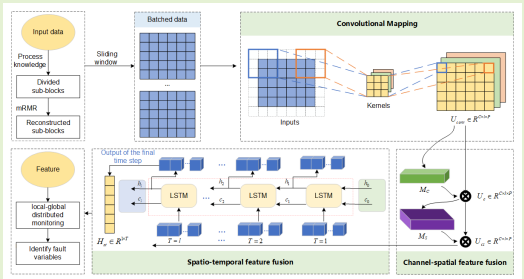
Spatio-temporal deep feature fusion network-based distributed quality-related anomaly monitoring for large-scale manufacturing processes

Kaixiang Peng, *Member, IEEE*, Qingchen Fan, Tie Li, Xin Qin and Jie Dong

***Abstract*— Large-scale manufacturing processes are characterized by complex process, dynamic coupling between variables, and quality heritability between sub-processes, which poses great challenges to traditional process monitoring. In this paper, a novel distributed quality anomaly monitoring framework based on a designed spatio-temporal deep feature fusion network is proposed. First, sub-block division is performed based on the process knowledge, and minimal redundancy maximal relevance is introduced to select quality-related coupling variables between sub-blocks to achieve reconstruction and precise partitioning. Second, a new spatio-temporal deep feature fusion network is designed, combining the advantages of convolutional neural network and long short-term memory network in extracting spatial and temporal features, respectively. Canonical correlation analysis is employed to quantify the correlation between the extracted features and quality variables, constraining the extraction of deep spatio-temporal quality-related features. A two-stage training strategy is proposed to ensure the completeness of quality-related information. The convolutional block attention module is introduced to dynamically focus on important quality-related information while accelerating network convergence. Then, based on the deep spatio-temporal features, we design the local monitoring statistic. Bayesian inference is used to obtain the global quality anomaly monitoring results, and combined with variable forgetting contributions to achieve visualization of fault propagation in large-scale sequential manufacturing processes. Finally, the effectiveness of the proposed method is validated by the actual hot strip mill process (HSMP) data.**



***Index Terms*— Process monitoring, Distributed quality anomaly detection, Spatio-temporal deep feature fusion, Large-scale manufacturing process.**

I. Introduction

With the development of science and technology, typical manufacturing processes, such as metallurgy, chemical engineering, and building materials, are becoming increasingly complex and large-scale. These process exhibit complex characteristics, such as strong coupling between operating units, highly correlated process variables and inherent quality information [1], [2]. Due to the continuity of the production process, the output of the previous sub-process is often the input for the next one. These sub-processes are serially connected, presenting obvious spatial distribution characteristic, and spreading and connecting through information and quality flow. In addition, when one of the process variables is abnormal, due to the time lag of the variables between the sub-processes, it does not immediately affect the final product quality, but gradually leads to abnormalities in the subsequent process variables and quality variables, presenting process and quality heritability, and the variables in the sub-processes are dynamically coupled. When a fault occurs in one of the processes, with the transmission of fault information, it will spread, accumulate, and amplify among the sub-processes, eventually leading to abnormalities in the entire process, affecting the final product quality, and even serious safety accidents [3]. In this case, there is an urgent need for accurate quality anomaly diagnostics to determine the operating status of the equipment and process quality in real-time, to ensure safe and stable production while providing field engineers with key information for subsequent adjustment and maintenance.

For quality-related fault diagnosis, data-driven approaches are considered the mainstream approaches, which do not require precise mathematical models and expert experience

This paper was supported by the National Natural Science Foundation of China (NSFC) under Grants (U21A20483, 62373040, 62273031).(Corresponding author: Kaixiang Peng.)

Kaixiang Peng, Qingchen Fan, Tie Li, Xin Qin and Jie Dong 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: kaixiang@ustb.edu.cn; m202220700@xs.ustb.edu.cn; d202440126@xs.ustb.edu.cn; d20231-0360@xs.ustb.edu.cn; dongjie@ies.ustb.edu.cn).

but extract meaningful information from process data [4], [5]. Traditionally, canonical correlation analysis (CCA), principal component regression (PCR), and partial least squares (PLS) have been the most popular quality-related fault diagnosis approaches [6]–[8]. Related scholars have proposed various extension methods to cope with the nonlinear and dynamic characteristics of the process. Manifold learning methods [9], Kernel-based [10], and neural network-based [11] nonlinear processing methods have been derived, and dynamic internal methods [12], time-series-based correlation analysis methods [13], [14], and state-space modeling methods have been proposed to solve the dynamic problems in the processes. However, these methods only use a shallow network structure and focus on shallow learning, which have limitations in learning ability and extracting features, and insufficient ability to characterize data in large-scale complex processes.

To well extract deep features, deep learning methods have received more and more attention and applications in the field of fault diagnosis [15]. Multiple layers of stacked nonlinear networks are capable in theory of approximating any complex function, effectively capturing the nonlinear relationships among variables and thereby rendering the extracted features much robust and reliable. For example, the stacked autoencoder (SAE), a classical unsupervised neural network, has excellent performance and good scalability, which can learn data features well, and achieve the prediction of quality variables [16], [17]. By adjusting the correlation weights of the process and quality variables, as well as increasing the reconstruction of the quality variables, it is possible to extract quality-related features from the SAE extension methods [16], [18], [19]. Deep canonical correlation analysis (DCCA), as a nonlinear extension of CCA, has also received extensive attention from scholars in various fields [20], which enhances the correlation between features extracted from multiple stacked nonlinear layers, by adding a CCA layer to constrain the deep representation. The scholars combined SAE and DCCA to construct a data-driven individual association learning framework that describe the relationships between process input and output variables to enable process monitoring [21]. Additionally, multi-scale fusion network and spatio-temporal feature extraction network have also been developed and applied to fault diagnosis [22]–[24]. The above works focus on building a centralized monitoring model, ignoring the local information and correlation among processes, which ultimately affects the monitoring and diagnostic results.

The large-scale and multi-process of the manufacturing process pose great challenges to quality anomaly diagnosis, specifically when sub-processes are coupled and process variables interact with each other. To address this challenge, scholars have proposed decentralized and multi-block diagnostic strategies [25], [26]. It consists of three main steps. Firstly, the appropriate process decomposition, informed by empirical knowledge and data analysis, aids in extracting local information and facilitates a much rational approach to local diagnostic analysis. Then, local monitoring models are established based on process data characteristics to reduce the computational burden of the models. Finally, reliable monitoring statistics are designed to fuse the information from each sub-block and provide the final monitoring decision, generally using the Bayesian fusion method [26]. The above methodologies can well describe the local characteristics of large-scale manufacturing processes, simplify model complexity, and achieve reliable monitoring results. Chen et al. [27] proposed a distributed CCA-based approach, which uses correlation information from the neighboring nodes to reduce uncertainties and decreases the cost of the data transmission between network nodes. However, as a multivariate statistical approach, complex deep features within and between sub-processes are often ignored. Combining deep learning methods and distributed frameworks to extract the deep features and mine the important information is helpful in improving the monitoring performance. Yao and Ge [28] proposed the distributed and parallel extreme learning machine (dp-ELM) and hierarchical extreme learning machine (dp-HELM) for multimode process quality prediction with big data. The efficient ELM algorithm is transformed into the distributed and parallel modeling form according to the map-reduce framework. Rong et al. [29] designed a large-scale supervisory process monitoring based on quality indicators using the idea of community segmentation of complex networks. Jiang et al. [30] extended SAE and DCCA to whole process monitoring for local-global modeling and distributed computing, and used mutual information and CCA as the information correlation constraints between sub-blocks when building the monitoring model. These methods consider the interactions of sub-process and use deep learning methods to mine the features. However, the consideration of spatio-temporal information is insufficient. More importantly, there is little attention to the transfer of quality heritability information between relevant processes during production, where the quality information of the upstream is transferred to the downstream along with the manufacturing process.

Based on the spatio-temporal distribution, dynamic coupling between variables, and quality heritability characteristics of large-scale sequential manufacturing processes, a novel Spatio-Temporal Deep Feature Fusion Network-based (STDFFN) distributed quality-related anomaly monitoring framework is proposed in this paper. The proposed STDFFN model fully considers the spatio-temporal information in process variables by CNN and LSTM networks. During the training process, CCA is used to quantify the correlation between deep features and quality variables, constraining the model's extraction of deep spatio-temporal quality-related features. An attention mechanism is employed to dynamically focus on the most important quality-related information and accelerate network convergence. The distributed quality anomaly monitoring framework and variable oblivion contribution makes it easier to analyze and observe the quality anomaly monitoring results and fault propagation in large-scale sequential manufacturing processes. The key contributions are outlined as follows:

1) A distributed quality anomaly monitoring framework based on STDFFN is proposed to address spatio-temporal distribution, variable coupling, and quality transmission in large-scale sequential manufacturing processes.

2) A new spatio-temporal deep feature fusion network is designed, which can extract deep spatio-temporal quality-related features by quantifying the correlation between deep

features and quality variables.

3) Based on Bayesian inference and variable forgetting contribution, the global quality anomaly monitoring results and fault propagation visualization of large-scale sequential manufacturing processes have been achieved.

The proposed distributed quality anomaly monitoring framework has been validated in actual HSMP, verifying its effectiveness and feasibility.

The rest of the paper is organized as follows. Section II briefly reviews the basic ideas of CCA and DCCA algorithms. The framework for detecting quality anomalies in large-scale sequential manufacturing processes is detailed in Section III, and its effectiveness is verified by the case study of an actual HSMP in Section IV. Section V provides the conclusion of the paper.

II. Preliminaries

In this section, we provide a concise review of the foundational theories employed in the model and the motivation of the paper.

A. Canonical Correlation Analysis

CCA [31], [32] has been implemented in process monitoring for extracting correlated representations, which explores and characterizes the linear correlation between variables. Suppose $u \in \Re^p$ and $y \in \Re^q$ are two sets of random variables with covariance matrices Σ_u, Σ_y, Σ_{uy}. CCA aims to determine two linear mappings, P_u and P_y, to maximize the correlation between $P_u^{\mathrm{T}}u$ and $P_y^{\mathrm{T}}y$. The corresponding optimization problem is

$$\begin{aligned}(P_u^*, P_y^*) &= \arg \max_{(P_u, P_y)} \operatorname{corr}\left(P_u^{\mathrm{T}}u, P_y^{\mathrm{T}}y\right) \\ &= \arg \max_{(P_u, P_y)} \frac{P_u^{\mathrm{T}}\Sigma_{uy}P_y}{\left(P_u^{\mathrm{T}}\Sigma_u P_u\right)^{1/2}\left(P_y^{\mathrm{T}}\Sigma_y P_y\right)^{1/2}}\end{aligned} \tag{1}$$

To solve Eq. (1), construct the matrix Σ and perform the singular value decomposition as

$$\Sigma = \Sigma_u^{-1/2}\Sigma_{uy}\Sigma_y^{-1/2} = U\Lambda V^{\mathrm{T}} \tag{2}$$

where $\Lambda \in \Re^{p\times q}$ is the diagonal matrix, leading to the vectors P_u and P_y as

$$P_u = \Sigma_u^{-1/2}U_r \in \Re^{p\times p} \tag{3}$$

$$P_y = \Sigma_y^{-1/2}V_r \in \Re^{q\times q} \tag{4}$$

where r denotes the number of non-zero singular values, $r = rank(\Lambda) \le \min\{p, q\}$, and U_r and V_r are the first r columns. CCA can effectively captures the linear correlations between variables but fails to adequately characterize the nonlinear relationships among them.

B. Deep Canonical Correlation Analysis

DCCA, as a nonlinear extension of CCA, has garnered widespread interest among researchers across various disciplines [20]. DCCA employs two fully connected neural networks to process two inputs, u and y, resulting in their nonlinear representations H_1 and H_2 (as shown in Fig.1). The training objective is to jointly optimize the parameters of both neural networks to maximize the correlation between the output features. That is

$$\begin{aligned}(\theta_1^*, \theta_2^*) &= \arg \max_{(\theta_1, \theta_2)} corr\left(f_1\left(u;\theta_1\right), f_2\left(y;\theta_2\right)\right) \\ &= \arg \max_{(\theta_1, \theta_2)} corr(H_1, H_2)\end{aligned} \tag{5}$$

where, θ_1 and θ_2 represent the parameters of the two networks, including weights and biases. The sum of the first k singular values of the matrix $\Sigma = \Sigma_{H_1}^{-1/2}\Sigma_{H_1H_2}\Sigma_{H_2}^{-1/2}$ represents the total correlation of the first k components of H_1 and H_2. When considering all components, it corresponds to the trace of Σ, i.e.,

$$corr\left(H_1, H_2\right) = ||\Sigma||_{tr} = tr(\Sigma^{\mathrm{T}}\Sigma)^{1/2} \tag{6}$$

$corr\left(H_1, H_2\right)$ serves as the objective function for training the networks. The parameters (θ_1^*, θ_2^*) are optimized using gradient descent method. When $corr\left(H_1, H_2\right)$ reaches its maximum value, it means that the total correlation of the network's output features is maximized.

C. Motivation and Problem Formulation

Facing complex industrial processes, Jiang et al. [32] extended CCA-based fault diagnosis to the distributed monitoring of whole process industries. They considered the interrelated information between adjacent units to establish local monitors and employed genetic algorithm (GA) to identify the relevant variables. This approach reduces communication costs and makes the distributed monitoring scheme more effective than centralized monitoring in detecting faults across the whole process. However, the linear CCA-based process monitoring method does not yield satisfactory results when dealing with nonlinear processes. Jiang and Yan [33] proposed a detection method based on deep correlation representation learning, where deep representations generated by a deep confidence neural network were used for CCA to explore the correlation between two sets of representations. DBN takes into

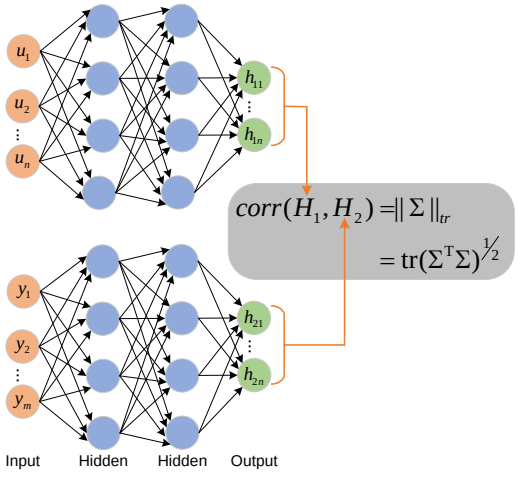


Fig. 1. Schematic diagram of DCCA, consisting of two fully connected neural networks where the network outputs (top layer of each network) are maximally correlated.

account the nonlinear relationship between variables, but the deep representations are not necessarily correlated with each other. Jiang et al. [21], [30] developed a DCCA neural network framework combining two Stacked Autoencoders (SAEs) and CCA to maximize the correlation of deep representations. They applied SAEs to extract nonlinear features from local units for internal monitoring and used mutual information to filter neighborhood variables. DCCA was then employed to jointly learn external correlation features between local and neighborhood variables, enabling the construction of both external and global monitoring models. However, the correlation information of neighboring sub-blocks was not adequately considered when constructing the local monitoring model. In large-scale manufacturing processes, the interconnection of sub-processes means that there is significant information in the process variables of one sub-block that can influence those of another. When a fault occurs, it accumulates and amplifies as it propagates through the sub-processes. Additionally, these multiple sub-processes exhibit clear spatial distributions, dynamic interactions among the process variables, and quality heritability characteristics across the sub-processes. Hence, for the quality-related monitoring of large-scale manufacturing processes, the following challenges must be addressed: 1) incorporating correlation information between sub-processes in the development of monitoring models, 2) collaborative extraction of temporal and spatial quality-related features, and 3) achieving efficient local-global monitoring and fault diagnosis.

III. Methodologies

This section comprehensively describes the proposed distributed quality-related fault monitoring framework that leverages a STDFFN. Firstly, the STDFFN is presented as extracting spatio-temporal quality-related features. Subsequently, a distributed STDFFN approach in the framework is proposed for large-scale manufacturing processes, enabling comprehensive local-global quality-related fault monitoring across the whole process.

A. Spatio-temporal Deep Feature Fusion Network

Considering the coupling relationship between process variables and extracting the temporal and spatial features of the data, we adopt a two-stage spatio-temporal feature extraction method based on STDFFN. The basic components of the model include CNN, LSTM neural network, and two CCA layers located in the top layers of CNN and LSTM, which is shown in Fig.2. During the training process, the network parameters of CNN and LSTM are learned in two stages. Typically, the process and quality samples are designated as u and y, respectively. Following this, the historical process and quality data from the normal operating state are divided into two sets, $U = [u_1, u_2, \ldots, u_N] \in \Re^{N\times P}$ and $Y = [y_1, y_2, \ldots, y_N] \in \Re^{N\times Q}$, where P and Q are the numbers of the variables, respectively, and N is the number of training samples, and they are normalized as follows:

$$u_i^* = \frac{u_i - \bar{u}}{\sigma_u} \tag{7}$$

$$y_i^* = \frac{y_i - \bar{y}}{\sigma_y}, \quad i = 1, 2, \ldots, N \tag{8}$$

where $\bar{u}$ and $\bar{y}$ are the means of the training set, σ_u and σ_y are the variances of the training set. For good extraction of temporal features, the data are batched using a sliding window.

$$U_{bat} = [U_1, U_2, \ldots, U_{N-l+1}] \in \Re^{(N-l+1)\times l\times P} \tag{9}$$

$$Y_{bat} = [Y_1, Y_2, \ldots, Y_{N-l+1}] \in \Re^{(N-l+1)\times l\times Q} \tag{10}$$

$$U_i = [u_{i-l+1}, u_{i-l+2}, \ldots, u_i] \in \Re^{l\times P} \tag{11}$$

$$Y_i = [y_{i-l+1}, y_{i-l+2}, \ldots, y_i] \in \Re^{l\times Q} \tag{12}$$

i denotes the current moment and l is the window size.

In the first-stage, the preprocessed process variable data are used as the input to the CNN to extract spatial features.

$$U_{conv} = \sigma(f_{conv}(U_i)) \tag{13}$$

where σ is the activation function, f_{conv} represents the convolution operation, $U_i \in R^{1\times l\times P}$ is the convolutional input, where we set one channels of the input data, $U_{conv} \in R^{C\times l\times P}$ is the convolutional output, and C is the number of convolutional channels. Recognizing that convolutional operations capture features by integrating cross-channel and spatial information, a CBAM is incorporated following the convolutional layer to bolster the representational power of CNNs. This module sequentially determines the attention weights for the channel and spatial dimensions separately and then applies these attention weights to the convolutional outputs to refine features. This process enhances the focus on salient features and diminishes the impact of less relevant ones. The entire attention mechanism can be encapsulated as

$$U_c = M_c(U_{conv}) \otimes U_{conv} \tag{14}$$

$$U_{cs} = M_s(U_c) \otimes U_c \tag{15}$$

where $M_c \in R^{C\times 1\times 1}$ and $M_s \in R^{1\times l\times P}$ are the channel and spatial attention weight matrix, respectively, and $\otimes$ denotes the element-by-element multiplication, with $U_c \in R^{C\times l\times P}$ being the output of the channel attention block, and $U_{cs} \in R^{C\times l\times P}$ being the final output of the CBAM block.

To effectively integrate global feature representations and quantify the correlation between spatial features and quality variables, the spatial features U_{cs} are expanded and serve as the inputs to a fully connected layer, which further reduces the dimensionality to produce the spatial feature $H_s \in R^{1\times S}$, where S represents the output dimension of the fully connected layer. Y_i is reconstructed to $Y_i \in R^{1\times W}$, with $W = l \times Q$. Eq. (6) quantifies the correlation between the spatial feature H_s and the batch quality variable. Subsequently, the first-stage optimization objective function is defined as

$$\theta_{conv}^* = \arg\max_{\theta_{conv}} corr(H_s, Y_i) \tag{16}$$

The parameters θ_{conv} encompass those of the CNN, CBAM, and the fully connected neural network, which are trained via the gradient descent method.

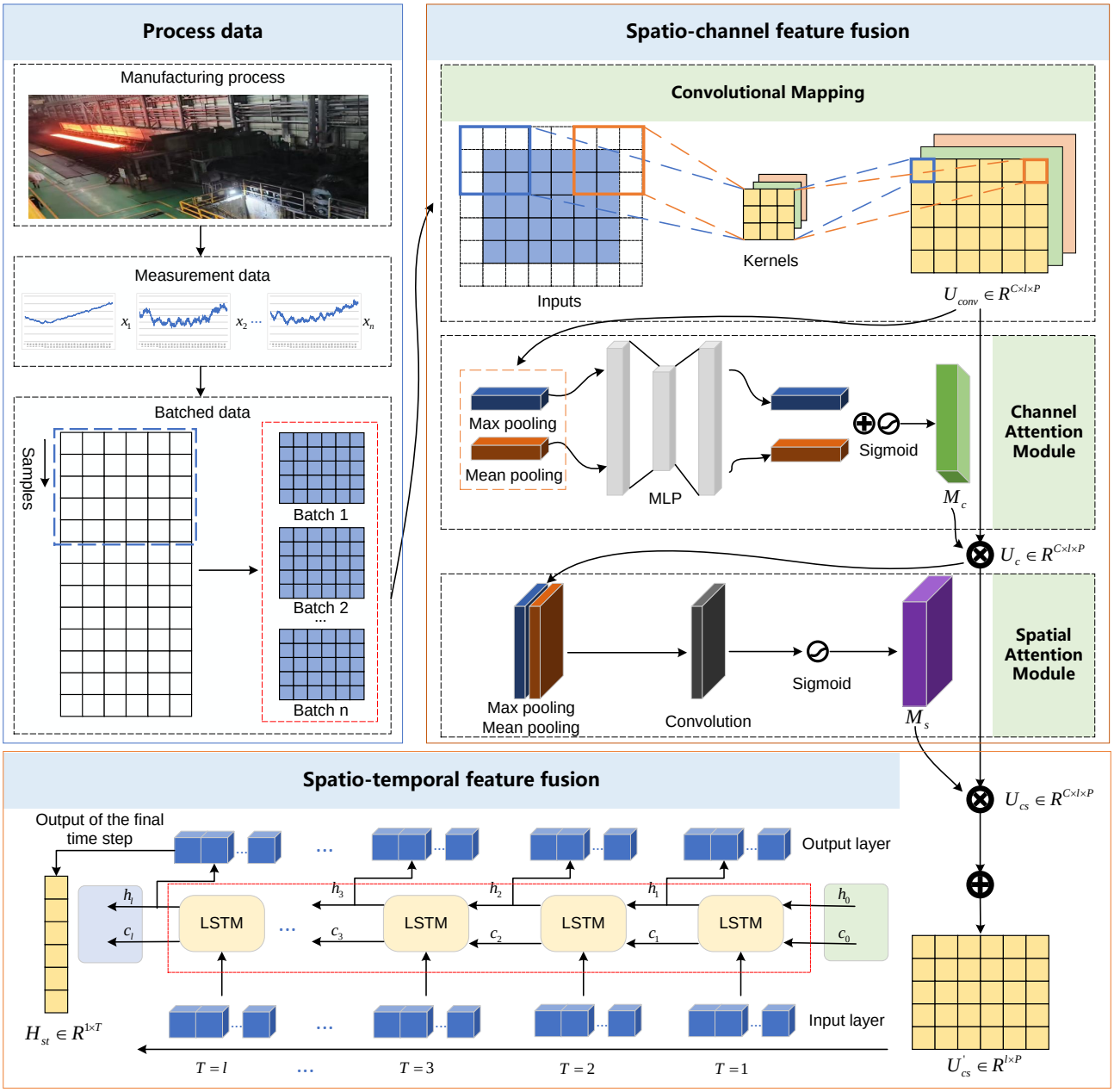


Fig. 2. The structure of the spatio-temporal deep feature fusion network.

The CNN and CBAM attention modules derived from the initial training phase are kept constant in the subsequent stage. The outputs, U_{cs}, are reconstructed to yield $U'_{cs} \in R^{l\times P}$ as the input for the LSTM (Remark 1), thereby extracting the spatio-temporal features of the process variables. Given the focus on the impact of process variables on the immediate output in actual production processes, we specifically obtain the output from the final time step of the LSTM as the ultimate set of spatio-temporal features acquired.

$$H_{st} = f_{lstm}\left(U'_{cs}\right) \tag{17}$$

where $H_{st} \in R^{1\times T}$, and T represents the output dimension of the LSTM, f_{lstm} means the LSTM operation. The overall correlation between the quality variable y_i and the spatio-temporal feature H_{st} at the current moment is calculated, which is used to determine the optimization objective function for the second stage.

$$\theta^*_{lstm} = \arg\max_{\theta_{lstm}} corr(H_{st}, y_i) \tag{18}$$

The network parameters θ^*_{lstm} are optimized using the gradient descent method.

Remark 1: U'_{cs} is derived from the reconstruction of U_{cs} by sequentially summing the multi-channel data along the channel axis, resulting in a data dimension of $R^{1\times l\times P}$. Since the LSTM requires two-dimensional input data, the reconstruction process set the channel dimension of 1 to yield U'_{cs}.

B. Local-global Monitoring Framework for Large-scale Manufacturing Processes

Large-scale manufacturing process is distinguished by multiple operation units, involving numerous variables and interactions between different processes. The operating units within the production process exhibit complex characteristics. The whole process is decomposed into various sub-blocks which can enhance local monitoring effectiveness and reduce the computational complexity of the models. Effective process decomposition should consider the physical constraints, topological structure, and the flow of quality-related information among operational units. First, the whole process is divided into B sub-blocks based on process knowledge, considering the sequence and spatial arrangement of the processes, as well as the inter-variable coupling. Within each sub-block, multiple sensors are deployed to gather diverse data. For instance, in a manufacturing process that comprises 7 operating units, $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \rightarrow G$, these units are empirically divided into 3 sub-blocks, denoted as (A, B, C), (D, E), (F, G). Due to the serial nature of the process, the process and quality information in sub-block $b-1$ will be transmitted to sub-block b and affect its process variables. There are coupling variables related to sub-block b in sub-block $b-1$, and the process variables that have a significant impact on quality variables are the primary focus of attention. Therefore, the mRMR criterion is employed to select quality-related coupling variables from sub-block $b-1$ for reconstruction of sub-block b, ensuring that the selected coupling variables have the strongest correlation with the quality variables, while introducing minimal redundancy to sub-block b.

Assume that the historical process variable data are $U = [\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_B] \in \Re^{N \times M}$, where N is the number of samples, and M is the number of variables in the whole production process. $\hat{U}_b = [u_{b,1}, u_{b,2}, \ldots, u_{b,N}] \in \Re^{N \times P_b} (b = 1, 2, \ldots, B)$ is the process variable of the bth sub-block and P_b is the number of variables in sub-block b , $\sum_{b=1}^{B} P_b = M$. $Y \in R^{N \times Q}$ represents the matrix of quality variables, with Q being the number of quality variables. The optimization problem for mRMR, utilizing the incremental search approach, is as follows:

$$\max_{u_{b,j} \in \hat{U}_{b-1}} \left[I\left(u_{b,j}; Y\right) - \frac{1}{m} \sum_{u_{b,i} \in \hat{U}_b} I\left(u_{b,j}; u_{b,i}\right) \right] \tag{19}$$

where $I(;)$ denotes mutual information, and m represents the number of features in the sub-block b. Given the existing sub-block b, Eq. (19) is used to compute and rank each feature in the preceding sub-block $b-1$. The top n variables are then chosen as the relevant variables for reconstructing the sub-block b.

Models are developed utilizing the local process variable $\hat{U}_b$ and the quality variable Y to capture the complex relationships within the local units. Eq. (16) and (18) serve as the quality-related constraints for training the network model. Let the features extracted by STDFFN be denoted as $H_{st}^b \in R^{(N-l+1) \times T_b}$, where T_b represents the dimensionality of the features extracted by the bth sub-block, encompassing quality-related information from the process variable $\hat{U}_b$. We construct the following statistics to monitor the status of the local unit.

$$T_b^2 = \left(H_{st}^b - \overline{H_{st}^b}\right)^{\mathrm{T}} \Lambda_{H_{st}}^{b\,-1} \left(H_{st}^b - \overline{H_{st}^b}\right) \tag{20}$$

where $\overline{H_{st}^b}$ and $\Lambda_{H_{st}}^b$ are the mean and variance of the acquired features, respectively. Kernel density estimation (KDE) is employed to infer an arbitrary probability density distribution from the data [35]. KDE is selected to establish the monitoring threshold T_{bth}^2 for sub-block b, utilizing statistical data from historical normal operations.

Following the aforementioned steps, we can derive the quality-related information and local monitoring statistics for each sub-block, enabling the surveillance of variable fluctuations across multiple sub-blocks. However, the monitoring outcomes from different sub-blocks may not always align. To synthesize the monitoring results from multiple sub-blocks, Bayesian inference is utilized to develop a comprehensive decision indicator [35]. The state of the samples can be assessed based on the conditional probability, that is

$$P_{T_{bth}^2}^b\left(F|u_b\right) = \frac{P_{T_{bth}^2}^b\left(u_b|F\right) P_{T_{bth}^2}^b\left(F\right)}{P_{T_{bth}^2}^b\left(u_b\right)} \tag{21}$$

$$P_{T_{bth}^2}^b\left(u_b\right) = P_{T_{bth}^2}^b\left(u_b|N\right) P_{T_{bth}^2}^b\left(N\right) + P_{T_{bth}^2}^b\left(u_b|F\right) P_{T_{bth}^2}^b\left(F\right) \tag{22}$$

$P_{T_{bth}^2}^b(F)$ and $P_{T_{bth}^2}^b(N)$ represent the prior probabilities of normal and faulty samples, respectively, typically set at confidence levels α and $(1-\alpha)$. Let $u_b \in R^{1 \times P_b}$ represent a single sample from sub-block b. $P_{T_{bth}^2}^b\left(u_b|F\right)$ and $P_{T_{bth}^2}^b\left(u_b|N\right)$ are the conditional probabilities of the samples being in faulty or normal states.

$$P_{T_{bth}^2}^b\left(u_b|F\right) = \exp\left(-\frac{T_{bth}^2}{T_b^2}\right) \tag{23}$$

$$P_{T_{bth}^2}^b\left(u_b|N\right) = \exp\left(-\frac{T_b^2}{T_{bth}^2}\right) \tag{24}$$

Based on Eq. (21), (22), and (23), the integrated monitoring statistics are as follows:

$$T_{global}^2 = \frac{\sum_{b=1}^{B} \left(P_{T_{bth}^2}^b\left(u_b|F\right) P_{T_{bth}^2}^b\left(F|u_b\right)\right)}{\sum_{b=1}^{B} P_{T_{bth}^2}^b\left(u_b|F\right)} \tag{25}$$

The confidence level α serves as the control limit for the integrated monitoring statistics.

During the online monitoring phase, the status of the current sample is determined using the following strategy:

1) $T_b^2 \geq T_{bth}^2$ and $T_{global}^2 \geq \alpha$ means a fault has occurred in the production process and is located in subblock b;

2) $T_b^2 \geq T_{bth}^2$ and $T_{global}^2 < \alpha$ means that sub-block b has a fault.

3) $T_b^2 < T_{bth}^2$ and $T_{global}^2 < \alpha$ means the production process is operating under normal conditions.

When quality-related faults are detected within a continuous production process, the variable oblivion contribution plot-based (VOB-CP) is utilized to locate the faulty variable [36]. $u_{b,i}$ is the ith variable in the bth sub-block. To determine the contribution of $u_{b,i}$, we set the values of all other variables

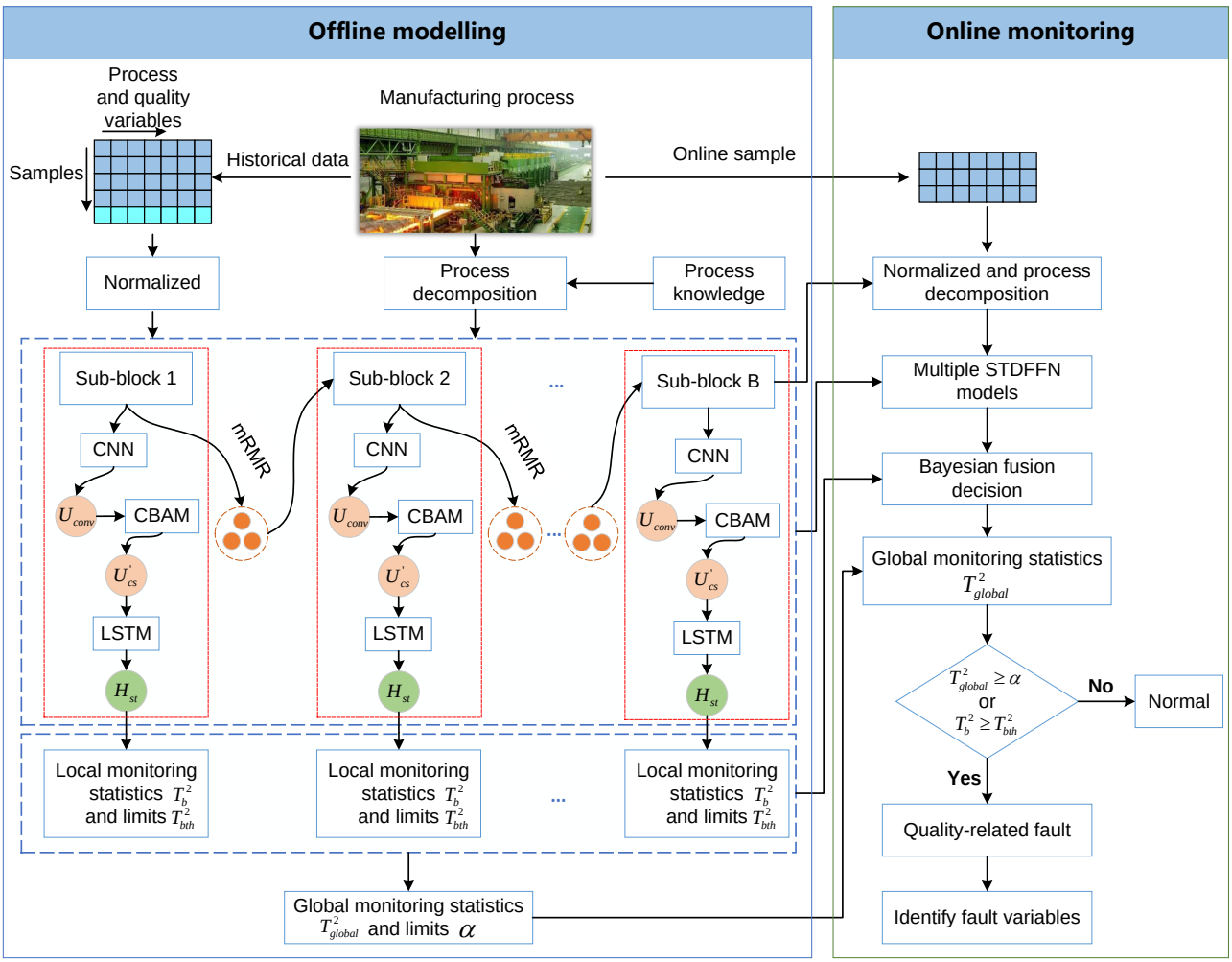


Fig. 3. STDFFN-based local-global monitoring framework for large-scale manufacturing processes.

to zero, thereby making the statistic $T^2_{b,i}$ focus exclusively on $u_{b,i}$. The contribution is then assessed by comparing the variations under normal and faulty conditions.

$$Con^b_{i,j} = \frac{T^2_{b,i,j} - AVE(T^2_{b,i,normal})}{VAR(T^2_{b,i,normal})} \tag{25}$$

where $T^2_{b,i,j}$ is the statistic for the jth sample, $AVE(T^2_{b,i,normal})$ and $VAR(T^2_{b,i,normal})$ are the mean value and the standard deviation of $T^2_{b,i}$ under normal operating conditions, repectively.

The whole monitoring scheme utilizing STDFFN is depicted in Fig.3. During the offline phase, individual monitoring models are developed and control limits are established for each sub-block. Simultaneously, fusion monitoring statistics are formulated as well as the global control limit is determined. In the online monitoring phase, the latest sample's monitoring statistics are computed, and the presence of local and global faults is assessed based on the established decision-making logic.

IV. Case Study

In this section, the effectiveness of the proposed distributed quality-related monitoring method is verified by an actual HSMP and compared with other base-line methods. All experiments were conducted on Python 3.8.18, PyTorch 1.11.0, and CUDA 11.3, with the computer configured as follows: Windows 10 operating system, Intel (R) Core (TM) i5-8300H CPU @ 2.30GHz, NVIDIA GeForce GTX 1050Ti GPU, and 8GB of RAM.

A. Description of HSMP

HSMP is a typical large-scale sequential manufacturing process, primarily composed of furnaces, roughing mills, transfer tables, finishing mills, laminar cooling, and coilers in series. The sub-processes are geographically dispersed and interconnected through information and quality flows, resulting in dynamic interdependencies among variables and a cascading effect on quality across sequentially linked sub-processes [37], [38]. Among them, the finishing mill process (FMP) has an important influence on the quality indicators such as thickness and convexity of the final product. Therefore, FMP is used as

TABLE I

DESCRIPTION OF FMP PROCESS AND QUALITY VARIABLES

Variables	Type	Description	Unit
G1∼G7	Process Variables	Average gap of the ith stand, $i = 1, 2, \ldots, 7$	mm
F1∼F7	Process Variables	Total force of the ith stand, $i = 1, 2, \ldots, 7$	MN
B2∼B7	Process Variables	Work roll bending force of the ith stand, $i = 2, \ldots, 7$	MN
y	Quality Variable	Exit strip thickness of the finishing mill	mm

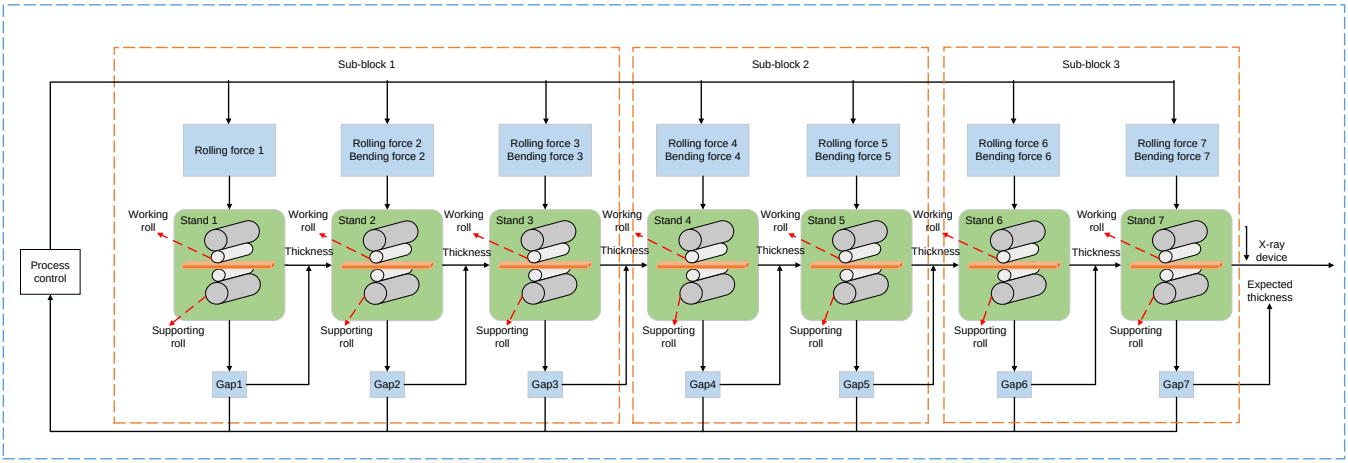


Fig. 4. Flow diagram of the FMP.

TABLE II

DESCRIPTION OF TYPICAL QUALITY-RELATED FAULTS IN FMP

Fault No.	Description
Fault 1	Roll gap control actuator fault of the 4th stand
Fault 2	Cooling water valve fault between the 2nd and 3rd stands

TABLE III

SUB-BLOCK DIVISION RESULTS IN FMP

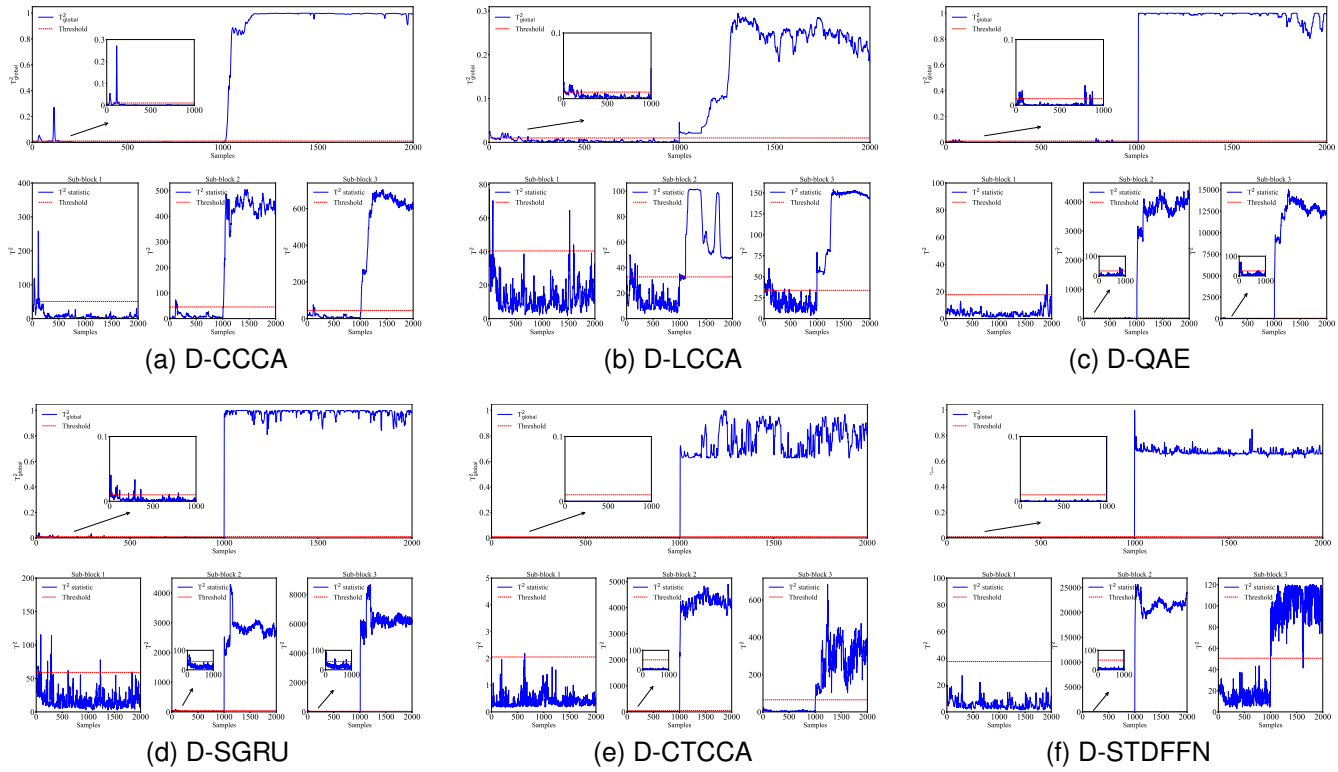
Sub-block No.	Variables
Block 1	G1,G2,G3,F1,F2,F3,B2,B3
Block 2	F1,F2,F3,G4,G5,F4,F5,B4,B5
Block 3	G4,G5,F4,F5,G6,G7,F6,F7,B6,B7

the verification background in this paper. The FMP consists of seven finishing stands in series, each consisting of rolls, loopers, and hydraulics, including key process variables such as rolling forces, roll gaps, and bending roll forces, and a typical FMP is shown in Fig.4. To validate the effectiveness of the proposed monitoring framework, normal and fault data of the FMP during the actual production process were collected with a sampling period of 0.01 s. Table I provides the description of the quality and process variables, and the two typical quality-related faults are shown in Table II.

B. Monitoring Results and Analysis

The strip pass through the FMP, sequentially rolling through seven finishing stands, ultimately achieving the desired product quality specifications (such as final exit thickness, etc.). The finishing mill stands are arranged in a sequential chain, where a malfunction in one stand can impact the operations of the subsequent stands. Therefore, the spatial arrangement of the FMP is crucial to consider. As a large-scale manufacturing process, dividing the FMP into sub-blocks aids in capturing the temporal and spatial features of the process data, thereby enhancing the monitoring performence. Based on the above conditions, the FMP is divided into sub-blocks using a combination of process expertise and data analysis. Specifically, the seven stands are organized into three sub-blocks: Stands 1-3 as the first sub-block, stands 4-5 as the second sub-block, and stands 6-7 as the third sub-block. In addition, the sequentially linked FMP sub-processes exchange information and quality flows, exhibiting distinct dynamic coupling and quality heritability among variables. The initial division outcomes are refined through reconstruction to achieve the final sub-block division, as detailed in Table III.

To assess the efficacy of the proposed approach, the Distributed STDFFN (D-STDFFN) is benchmarked against Distributed Convolutional CCA (D-CCCA), Distributed LSTM-CCA (D-LCCA), Distributed Convolutional TCN-CCA (D-CTCCA), Distributed Quality-driven SAE (D-QAE) [18] and Distributed Siamese Gated Recurrent Unit (D-SGRU) [38]. For the model described in this paper, a three-layer 2D-CNN model is employed with the number of channels set to 4 and the kernel size set to 3. To maintain consistent input and output dimensions for the 2D-CNN, the convolution step length is set to 1 and padding is set to 1. The channel attention module of CBAM adopts two two-layer fully connected neural networks, with input and output dimensions set to 4. The spatial attention module adopts one layer of 2D-CNN, whose kernel size is set to 3, and the step length and padding are both 1. The hidden layer dimension of the LSTM is set to 16, the time step is set to 50, and the number of layers is set to 2. D-CCCA and D-LCCA are ablation experiments, where the LSTM and 2D-CNN are removed, respectively.



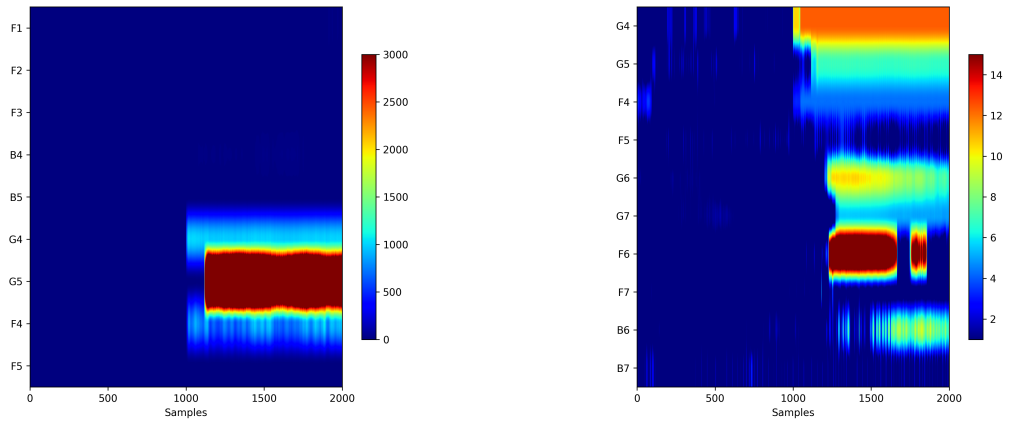
(a) D-CCCA (b) D-LCCA (c) D-QAE

(d) D-SGRU (e) D-CTCCA (f) D-STDFFN

Fig. 5. Monitoring results of different methods for Fault 1.

During the offline training phase, due to the faults occurring in different target thickness production processes, 2000 and 1500 samples from two distinct normal production conditions were collected for offline modeling, with each sample comprising 20 process variables and 1 quality variable. The size of the sliding window used is $l = 50$. Additionally, the confidence level is set to 99%. The activation function used is the ReLU function, and the model employs the Adam gradient update strategy with a learning rate of $lr = 0.001$. In the online monitoring stage, two sets of process variable samples were collected for online monitoring, with data dimensions of [2000, 50, 20] and [1500, 50, 20], respectively, where 2000 and 1500 are batch quantities, 50 is the sliding window size, and 20 is the number of process variables. Within this study, the False Alarm Rate (FAR), Fault Detection Rate (FDR) [27], and Fault Detection Delay (FDD) are employed as key performances for monitoring. The FDD is defined as the time lag between the initial detection of a fault and the emergence of the first defective sample [39]. And we evaluate the average detection performance of the method for multiple faults using average FAR, average FDR, and average FDD.

1) Case 1: Fault 1 is caused by abnormal roll gap control



(a) Sub-block 2 (b) Sub-block 3

Fig. 6. Fault diagnosis results of Fault 1.

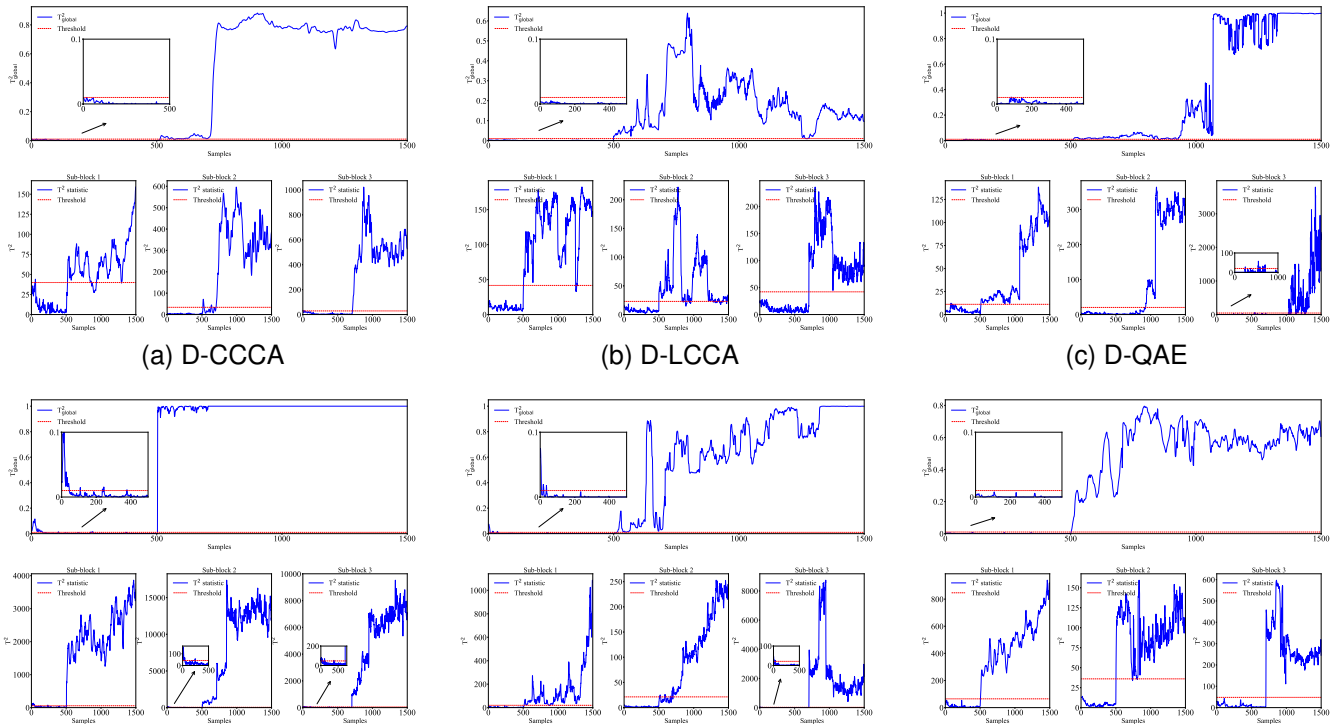


Fig. 7. Monitoring results of different methods for Fault 2.

TABLE IV
MONITORING RESULTS OF COMPARATIVE METHODS

Method	Fault 1			Fault 2			Average		
	FAR(%)	FDR(%)	FDD(s)	FAR(%)	FDR(%)	FDD(s)	FAR(%)	FDR(%)	FDD(s)
D-CCCA	5.30	98.60	0.14	**0.00**	97.90	0.14	2.65	98.25	0.14
D-LCCA	8.20	**100.0**	**0.00**	**0.00**	99.30	**0.00**	4.10	99.65	**0.00**
D-QAE	3.60	98.90	0.11	1.00	97.90	0.11	2.30	98.40	0.11
D-SGRU	5.40	99.90	0.01	11.20	99.60	0.04	8.30	99.75	0.025
D-CTCCA	**0.00**	99.60	0.04	3.20	98.70	0.13	1.60	99.15	0.085
D-STDFFN	0.10	**100.00**	**0.00**	**0.00**	**99.70**	0.03	**0.05**	**99.85**	0.015

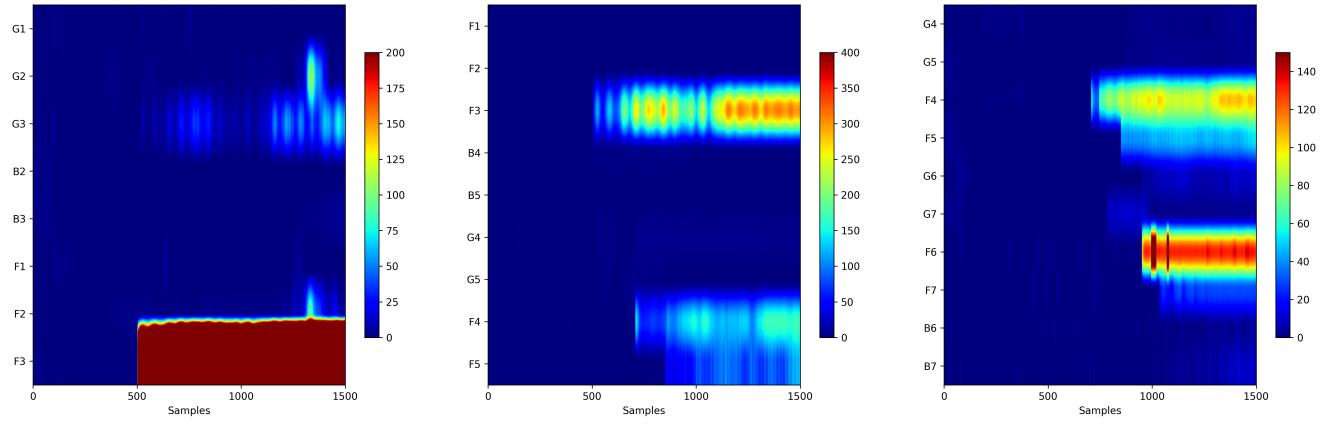


Fig. 8. Fault diagnosis results of Fault 2.

in stand 4, which directly affects the values of the rolling gap G4 and the rolling force F4. Due to the process characteristics of FMP, subsequent process variable sampling values will gradually be affected, ultimately affecting product quality.

TABLE V

MONITORING RESULTS OF PARAMETER SENSITIVITY EXPERIMENTS FOR STDFFN

Case type	Parameter setting	Fault 1		Fault 2		Average	
		FAR(%)	FDR(%)	FAR(%)	FDR(%)	FAR(%)	FDR(%)
Case 1	lr=0.0001	1.90	**100.00**	4.40	98.50	3.15	99.25
	lr=0.001	**0.10**	**100.00**	**0.00**	**99.70**	**0.05**	**99.85**
	lr=0.01	4.20	**100.00**	12.20	95.70	8.20	97.85
	lr=0.1	4.70	**100.00**	**0.00**	78.80	2.35	89.40
Case 2	kernel size=(3, 3), padding=1	**0.10**	**100.00**	**0.00**	**99.70**	**0.05**	**99.85**
	kernel size=(5, 5), padding=2	3.80	**100.00**	7.20	98.10	5.50	99.05
	kernel size=(7, 7), padding=3	3.30	**100.00**	11.80	98.40	7.55	99.20
Case 3	LSTM Hidden Layer Dimension=8	3.90	**100.00**	4.40	99.70	4.15	99.85
	LSTM Hidden Layer Dimension=16	**0.10**	**100.00**	**0.00**	**99.70**	**0.05**	**99.85**
	LSTM Hidden Layer Dimension=24	3.70	**100.00**	4.40	99.60	4.05	99.80
	LSTM Hidden Layer Dimension=32	4.90	**100.00**	8.60	98.90	6.75	99.45

Therefore, the fault appears in sub-blocks 2 and 3. Fault 1 starts from the 1001st sample point, and after sliding window processing, it will include faulty samples from the 1001st batch and continue until the end of sampling. Fig.5 presents the local and global quality anomaly monitoring results of six different methods. As ablation experiments, D-CCCA demonstrates the lowest FDR and the highest FDD, while D-LCCA shows the highest FAR, particularly severe in sub-block 3. The results of D-QAE show some improvement compared to D-CCCA, but the difference remains marginal. It is evident that when temporal information is neglected, both methods suffer from significant drawbacks in terms of FDR and FDD. D-SGRU significantly improves the FDR but still maintains a relatively high FAR. These methods, which only focus on either spatial or temporal features, exhibit varying impacts on FAR, FDR, and FDD. In contrast, D-CTCCA and D-STDFFN, which simultaneously extract both temporal and spatial features of process variables, effectively enhance the FDR and reduce the FAR. However, D-CTCCA still exhibits noticeable detection delays, likely due to the limitation of TCN in capturing temporal features. The proposed monitoring framework in this study achieves the best local and global quality anomaly monitoring results, which significantly aids in on-site maintenance and decision-making. Since the fault occurs in sub-block 2 and sub-block 3, we calculated the variable oblivion contributions for these sub-blocks. As shown in Fig.6, the contribution plot for sub-block 2 reveals that G4 and F4 are initially identified as fault variables, subsequently affecting variable F5, and ultimately leading to anomalies in F6, G6, G7, and B6 in sub-block 3. This observation aligns with the actual scenario of Fault 1 and clearly illustrates the propagation of the fault across sub-blocks.

2) Case 2: Fault 2 involves an anomaly in the cooling water controller between stands 2 and 3, leading to a fault in the cooling water valve, which further affects the rolling force of stand 3 and subsequent stands, causing changes in the rolling gap and ultimately affecting product quality. Therefore, the fault appears in all three sub-blocks. Fault 2 starts from the 501st sample and will include faulty samples from the 501st window onwards, continuing until the end of sampling. Fig.7 shows the local and global monitoring results of six methods for Fault 2, with D-CCCA and D-QAE having the worst FDRs. Although D-QAE has improved FDD, the effect is not ideal. D-LCCA and D-SGRU significantly improve FDRs and FDDs by extracting temporal features. However, D-SGRU has a large number of false alarm before detecting faults, which is not benefit to the overall monitoring. D-CTCCA still shows shortcomings in capturing temporal features, as evidenced by significant delays in the monitoring results of the sub-block 2. This is similar to the monitoring results of D-CCCA. The proposed method has the best FAR and FDR, and also has a good improvement effect on FDD. We calculated the variable oblivion contributions of three sub-blocks as shown in Fig.8. F3 and G3 are identified as initial fault variables in sub-block 1, further affecting F4 and F5, and ultimately leading to anomalies in F6 and F7. This also corresponds to the actual propagation of the fault in FMP.

Table IV shows the monitoring results of different methods. It can be seen that the proposed method in this paper is almost optimal in FAR, FDR, and FDD, demonstrating the best monitoring performance for quality-related faults. The global FDR for Fault 1 is 100%, and for Fault 2, it is 99.70%.

C. Parameter Sensitivity Analysis

In order to objectively verify the effectiveness of the proposed method, we conducted further parameter sensitivity experiments based on the aforementioned normal and fault data. The parameter sensitivity analysis primarily focuses on the learning rate, 2D-CNN kernel size, and LSTM hidden layer dimensions. The confidence level, size of sliding window, network layers, and other parameter settings are consistent with the above validation experiments. When selecting parameters, we pay more attention to FAR and FDR which are used as evaluation indexes, and the average FAR and average FDR of multiple faults are considered as well. The specific operations are as follows:

1) Case 1: 2D-CNN Kernel Size = (3, 3), LSTM Hidden Layer Dimensions = 16, Varying Learning Rate lr.

2) Case 2: LSTM Hidden Layer Dimensions = 16, Learning Rate $lr = 0.001$, Varying 2D-CNN Kernel Size and Padding. (Ensure that the input and output dimensions of 2D-CNN are consistent.)

3) Case 3: 2D-CNN Kernel Size = (3, 3), Learning Rate $lr = 0.001$, Varying LSTM Hidden Layer Dimensions.

Table V presents the monitoring results of parameter sensitivity experiments for STDFFN. With different parameter

changes, STDFFN maintains good FDR on Fault 1, while FAR shows varying degrees of increase. However, the monitoring results of STDFFN on Fault 2 show more significant changes. As the learning rate changes, there is a fluctuation in the FAR of Fault 2. FDR shows a significant decrease, which may be due to an inappropriate learning rate that prevented the model from fully converging. When we try to increase the kernel size of the 2D-CNN to increase the receptive field, both the FAR and FDR of Fault 2 change in a negative direction. The excessively large convolution kernel captures more redundant features, reducing the generalization ability of the model. We also observe the monitoring effect of different numbers of hidden layer dimensions in LSTM. When the number of hidden layer dimensions changes within a small range, there is no significant impact on the FDR, but the FAR increases. When this parameter significantly increases, both the FAR and FDR of Fault 2 show obvious poor performance.

According to the comprehensive analysis of parameter sensitivity experimental results, it can be seen that the parameter adjustments we made during model training are more conducive to the convergence of the model and the improvement of the monitoring effect. While providing an objective explanation for our hyperparameter settings, the effectiveness of the proposed method is also validated.

V. Conclusion

In this paper, a distributed quality anomaly monitoring framework for large-scale manufacturing processes based on a spatio-temporal deep feature fusion network is proposed. Firstly, sub-blocks are divided based on process knowledge, and the mRMR is used to reconstruct. The sample data are batch-processed using a sliding window to construct the relevant training dataset, so that the local information fully considers the coupling association between sub-blocks and the temporal dynamic information, and solves the problem of quality heritability among sub-blocks. Then, the STDFFN model is designed to achieve collaborative extraction of spatio-temporal quality-related features for large-scale manufacturing processes. Finally, local and global monitoring results are obtained using Bayesian inference. The contributions of relevant variables based on variable oblivion contribution plots are calculated to achieve efficient local-global integrated monitoring and fault diagnosis. The feasibility and effectiveness of the proposed framework are verified through the HSMP.

Although the proposed method achieves excellent monitoring performance, it has a high computational complexity during training phase, which could be a challenge for real-time monitoring in some resource-constrained industrial environments. In the multimodal manufacturing process, the dynamic changes in process variables and the complexity of inter-sub-block correlations will significantly affect the precise division and reconstruction of sub-blocks, affecting the stability and reliability of monitoring results. For future work, we will study lightweight network structures to reduce the computational complexity of models and consider using transfer learning methods to address the dynamic changes of process variables in multimodal manufacturing processes. Moreover, we will devote ourselves to the problem of fault root cause diagnosis and fault propagation in the case of quality anomaly in large-scale manufacturing processes.

References

[1] C. Zhang, J. Dong, H. Zhang, and K. Peng, "Distributed Quality-Related Process Monitoring Framework Using Parallel DVIB-VAE-mRMR for Large-Scale Processes," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-11, Jan. 2023.

[2] 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, Oct. 2023.

[3] S. X. Ding, S. Yin, K. Peng, H. Hao, and B. Shen, "A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2239–2247, Nov. 2013.

[4] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," *Control Eng. Pract.*, vol. 62, pp. 11–21, Mar. 2017.

[5] X. Cao and K. Peng, "Multiphase Degradation Modeling and Remaining Useful Life Prediction Considering Aleatory and Epistemic Uncertainty," *IEEE Sensors J.*, vol. 23, no. 22, pp. 27757-27770, Nov. 2023.

[6] X. Xie, W. Sun, and K. C. Cheung, "An advanced PLS approach for key performance indicator related prediction and diagnosis in case of outliers," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2587–2594, Apr. 2016.

[7] K. X. Peng, L. Ma, and K. Zhang, "Review of quality-related fault detection and diagnosis techniques for complex industrial processes," *Acta Automatica Sinica*, vol. 43, pp. 349–365, 2017.

[8] Q. 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, Jul. 2017.

[9] Y. J. Liu, T. Chen, and Y. Yao, "Nonlinear process monitoring and fault isolation using extended maximum variance unfolding," *J. Process Control.*, vol. 24, no. 6, pp. 880–891, Apr. 2014.

[10] L. J. Luo, S. Y. Bao, J. F. Mao, and D. Tang, "Nonlinear process monitoring based on kernel global-local preserving projections," *J. Process Control.*, vol. 38, pp. 11–21, Dec. 2015.

[11] P. Tang, K. X. 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, Jan. 2021.

[12] J. L. Zheng, C. H. Zhao, and F. R. Gao, "Retrospective comparison of several typical linear dynamic latent variable models for industrial process monitoring," *Comput. Chem. Eng.*, vol. 157, 2022, Art. no. 107587.

[13] P. E. P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 36–45, Feb. 2010.

[14] 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.

[15] D. Li, H. Zhang, L. An, J. Dong and K. Peng, "A Quality-Related Fault Detection Method Based on Deep Spatial–Temporal Networks and Canonical Variable Analysis for Industrial Processes," *IEEE Sensors J.*, vol. 24, no. 18, pp. 29047-29055, Sept. 2024.

[16] X. F. Yuan, C. Ou, Y. L. Wang, C. H. Yang, and W. H. Gui, "Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE," *Neural Comput.*, vol. 396, pp. 375–382,Apr. 2019.

[17] L. J. Feng, C. H. Zhao, and Y. X. Sun, "Dual attention-based encoder-decoder: A customized sequence-to-sequence learning for soft sensor development," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3306–3317, Aug. 2021.

[18] X. F. Yuan, J. Zhou, B. Huang, Y. L. Wang, C. H. Yang, and W. H. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 375–3730, Jun. 2020.

[19] Q. Q. Sun and Z. Q. Ge, "Gated stacked target-related autoencoder: A novel deep feature extraction and layerwise ensemble method for industrial soft sensor application," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3457–3468, May 2022.

[20] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep Canonical Correlation Analysis," *Proc. Int. Conf. Mach. Learn.*, pp. 1247–1255, Jun. 2013.

[21] Q. Jiang, S. Yan, X. Yan, S. Chen, and J. Sun, "Data-driven individual–joint learning framework for nonlinear process monitoring," *Control Eng. Pract.*, vol. 95, Nov. 2019.
[22] X. Yan, D. Jiang, L. Xiang, Y. Xu, Y. Wang, "CDTFAFN: A novel coarse-to-fine dual-scale time-frequency attention fusion network for machinery vibro-acoustic fault diagnosis," *Inform. Fusion*, vol. 112, Dec. 2024.
[23] C. Zhang, J. Dong, K. Peng, R. Sun, "Spatio-temporal feature extraction network based multi-performance indicators synergetic monitoring method for complex industrial processes," *Expert Syst. Appl.*, vol. 257, Dec. 2024.
[24] J. Liu, J. Mu, H. Sun, C. Dai, Z. Ji and I. Ganchev, "BFG&MSF-Net: boundary feature guidance and multi-scale fusion network for thyroid nodule segmentation," *IEEE Access*, vol. 12, pp. 78701-78713, May 2024.
[25] J. L. Zhu, Z. Q. Ge, and Z. H. Song, "Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with Big Data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1877–1885, Aug. 2017.
[26] C. Li, G. Li, X. Chen, P. Zhou, and X. He, "A multiblock kernel dynamic latent variable model for large-scale industrial process monitoring," *IEEE Trans. Instrum. Meas.*, vol. 71, Oct. 2022.
[27] Z. W. Chen et al., "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2710–2720, May 2019.
[28] L. Yao and Z. Ge, "Distributed parallel deep learning of hierarchical extreme learning machine for multimode quality prediction with big process data," *Eng. Appl. Artif. Intell.*, vol. 81, pp. 450–465, May 2019.
[29] M. Rong, H. Shi, and S. Tan, "Large-scale supervised process monitoring based on distributed modified principal component regression," *Ind. Eng. Chem. Res.*, vol. 58, no. 39, pp. 18223–18240, Oct. 2019.
[30] Q. Jiang, S. Yan, H. Cheng and X. Yan, "Local–Global Modeling and Distributed Computing Framework for Nonlinear Plant-Wide Process Monitoring With Industrial Big Data," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 8, pp. 3355-3365, Aug. 2021.
[31] Z. Chen, S. X. Ding, K. Zhang, Z. Li, and Z. Hu, "Canonical correlation analysis-based fault detection methods with application to alumina evaporation process," *Control Eng. Pract.*, vol. 46, pp. 51–58, Jan. 2016.
[32] Q. Jiang, S. X. Ding, Y. Wang, and X. Yan, "Data-driven distributed local fault detection for large-scale processes based on the GA-regularized canonical correlation analysis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 8148–8157, Oct. 2017.
[33] Q. Jiang and X. Yan, "Learning Deep Correlated Representations for Nonlinear Process Monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6200-6209, Dec. 2019.
[34] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *in Lect. Notes Comput. Sci.*, 2018, pp. 3–19.
[35] 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.
[36] J. Zhu, H. Shi, B. Song, Y. Tao and S. Tan, "Convolutional Neural Network Based Feature Learning for Large-Scale Quality-Related Process Monitoring," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4555-4565, Jul. 2022.
[37] L. Ma, J. Dong, K. X. Peng, and C. F. Zhang, "Hierarchical monitoring and root-cause diagnosis framework for key performance indicator-related multiple faults in process industries," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2091–2100, Apr. 2019.
[38] C. F. Zhang, K. X. Peng, J. Dong, and L.Miao, "A comprehensive operating performance assessment framework based on distributed Siamese gated recurrent unit for hot strip mill process," *Appl. Soft Comput.*, vol. 132, Dec. 2022.
[39] K. Zhang, Y. A. W. Shardt, Z. W. Chen, and K. X. Peng, "Using the expected detection delay to assess the performance of different multivariate statistical process monitoring methods for multiplicative and drift faults," *ISA Trans.*, vol. 67, pp. 56–66, Nov. 2016.