CCNs

For the local–global fault diagnosis strategy of large-scale complex processes, root cause recognition is divided into two steps. First, when a fault occurs that affects the global characteristics of multiple subsystems or processes, T_b^2 or Q_b statistic exceeds the control limits. The results of the subsystem local detection can be used to find the subsystems where the fault occurs. Second, we select attention weights as a measure of the variable's contribution to compensate for the limitation of the traditional contribution plot method that cannot calculate data contribution after extracting deep features.

In Section III-B, the bth subsystem process data $\mathbf{X}_b$ are first upsampled by (1). Since the outermost layer of the autoencoder retains the fullest information in the process data, which is more conducive to the identification of fault variables, $\mathbf{W} = (w_1, w_2, \ldots, w_{p_b})$ in (1) is the attention weight of different variables in the bth subsystem. Taking into account the reliability of identification and the differences in contribution between samples, the effect of too much sample variation on contribution is first eliminated. Calculate the difference between the rth online sample $x_{b,i}^{r,\mathrm{new}}$ of the ith process variable of the bth subsystem and the mean of the normal sample

$$\vartheta_{b,i}^{r} = \frac{\left|x_{b,i}^{r,\mathrm{new}} - \bar{x}_{b,i}\right|}{S_{b,i}} \tag{27}$$

where $S_{b,i}$ denotes standard deviation. Then, add attention weight

$$\mathrm{diff}_{b,i}^{r} = w_i\vartheta_{b,i}^{r}. \tag{28}$$

Define sample weight contribution SWC to explain the contribution of the sample

$$\mathrm{SWC}\left(x_{b,i}^{r,\mathrm{new}}\right) = \zeta_{b,i}^{r} = \begin{cases}1, & \mathrm{diff}_{b,i}^{r} \ge 1\\ \mathrm{diff}_{b,i}^{r}, & \mathrm{diff}_{b,i}^{r} < 1\end{cases} \tag{29}$$

where $\zeta_{b,i}^{r}$ denotes SWC of $x_{b,i}^{r,\mathrm{new}}$. A further weighted value $\omega_g = 1$ is introduced, rounding off the contribution of the

statistic that does not exceed the control limit. Correct the variable contributions to

$$\mathrm{WC} = \omega_g \times \mathrm{SWC} \tag{30}$$

where SWC is the contribution obtained from the calculation of (29). When the statistic exceeds the control limit, $\omega_g = 1$; otherwise, $\omega_g = 0$. In this way, samples of variables that are malfunctioning are calculated and located.

After the fault is located in the variables, it is also necessary to analyze the interrelationships between the variables to analyze the root cause of the fault. The connecting lines in CCNs are obtained from the mechanism models, which express the causal relationship between the variables. After filtering out the occurring fault variables, analyzing the characteristics such as the trend of these variables at the time of the fault, and querying the relationship among these variables and other variables in the CCNs, the root cause variables of the fault can be located. It is worth noting that the CCNs constructed through the mechanism models are completed in the offline stage, and in the online stage, only the fault variables need to be obtained to achieve highly interpretable fault root cause identification in conjunction with the CCN analysis.

IV. Case Study

To verify the effectiveness of the proposed method in this article, HSMP is selected for the case study. HSMP is a key process in the steel production process, which consists of several key steps as follows. First, the steel loading machine loads the continuous casting billet into the step furnace for heating. After the heated slabs are descaled by high-pressure water, they are sent to the roughing mill (the thickness of the slabs is about 200 mm) to be rolled for 3–5 passes, and then they are sent to the finishing mill through the processes of heat preservation, flying shear, and high-pressure water descaling. The finishing mill process (FMP) generally consists of six or seven rolling mills, with a looper between the stands, and a hydraulic automatic Gauge control (AGC) system is used to ensure the strip thickness accuracy and strip shape requirement. After FMP, accelerated cooling is controlled by a laminar cooling system, and the final coil is wound into steel coils.

FMP directly affects the product quality performances of the entire FMP product thickness, strip shape, and so on. Once a fault occurs in the process, it is difficult to ensure the product quality only by a delayed feedback control strategy. Therefore, we carry out a process monitoring study on the finishing stage. It focuses on two quality indicators exit thickness and strip shape. A large amount of data reflecting the production process is collected at the 1700-mm HSMP line of Anshan Iron and Steel Group Company Ltd., Liaoning, China. The CCNs are shown in Fig. 4.

According to the mechanism models, the FMP is divided into thickness control model, strip shape control model, temperature control model, and looper control model. More than 200 variables associated with the models are selected in the CCNs to be connected together. The coupling between the models and the correlation between the stands can be clearly seen. The most critical variables in the FMP are selected

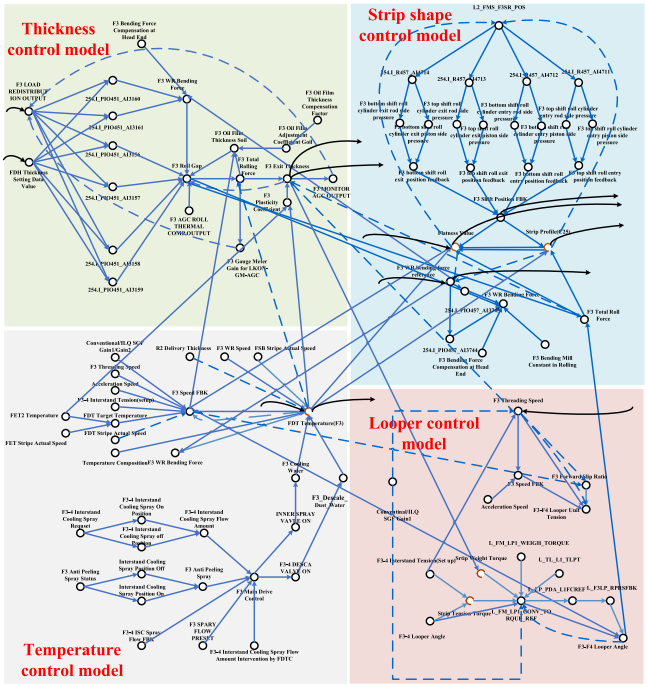

Fig. 4. Local CCNs diagram of the FMP (the third stand).

to construct the dataset. 4000 sets of normal historical data are trained offline to establish a quality monitoring model, and 4000 sets of real-time data with different faults are utilized, respectively, for fault diagnosis. The details of the relevant variables are shown in Table I, and the other variables involved in Fig. 4 will be used as an aid in guiding the process decomposition and root cause recognition.

The seven stands can be roughly categorized into three sections—1 and 2 stands, 3–5 stands, and 6 and 7 stands—based on the spatial locations where the process variables are located. According to the observation of the FMP CCNs, the increase in the bending force will make the rolling gap decrease, and the decrease in the rolling gap will increase the rolling force. Rolling force will finally directly affect the exit thickness of the strip. From the interstand correlation analysis in the CCNs, it can be seen that the bending force of the former stand will have an effect on the bending force of the latter stand, and the exit thickness of the former stand will have an effect on the AGC of the next stand, which, in turn, indirectly affects the roll gap and rolling force of the next stand. Therefore, the process is decomposed and the results are shown in Table II.

Three typical faults are selected for simulation verification, and the information about the faults is shown in Table III.

Ablation and comparative experiments are conducted to verify the effectiveness of the fault diagnosis method. The 1DCAE-RNPE-DCCA algorithm which is a conventional process decomposition method, the CCN-DCCA algorithm without extracting spatiotemporal features, the CNN–long short-term memory (CNN-LSTM) [35] algorithm which is a deep learning algorithm and takes into account spatiotemporal characteristic of the data, parallel spatiotemporal information analysis (PSTIA) [36] which is a parallel spatiotemporal

TABLE I

SELECTION OF MONITORING PROCESS VARIABLES IN THE FMP

Variable No.	Category	Description	Units
1,7,15,23,31,39,47	Process variable	Rolling force of stand i, $i = 1, ..., 7$	KN
2,8,16,24,32,40,48	Process variable	Differential rolling force of stand i, i=1,...,7	KN
3,9,17,25,33,41,49	Process variable	Rolling power of stand i, $i = 1, ..., 7$	KW
4,10,18,26,34,42	Process variable	Inter-stand tension of stand i and $(i + 1), i = 1, ..., 6$	Mpa
11,19,27,35,43,50	Process variable	Balance force of stand i, i=2,...,7	KN
12,20,28,36,44,51	Process variable	Bending force of stand i, i=2,...,7	KN
5,13,21,29,37,45,52	Process variable	Average roll gap of stand i, i=1,...,7	mm
6,14,22,30,38,46,53	Process variable	Roller speed of stand i, i=1,...,7	m/s
54	Quality variable	Crown	μm
55	Quality variable	Finishing exit thickness	mm

TABLE II

SUBSYSTEM DIVISION RESULTS IN THE FMP

No.	Variables
Subsystem 1	1,2,3,4,5,6,7,8,9,10,11,12,13, 14,15,16,17,18,19,20,21,22
Subsystem 2	7,8,9,10,11,12,13,14,15,16,17,18,19,20, 21,22,23,24,25,26,27,28,29,30,31,32,33,34, 35,36,37,38,39,40,41,42,43,44,45,46
Subsystem 3	31,32,33,34,35,36,37,38,39,40,41,42,43, 44,45,46,47,48,49,50,51,52,53

TABLE III

FAULT DESCRIPTION OF THE FMP

Fault No.	Fault description	Type	Durations
Fault 1	Hardness fault of incoming material	Quality-related	1501-4000
Fault 2	Bending force measurement sensor fault of stand 5	Quality-related	501-3000
Fault 3	Condensate valve blocked fault between stands 3 and 4	Quality-unrelated	501-3000

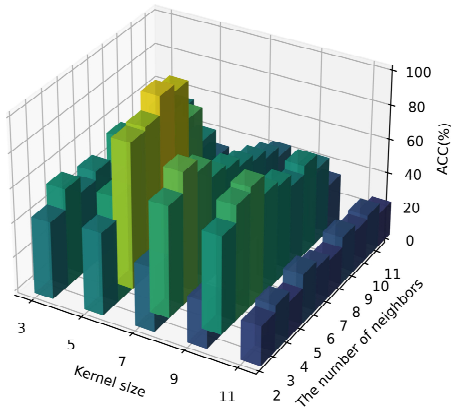

Fig. 5. Parameter selection results.

feature extraction method, and distributed conditional variational recurrent autoencoder (D-CVRAE) [37] which transfers the model in spatiotemporal dimensions, are chosen to compare with the proposed method. Based on the amount of data and [35], the optimal parameters are found in the corresponding range. Using ACCuracy (ACC) [38] as a measure, the results of the experiment are shown in Fig. 5. It can be seen that the highest fault diagnosis accuracy is achieved when the kernel size is 5 and the number of neighbors is 5. Table IV presents the parameters for the proposed method.

TABLE IV

PARAMETERS OF THE PROPOSED METHOD

1DCAE-RNPE-DCCA	
Kernel size	5
Number of convolution kernels	128, 256
Batch size	64
Epoch	120
Hidden size	30
Optimizer	Adam
Activation function	ReLU
The number of neighbors	5

We illustrate the effectiveness of our method using Faults 1 and 3 as examples. Fig. 6 illustrates the results for Fault 1. It can be seen from Fig. 6(a) that both T^2 and Q can detect the fault, and the fault is judged to be a quality-related fault. In addition, the proposed method detects the fault immediately at the 1501st samples with a false alarm rate (FAR) [9] of 0. From Fig. 6, it can be seen that the proposed method is significantly superior to the other methods in terms of FDR [9] and FAR.

From Fig. 7, it can be seen that all three subsystems simultaneously monitored the fault at the 1501st sample for Fault 1. It can be concluded that this type of fault affects all subsystems at the same time.

Fig. 8 illustrates the results for quality-unrelated Fault 3. It can be seen from Fig. 8(a) that quality-related statistics T^2 did not detect any faults, while the quality-unrelated statistics Q detected the faults at the 501–3000 sample points. Fig. 8(b) and (c) are not as good as the proposed method due to irrational process decomposition and lack of consideration of spatiotemporal characteristics, respectively. It can be seen from Fig. 8(d) that the method based on CNN–LSTM did not detect Fault 3. The method predicts quality variables by extracting spatiotemporal features from the process variables and compares the predicted values with the actual values to achieve online fault detection, therefore, cannot detect quality-unrelated faults. The detection results only indicate that no quality-related faults occur and do not indicate the existence of quality-unrelated faults. Therefore, this method has limited assistance in actual on-site production and poses challenges for root cause recognition. The detection results of PSTIA and D-CVRAE methods in Fig. 8(e) and (f) are slightly inferior to the proposed method due to that they do not consider deep spatial characteristics. And they obtain poorly interpretable detection results.

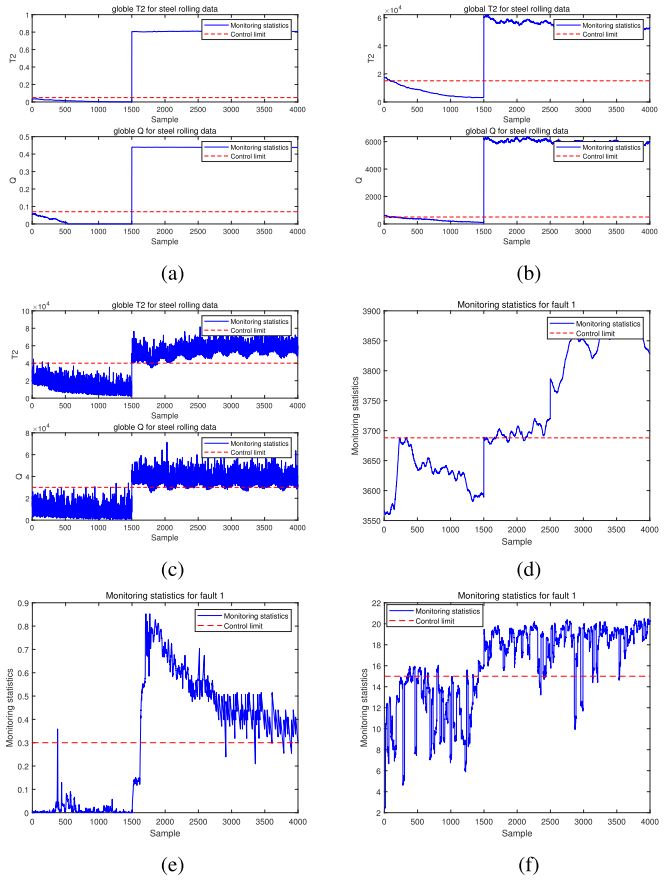

Fig. 6. Global monitoring results of Fault 1. (a) Proposed method. (b) 1DCAE-RNPE-DCCA. (c) CCN-DCCA. (d) CNN–LSTM. (e) PSTIA. (f) D-CVRAE.

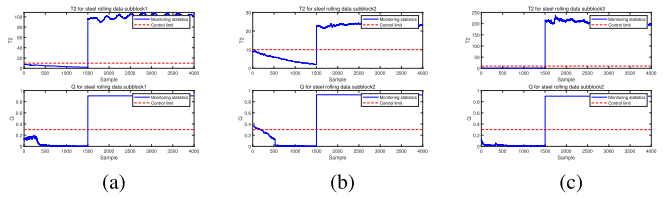

Fig. 7. Local monitoring results of the proposed method Fault 1. (a) Subsystem 1. (b) Subsystem 2. (c) Subsystem 3.

Fig. 9 shows the local monitoring results for Fault 3 of the proposed method where subsystem 3 did not detect faults.

Table V quantifies the results of the different methods by FDRs. The proposed method can distinguish whether the faults are quality-related faults or not. At the same time, the FDRs outperform the other comparative methods across the board.

Table VI demonstrates the superiority of the proposed method in this article in terms of FARs.

It can be seen from Tables V and VI that RNPE-1DCAE-DCCA uses a conventional data-driven process decomposition-based method, which is not highly interpretable and does not mine the correlation between subsystems thoroughly. CCN-DCCA ignores the correlation among spatiotemporal features and quality variables, and thus the detection results are not as good as the proposed method. CNN–LSTM takes into account the correlation among spatiotemporal features and

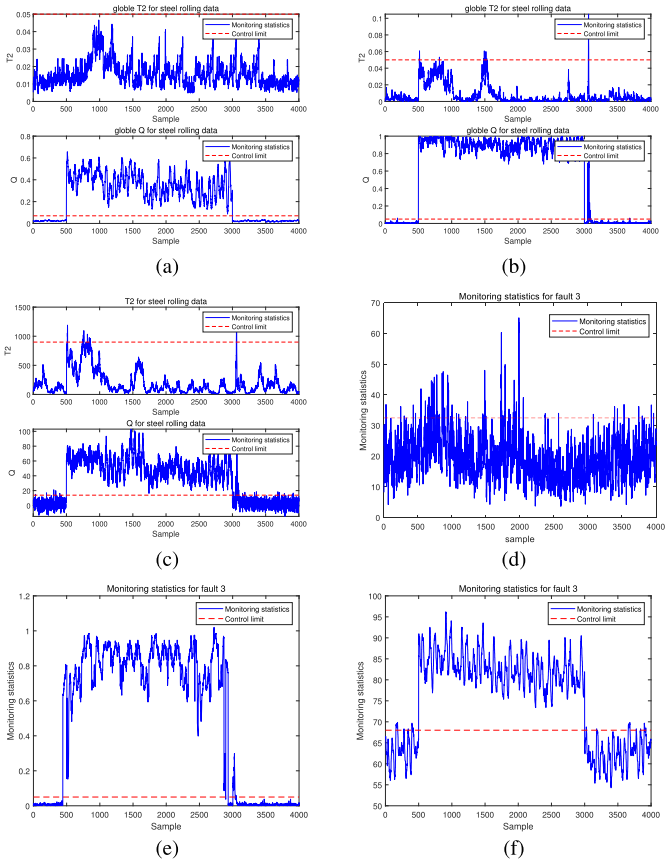

Fig. 8. Global monitoring results of Fault 3. (a) Proposed method. (b) 1DCAE-RNPE-DCCA. (c) CCN-DCCA. (d) CNN–LSTM. (e) PSTIA. (f) D-CVRAE.

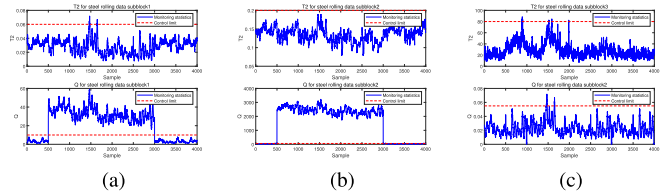

Fig. 9. Local monitoring results of the proposed method Fault 3. (a) Subsystem 1. (b) Subsystem 2. (c) Subsystem 3.

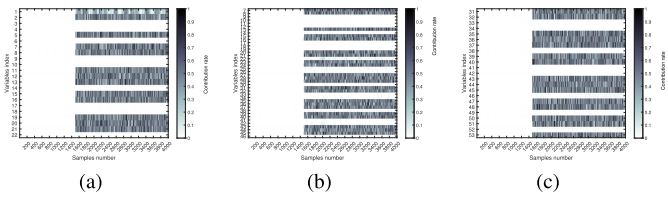

Fig. 10. Contribution plots for Fault 1. (a) Subsystem 1. (b) Subsystem 2. (c) Subsystem 3.

quality variables, but the method is not sufficient for mining spatiotemporal features and cannot detect quality-unrelated faults compared to the proposed method.

After obtaining the accurate detection results, the root cause recognition is further achieved with the help of CCNs. Contribution plots for Fault 1 are given in Fig. 10.

As can be seen in Fig. 10, the root cause variables for Fault 1 are the total rolling forces in stands 1–7 as well as the

TABLE V

FDRs OF COMPARATIVE METHODS (%)

Fault No.	RNPE-1DCAE -DCCA		CCNs -DCCA		CNN -LSTM	PSTIA	D-CVRAE	CCNs-RNPE -1DCAE-DCCA	
	T^2	Q	T^2	Q				T^2	Q
Fault 1	**100**	**100**	97.8	93.3	89.8	95.8	88.3	**100**	**100**
Fault 2	97.3	99	93.2	97.8	88.3	92.6	89.9	**97.7**	**100**
Fault 3	1.2	98.5	0.8	96.2	90.3	97.8	98.3	**0**	**98.8**

TABLE VI

FARs OF COMPARATIVE METHODS (%)

Fault No.	RNPE-1DCAE -DCCA		CCNs -DCCA		CNN -LSTM	PSTIA	D-CVRAE	CCNs-RNPE -1DCAE-DCCA	
	T^2	Q	T^2	Q				T^2	Q
Fault 1	6.1	8.4	0.3	0.3	0.3	0.6	5.9	**0**	**0**
Fault 2	7.4	4.5	0.2	0.7	1.2	3.7	4.5	**0**	**0**
Fault 3	**0**	5.2	**0**	**0**	2.4	2.8	8.9	**0**	**0**

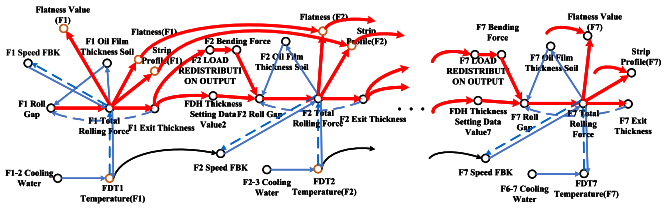

Fig. 11. Local CCNs for Fault 1.

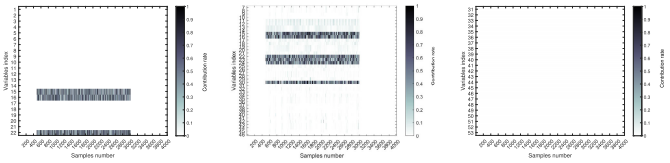

Fig. 12. Contribution plots for Fault 3. (a) Subsystem 1. (b) Subsystem 2. (c) Subsystem 3.

bending forces and the balance forces in stands 2–7. The fault variables start generating anomalies from the 1501st sample and continue until the 4000th sample and it can be seen that the hardness fault of incoming material affects all stands and is, therefore, detected in all subsystems. To better illustrate the process of root cause recognition, Fig. 11 gives the local CCNs for Fault 1.

As can be seen in Fig. 11, the incoming hardness abnormally first affects the roll gap of each stand, and the roll gap directly affects the total rolling force of each stand. At the same time, the changes in the roll gap lead to changes in the bending force. And the bending force and rolling force will have an impact on the strip shape. The change in rolling force will have an impact on the exit thickness of each stand. The sensor for detecting strip shape is only arranged at the exit position of the seventh stand. Therefore, the influence of each stand on the strip shape will end up in the crown and flatness at the exit of the seventh stand. While the thickness gauges are distributed in each stand. As a result, the exit thickness of each stand has a fault and is passed on to the next stand's load redistribution output, thickness setting data, and roller speed. In turn, these variables affect the bending force, roll gap, and total rolling force, which affect the exit thickness of the next stand. The above analysis is consistent with the actual fault cause analysis on-site.

For Fault 3, contribution plots are given in Fig. 12.

When Fault 3 occurs, the variables that cause the fault are the total rolling force, rolling force difference, and rolling speed of the 3 and 4 stands.

Combined with the local CCNs for Fault 3 shown in Fig. 13, it can be seen that Fault 3 is quality-unrelated.

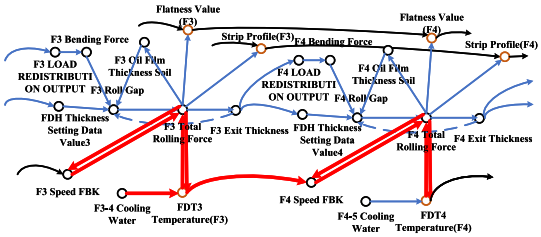

Fig. 13. Local CCNs for Fault 3.

Fault 3 is caused by a condensate valve blocked between stands 3 and 4. As can be seen in Fig. 13, the fault will first affect the output temperature of the third stand, and the temperature will have an impact on the total rolling force of the current stand. However, due to the feedback regulation, the fault does not affect exit thickness or strip shape. In addition, the output temperature change of the current stand will affect the next stand roller speed, and the same feedback controller exists between the roller speed and total rolling force. Therefore, Fault 3 will make the local stand temperature, roller speed, and total rolling force change, but it will not affect the thickness and strip quality variables used in this article. The above analysis is consistent with the actual fault cause analysis on-site, thus verifying the effectiveness of the root cause analysis method achieved by combining the relationship diagram with CCNs in this article, and the CCNs method increases the interpretability of fault diagnosis.

V. CONCLUSION

In this article, a new mechanism model-assisted spatiotemporal information fusion quality-related fault diagnosis method

is designed for large-scale industrial processes. On the basis of this method, CCNs are constructed by mechanism models first, which are enforced to decompose industrial processes and recognize root causes interpretably. Meanwhile, a 1DCAE-RNPE method is proposed to mine spatiotemporal features deeply. The correlations among spatiotemporal features and quality variables were considered and local statistics were constructed. Bayesian inference is enforced to establish global monitoring statistics. By introducing mechanism models to construct a data-driven fault diagnosis model, more interpretable process decomposition and fault root cause recognition results are obtained. Furthermore, the association between spatiotemporal features of large-scale processes and product quality is mined, which breaks through the barrier of interaction between spatiotemporal features and quality information in distributed monitoring systems. The feasibility of the proposed method is experimentally verified by the HSMP, and superior fault diagnosis performance has been attained compared to other competing methods.

Although the proposed method considers the correlation of spatiotemporal features with quality variables in conjunction with the mechanism models and achieves highly interpretable fault root cause recognition, it is only applicable to fault diagnosis of large-scale industrial processes with complete mechanism models. Moreover, the method did not take into account the uncertainty of the process. Future work will focus on building more lightweight CCNs, incorporating mechanism models into feature extraction for deep networks, designing new algorithms to automate the implementation of CCNs-based root cause recognition, as well as considering the effect of uncertainty when constructing models for industrial fault diagnosis.

REFERENCES

[1] K. Peng, L. Ma, and K. Zhang, "Review of quality-related fault detection and diagnosis techniques for complex industrial processes," *Acta Autom. Sinica*, vol. 43, no. 3, pp. 349–365, 2017.

[2] Q. Jiang, X. Yan, and B. Huang, "Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes," *Ind. Eng. Chem. Res.*, vol. 58, no. 29, pp. 12899–12912, Jul. 2019.

[3] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[4] B. Chen, C. Tao, J. Tao, Y. Jiang, and P. Li, "Bearing fault diagnosis using ACWGAN-GP enhanced by principal component analysis," *Sustainability*, vol. 15, no. 10, p. 7836, May 2023.

[5] Z. Lou, Y. Wang, Y. Si, and S. Lu, "A novel multivariate statistical process monitoring algorithm: Orthonormal subspace analysis," *Automatica*, vol. 138, Apr. 2022, Art. no. 110148.

[6] Q. Zhu, Q. Liu, and S. J. Qin, "Concurrent quality and process monitoring with canonical correlation analysis," *J. Process Control*, vol. 60, pp. 95–103, Dec. 2017.

[7] C. Liu, K. Wang, Y. Wang, S. Xie, and C. Yang, "Deep nonlinear dynamic feature extraction for quality prediction based on spatiotemporal neighborhood preserving SAE," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[8] D. Li, J. Dong, and K. Peng, "A novel adaptive STFT-SFA based fault detection method for nonstationary processes," *IEEE Sensors J.*, vol. 23, no. 10, pp. 10748–10757, May 2023.

[9] L. Ma, M. Wang, J. Dong, and K. Peng, "A novel distributed detection framework for quality-related faults in industrial plant-wide processes," *Neurocomputing*, vol. 492, pp. 126–136, Jul. 2022.

[10] C. Shang, F. Yang, B. Huang, and D. Huang, "Recursive slow feature analysis for adaptive monitoring of industrial processes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 11, pp. 8895–8905, Nov. 2018.

[11] P. Tang, K. Peng, and J. Dong, "Nonlinear quality-related fault detection using combined deep variational information bottleneck and variational autoencoder," *ISA Trans.*, vol. 114, pp. 444–454, Aug. 2021.

[12] X. Chen and C. Zhao, "Condition-driven soft transition modeling and monitoring strategy for complex nonstationary process," *IFAC-PapersOnLine*, vol. 54, no. 3, pp. 445–450, 2021.

[13] M. A. Rahim, M. M. Rahman, M. S. Islam, A. J. M. Muzahid, M. A. Rahman, and D. Ramasamy, "Deep learning-based vehicular engine health monitoring system utilising a hybrid convolutional neural network/bidirectional gated recurrent unit," *Expert Syst. Appl.*, vol. 257, Dec. 2024, Art. no. 125080.

[14] Y. Ren, R. Zhang, and F. Gao, "A network structure for industrial process fault diagnosis based on hyper feature extraction and stacked LSTM," *Chem. Eng. Sci.*, vol. 287, Apr. 2024, Art. no. 119745.

[15] H. Qiao, T. Wang, P. Wang, S. Qiao, and L. Zhang, "A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series," *Sensors*, vol. 18, no. 9, p. 2932, Sep. 2018.

[16] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, Dec. 2020, Art. no. 108086.

[17] S. R. Young, A. Davis, A. Mishtal, and I. Arel, "Hierarchical spatiotemporal feature extraction using recurrent online clustering," *Pattern Recognit. Lett.*, vol. 37, pp. 115–123, Feb. 2014.

[18] L. Ma, M. Wang, and K. Peng, "Bidirectional minimal gated unit-based nonlinear dynamic soft sensor modeling framework for quality prediction in process industries," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023.

[19] H. Zhao et al., "Multivariate time-series anomaly detection via graph attention network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 841–850.

[20] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, May 2021, pp. 4027–4035.

[21] S. Chen and Q. Jiang, "Distributed robust process monitoring based on optimized denoising autoencoder with reinforcement learning," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[22] X. Zhang, L. Ma, K. Peng, and C. Zhang, "A novel quality-related distributed fault diagnosis framework for large-scale sequential manufacturing processes," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 4397–4407, Mar. 2024.

[23] L. Xiang, X. Yang, A. Hu, H. Su, and P. Wang, "Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks," *Appl. Energy*, vol. 305, Jan. 2022, Art. no. 117925.

[24] Q. He, Y. Pang, G. Jiang, and P. Xie, "A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6875–6884, Oct. 2021.

[25] X. Sha, C. Luo, and N. Diao, "A robust fault detection method based on neighborhood preserving embedding," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[26] N. Zhang, Y. Tian, X. Wang, Y. Xu, Q. Zhu, and Y. He, "Novel bootstrap-based discriminant NPE integrated with orthogonal LPP for fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023.

[27] A. Miao, Z. Ge, Z. Song, and F. Shen, "Nonlocal structure constrained neighborhood preserving embedding model and its application for fault detection," *Chemometric Intell. Lab. Syst.*, vol. 142, pp. 184–196, Mar. 2015.

[28] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.

[29] S. X. Ding, *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Cham, Switzerland: Springer, 2008.

[30] M. Shi et al., "Deep hypergraph autoencoder embedding: An efficient intelligent approach for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 260, Jan. 2023, Art. no. 110172.

[31] G. Taguchi, *Introduction to Quality Engineering: Designing Quality Into Products and Processes*. Tokyo, Japan: Asian Productivity Organization, 1986.

[32] S. Shirmohammadi and H. Al Osman, "Machine learning in measurement part 1: Error contribution and terminology confusion," *IEEE Instrum. Meas. Mag.*, vol. 24, no. 2, pp. 84–92, Apr. 2021.

[33] H. Al Osman and S. Shirmohammadi, "Machine learning in measurement part 2: Uncertainty quantification," *IEEE Instrum. Meas. Mag.*, vol. 24, no. 3, pp. 23–27, May 2021.

[34] C. Li and K. Liu, "Path signature-based phase space reconstruction for stock trend prediction," *Int. J. Data Sci. Analytics*, vol. 14, no. 3, pp. 293–304, Sep. 2022.

[35] Z. Zhi, L. Liu, D. Liu, and C. Hu, "Fault detection of the harmonic reducer based on CNN-LSTM with a novel denoising algorithm," *IEEE Sensors J.*, vol. 22, no. 3, pp. 2572–2581, Feb. 2022.

[36] C. Zhang, J. Dong, K. Peng, and H. Zhang, "Quality-related spatio-temporal information analytics-based multiunit synergetic monitoring for plant-wide industrial processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 7268–7280, 2025.

[37] P. Tang, K. Peng, Z. Chen, and J. Dong, "A novel distributed CVRAE-based spatio-temporal process monitoring method with its application," *IEEE Trans. Ind. Informat.*, vol. 19, no. 11, pp. 10987–10997, Nov. 2023.

[38] Q. Li, Y. Wang, J. Dong, C. Zhang, and K. Peng, "Multi-node knowledge graph assisted distributed fault detection for large-scale industrial processes based on graph attention network and bidirectional LSTMs," *Neural Netw.*, vol. 173, May 2024, Art. no. 106210.