Causal Disentangled Graph Neural Network for Fault Diagnosis of Complex Industrial Process

Ruonan Liu, *Senior Member, IEEE*, Quanhu Zhang, Di Lin, *Member, IEEE*, Weidong Zhang, *Senior Member, IEEE* and Steven X. Ding

***Abstract*—Graph neural networks (GNN) are good at capturing the intricate topologies and dependencies among components and are outstanding in fault diagnosis tasks of complex industrial process. Bias substructures consisting of irrelevant sensor signals and noise data are simpler compared to causal substructures consisting of fault signals, and GNNs tend to utilize the letter to quickly achieve low loss. However, spurious correlations in the bias substructures will mislead predictions. To address this issue, this study takes the disentanglement of causal and bias substructures as the key to improve model stability. A causal disentangled graph neural network (CDGNN) is proposed. First, sensor signals are transformed into graph data employing an attention mechanism to capture the interactions between them. Then, a causal disentanglement learning module is designed to extract causal subgraphs from input graphs. Finally, causal subgraph features from different source machines are aggregated to form a complete graph representation. Experimental results on two complex industrial datasets indicate that CDGNN is an effective and stable method for fault diagnosis.**

***Index Terms*—Complex industrial process, Fault diagnosis, Graph neural networks, Causal disentanglement.**

I. Introduction

FAULT diagnosis is one of the key techniques for prognostics and health management, serving a vital function in guaranteeing the secure and efficient operation of complex systems as well as reducing maintenance costs [1]. Deep learning-based intelligent fault diagnosis techniques can adaptively extract features from vibration signals, providing superior diagnosis results while significantly saving time and effort. Data-driven fault diagnosis methods has become a widely adopted approach to address challenges posed by increasing complexity of industrial systems [2].

This research was partly supported by the National Key R&D Program of China under Grant No. 2022ZD0119900, the National Natural Science Foundation of China under Grant Nos. 62206199 and U2141234, Shanghai Science and Technology Program under Grant No. 22015810300, Tianjin Applied Basic Research Project under Grant No. 22JCQNJC00410, Young Elite Scientist Sponsorship Program under Grant No. YESS20220409. (Corresponding author is Di Lin).

Ruonan Liu is with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (E-mail: ruonan.liu@sjtu.edu.cn).

Quanhu Zhang and Di Lin are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China, and Key Laboratory for Machine Learning of Tianjin, Tianjin 300350, China (E-mail: quanhuzhang@tju.edu.cn, Ande.lin1988@gmail.com).

Weidong Zhang is with School of Information and Communication Engineering, Hainan University, Haikou 570228, Hainan, P.R. China. He is also with the Department of Automation, Shanghai Jiaotong University, Shanghai 200240, P.R. China. (e-mail: wdzhang@sjtu.edu.cn)

Steven X. Ding is with the School of Automation, University of Duisburg-Essen, 47057 Duisburg, Germany (e-mail: steven.ding@uni-due.de).

In industrial systems, numerous sensors are utilized for the implementation of variable control, quality monitoring control, and safety control. Their locations are carefully designed based on the functional causal relationships between equipment, process components, and subsystems. The potential relationships among these sensors form an implicit interaction network, which belongs to non-Euclidean data. The majority of existing fault diagnosis methods employ gridded data, thereby failing to take into account the topology and interaction relationships between sensors. Learning and utilizing the topological structure and interaction relationships between sensors is beneficial to grasp the root causes and propagation process of faults [3].

Considering that each sensor measurement is represented as a node and implicit interactions between these nodes are represented as edges, it is feasible to describe the implicit interaction network constituted by sensors can be described by a structural attribute graph. In this way, fault information and propagation processes are stored in this graph. Naturally, different faults correspond to different graph structures. Graph neural networks (GNN) [4] are capable of effectively modelling and analysing graph data using the inductive biases associated with non-Euclidean representation, and have achieved outstanding results in fault diagnosis tasks [5].

The core of fault diagnosis task is the identification of fault categories, and GNN-based fault diagnosis on multiple sensor signals is a graph classification task [6]. Graph classification is commonly based on identifying the key substructure instead of analyzing the whole graph [7]. For example, the functions and properties of proteins are largely determined by the amino acid series and 3D structure, rather than the structure of entire protein [8]. Ideally, GNN only extracts information about key substructures and makes correct predictions.

Because of structural interactions of components, faults will propagate through industrial processes and even affect the operation of other equipment, causing readings from multiple sensors to deviate from normal. Correlation analysis is a crucial technique for fault detection and process monitoring because it reveals the sensors associated with a fault. However, it can only analyze the correlation of data and is susceptible to noise and irrelevant data. In GNN-based fault diagnosis, the substructure consisting of nodes whose feature inputs are fault signals is a key subgraph for determining the type of fault. But more irrelevant sensor signals and noise data will interfere with the model to learn key fault information.

To effectively establish the graph structure for different faults and to achieve stable prediction by GNN using only key subgraphs, we model the GNN-based fault diagnosis according

to the causal theory. Based on the causal theory, data consists of inherent causal and confounding (i.e., bias) information. Causal learning is the process of extracting causal information from data and using its causal invariance to make predictions [9]. Fault signals reflect the state of equipment, correlate with the type of fault, and are regarded as causal variables. In contrast, irrelevant sensor signal and noise signal are regarded as bias variables. When constructing the graph, the causal substructure corresponding to true label consists of causal variables, whereas bias variables constitute meaningless bias substructures that lead to biased graphs. In addition, noise data vary under different working conditions, but fault signals corresponding to the same fault are usually similar.

For large industrial system with hundreds of sensors, such as a nuclear power system [10]. Even in the event of a fault, only a few sensors show anomalous readings and most are not affected. These irrelevant sensors form a simpler bias substructure than causal substructure. Unfortunately, GNNs tend to utilize this simple substructure to quickly achieve low loss [11]. However, the bias substructure is independent of fault, which will result in large prediction biases. To improve the robustness and reliability of GNN-based fault diagnosis methods, it is necessary to solve following challenges: 1) How to identify causal and bias substructures in a sensor correlation graph? 2) How to disentangle and extract causal substructures from input graphs? The causal substructures are commonly determined by the global properties of the whole graph. It is essential to build relationships between each graph before attempting to extract the causal substructures.

Based on above analysis, a causal disentangled graph neural network (CDGNN) framework for fault diagnosis in complex industrial process is proposed. Based on the causal relationships inherent to GNN-based fault diagnosis processes, this investigation employs causal theory to determine causal and bias variables in graphs, and then disentangle and extract causal substructures from graphs. GNN utilizes only the information of causal substructures for prediction, which mitigates the confounding effect of bias substructures and improves the stability of the model. The main contributions as:

1) We emphasize the stability and robustness issue of GNN-based fault diagnosis task. From the causal perspective, this issue is attributed to the confounding effect of the bias substructures.
2) We introduce a novel CDGNN strategy for fault diagnosis. First, the sensor signals are constructed into graph data through attention mechanism. These graphs are then causally disentangled to extract information from causal subgraphs, including three steps: causal estimation, causal disentanglement and causal aggregation.
3) Experimental results on two complex industrial process datasets show that CDGNN provides superior diagnosis results with better robustness and stability than current state-of-the-art methods.

The rest of the article are organized as follows. Problem formulation and preliminaries are presented in Section II. Section III details the proposed CDGNN. In Section IV, We conducted case studies on two complex industrial datasets. Finally, conclusions and discussion in Section V.

II. PROBLEM FORMULATION AND PRELIMINARIES

A. Problem Formulation

1) Sensor Signal Segments: In industrial systems, numerous sensors are used for system monitoring and process management, these sensor signals form n raw measured variables. For sensor signals over a period of time t, it can be represented as $\boldsymbol{s_i} = (s_i^1, s_i^2, .., s_i^t) \in \boldsymbol{S}$. Signal segments with large spans are usually difficult to handle, so it is necessary to obtain multiple signal segments through window sliding, which can be represented as $\boldsymbol{w_j} = (\boldsymbol{s_i^{t-m+1}}, ..., \boldsymbol{s_i^t}) \in \Omega$. m is the size of sliding window, which is determined according to the stability of sensor signals. These sensor signal segments are served as inputs for constructing graph structure.

2) Input Graph: A graph can be defined as $G = (V, E)$, where V and E are the sets of nodes and edges, respectively. $e_{ij} \in E$ donates an edge between node v_i and v_j. $\mathbf{X} \in \mathbb{R}^{n \times d_n}$ and $\mathbf{X}^e \in \mathbb{R}^{c \times d_c}$ denote nodes and edges attribute, respectively. In this study, the node features $\mathbf{x}_j$ are the attribute of sensor signal segments w_j, and the edge attribute $\mathbf{X}^e$ are learnt through an attention mechanism. The structure information of a graph can be described by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}, n = |V|$. The GNN-based fault diagnosis is a graph classification task: given a set of graphs $\{G_1, G_2, ..., G_N\} \in \mathcal{G}$ to identify fault types $\{y_1, y_2, ..., y_N\} \in Y$.

B. Graph Neural Networks

GNN captures features of graphs by aggregating and updating them among nodes through a message passing mechanism, and iteratively updates the results of message passing in the node representation [4]. The $k-$th message passing as:

$$\begin{aligned} \mathbf{a_i^{(k)}} &= Aggregate^{(k)}\left(\left\{\mathbf{h_j^{(k-1)}} : v_j \in \mathcal{N}(v_i)\right\}\right) \\ \mathbf{h_i^{(k)}} &= Update^{(k)}\left(\mathbf{h_i^{(k-1)}}, \mathbf{a_i^{((k))}}\right) \end{aligned} \tag{1}$$

In the $k-$th message passing, $Aggregate^{(k)}$ generates intermediate result $\mathbf{a_i^{(k)}}$ by aggregating the representations of the set of neighbors $\mathcal{N}(v_i)$ of node v_i , and $Update^{(k)}$ updates the representation $\mathbf{h_i^{(k)}}$ of node v_i by combining $\mathbf{a_i^{(k)}}$ and representation $\mathbf{h_i^{(k-1)}}$ of node v_i in the last round. The initial features of node v_i is $\mathbf{h_i^{(0)}} = \mathbf{x}_i$.

In the graph classification task, the graph representation needs to be obtained first and then GNN classifies labels based on it. A readout function $f_{readout}$ to get the graph representation h_G is defined as:

$$\mathbf{h}_G = f_{\text{readout}}\left(\left\{\mathbf{h}_i^{(k)} \mid v_i \in G\right\}\right). \tag{2}$$

C. Causal Learning

Traditional machine learning primarily focuses on the correlation at the data level, while causal learning attempts to identify causal relationships from data, i.e., how changes in

one event or variable result in changes in another event or variable. According to causal theory [12], data consists of inherent causal information and confounding (i.e., bias) information. Causal learning involves extracting causal information from data and using its causal invariance to make predictions.

Image classification models are susceptible to interference from bias information (e.g., background) that has spurious correlation with target, resulting in poor robustness of model. To solve this problem, [13] recognized the parts composed of causal and bias information as causal and bias substructures, respectively, and then proposed a disentangled features augmentation method to eliminate the spurious correlation between bias substructure and target. [14] analyzed the causal effects in graph classification tasks, and proposed a causal attention learning (CAL) strategy to utilize a backdoor adjustment method to assist model in eliminating spurious correlations, which enhances the capacity of GNN to generalize in the Out-Of-Distribution (OOD) environment.

Causal learning indeed provides a powerful tool for many tasks in machine learning. In image recognition, causal information in images is obvious, consistent and interpretable. However, in fault diagnosis, fault signals, irrelevant sensor signals, and noise signals are highly coupled. Different working conditions of machines introduce data biases, which may result in overfitting of fault diagnosis models and fail to generalize to other machines. To address this challenge, [15] proposed a causal consistency network (CCN) for mining constant causal data in individualised equipment. However, this method can only reduce data bias caused by working conditions, and cannot solve the interference of irrelevant sensor signals.

In fault diagnosis based on sensor signals, the causal information is invariant, i.e., fault signals. Irrelevant sensor signals and noise data will interfere with the model to extract fault features and are regarded as bias information. GNN-based fault diagnosis methods are able to learn the topology and interactions of equipment, but nodes with irrelevant sensor signals as feature inputs will cover fault signals in message passing. Environmental and operational noise can also make the same fault show different symptoms on different equipment. The majority of current causal learning methods are designed for image data and are not applicable to graph data. We identify nodes whose features are fault signals as causal nodes, which form causal substructures, and propose CDGNN to extract causal substructures from graph data.

III. PROPOSED FAULT DIAGNOSIS METHOD

In this section, we analyse causal relationships in the fault diagnosis according to causal theory and develop a structural causal model. To reduce the confounding effect of bias features, we employ deep networks to mine the causal information in signals, thereby facilitating more stable fault diagnosis. To meet the above requirements, CDGNN is proposed, which contains graph construction and causal disentanglement learning. The framework is shown in Fig. 2.

A. Attention Mechanism for Graph Construction

Industrial equipment state data is collected through numerous sensors, and the interaction pairs between sensors form an implicit sensor network. Considering each sensor as a node, this implicit network can be conceptualized as an explicit graph structure. Since the interaction between nodes varies with the type of faults, traditional methods such as correlation or K-nearest neighbour (KNN) are not effective in constructing sensor correlation graphs [5]. To accurately capture the physical topology of industrial systems from the data and differentiate various fault types, we employ an attention mechanism to automatically learn graph structure.

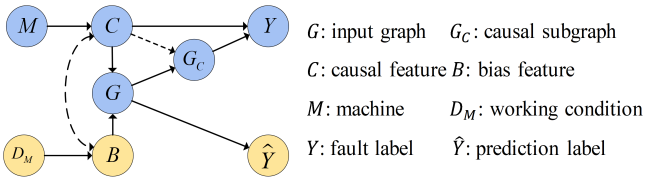

Fig. 1. Structure Causal Model in GNN-based Fault Diagnosis.

The sensor signals are divided into multiple equal-length segments $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})$ as node features. The edge attribute between nodes x_i and x_j is calculated to represent the interaction between them. Specifically, we execute the Pearson correlation coefficients [16] to calculate the correlation coefficient score. For feature matrix $\mathbf{x} = (x_i^1, x_i^2, ..., x_i^m)$, the learned correlation coefficient p_{ij} is defined as:

$$p_{ij} = f(x_i, x_j) \tag{3}$$

where $f(\cdot)$ is Pearson function. A large value of p_{ij} indicates that there is a strong correlation between two nodes and the edges between these nodes should be retained, otherwise they should be removed. We use attention-based dropout layer (ADL) [17] to filter the correlation coefficients and eliminate redundant interactions to obtain the input graph.

$$\begin{aligned} \mathbf{e} &= ADL(\mathbf{p}) \\ p_{ij} &= max\{0, p_{ij} - \tau(p_{ij})\} \end{aligned} \tag{4}$$

where $\tau(\cdot)$ is the threshold used to distinguish the strength of the correlation and all p_{ij} are below the threshold $\tau(p_{ij})$ will be truncated to 0. The correlation scores $\mathbf{e} \in \mathbb{R}^n$ are used as edge weights $\mathbf{X}^e$ and are involved in graph message passing.

The attention mechanism is employed to construct graphs, which can adaptively learn its structure and thus more accurately distinguish the graph structure of different faults. A larger correlation score between two nodes indicates a stronger physical association between the corresponding sensors.

B. Causal View on GNN-based Fault Diagnosis

We excavate the causality inherent to data generation and prediction process in GNN-based fault diagnosis to understand inherent mechanism of graph classification. To this end, we examine the causal relationships between the eight variables and construct a structural causal model (SCM) [9]. The SCM in GNN-based fault diagnosis is presented as Fig. 1, where each link represents a causal relationship.

• $\boldsymbol{C} \rightarrow \boldsymbol{G} \leftarrow \boldsymbol{B}$: using sensor signals for fault diagnosis, where causal feature $\boldsymbol{C}$ are fault signals of machine $\boldsymbol{M}$, and

bias feature $\boldsymbol{B}$ are irrelevant sensor signals and noise data determined by the working condition $\boldsymbol{D_M}$.

• $\boldsymbol{G} \rightarrow \boldsymbol{G_c} \rightarrow \boldsymbol{Y}$, and $\boldsymbol{C} \rightarrow \boldsymbol{G_c} \rightarrow \boldsymbol{Y}$: fault label $\boldsymbol{Y}$ is determined by causal subgraph $\boldsymbol{G_c}$, which is composed of $\boldsymbol{C}$.

• $\boldsymbol{C} \rightarrow \boldsymbol{Y}$: causal feature $\boldsymbol{C}$ is the sole determinant of fault label $\boldsymbol{Y}$. For instance, if $\boldsymbol{C}$ represents sensor signal generated when a fault occurs, then $\boldsymbol{Y}$ is the corresponding fault type.

• $\boldsymbol{B} \rightarrow \boldsymbol{G} \rightarrow \hat{\boldsymbol{Y}}$: bias feature $\boldsymbol{B}$ interferes with the prediction, resulting in the predicted label $\hat{\boldsymbol{Y}}$ is obtained from the graph $\boldsymbol{G}$ to differ from true fault label $\boldsymbol{Y}$.

• $\boldsymbol{C} \dashleftarrow\dashrightarrow \boldsymbol{B}$: fault signals (i.e., causal feature) and irrelevant sensor signals (i.e., bias feature) and noise are captured simultaneously. These signals are highly coupled and have spurious correlations that cannot be directly observed.

C. Inverse Probability Weighing

Compared to the causal substructure, the bias substructure is simpler and GNN can utilize the bias substructure to rapidly reach low loss. Therefore, GNN tends to utilize bias substructure when graphs exhibit bias [18]. However, the prediction via bias substructures is unstable. According to the SCM, we can identify two backdoor paths that may lead to spurious correlations between $\boldsymbol{B}$ and $\boldsymbol{Y}$: (1) $\boldsymbol{B} \rightarrow \boldsymbol{G} \rightarrow \boldsymbol{Y}$ (2) $\boldsymbol{B} \dashleftarrow\dashrightarrow \boldsymbol{C} \rightarrow \boldsymbol{Y}$. Confounding effects caused by bias feature will reduce the stability of predictions through these spurious correlations. Therefore, these two spurious paths must be blocked, and we propose to debias GNN from a causal view.

• $\boldsymbol{C} \leftarrow \boldsymbol{G} \rightarrow \boldsymbol{B}$ and $\boldsymbol{C} \rightarrow \boldsymbol{Y}$: to block path (1), it is necessary to disentangle $\boldsymbol{C}$ and $\boldsymbol{B}$ from $\boldsymbol{G}$, with all subsequent predictions being based solely on $\boldsymbol{C}$.

• $\boldsymbol{C} \dashleftarrow\times\dashrightarrow \boldsymbol{B}$: explicitly blocking path (2) is challenging. The path from $\boldsymbol{C}$ to $\boldsymbol{Y}$ is unchanging, but we can make $\boldsymbol{C}$ and $\boldsymbol{B}$ uncorrelated, achieving causal disentanglement.

In complex industrial fault diagnosis, causal and bias features cannot be directly observed, making it difficult to explicitly disentangle $\boldsymbol{C}$ and $\boldsymbol{B}$. Fortunately, causal theory provides a possible solution [19]: we can get rid of the backdoor path by using *do*-calculus on the causal variable $\boldsymbol{C}$ to estimate $P_m(Y|C) = P(Y|do(C))$. Backdoor adjustment as a causal intervention tool is theoretically sufficient for estimating causal relationships [14] between $\boldsymbol{C}$ and $\boldsymbol{B}$, but it involves summing probabilities over $\boldsymbol{B}$. If the number of possible values of $\boldsymbol{B}$ is very large, the summation will be very computationally intensive. To address this issue, we employ the inverse probability weighing (IPW) [20] approach to reweight graph data, which avoids the problem of probability summation over $\boldsymbol{B}$. According to Bayes rules, the IPW can be represented as:

$$\begin{aligned} P(Y|do(C)) &= P(Y|C) \\ &= \sum \frac{P(Y|C,B)P(B)P(C|B)}{P(C|B)} \quad (Basyes\ \ Rule) \\ &= \sum \frac{P(Y|C,B)}{P(C,B)} \quad (IWP) \end{aligned} \tag{5}$$

$P(Y|(C,B))$ denotes the given conditional probability of $\boldsymbol{C}$ and $\boldsymbol{B}$, and $P(B)$ donates the prior probability of $\boldsymbol{B}$.

D. Causal Disentanglement Learning

To eliminate spurious correlations, we propose causal disentanglement learning strategy. Firstly, a mask generator is designed to mask edges of the input graph, thereby obtaining two distinct subgraphs: a causal subgraph and a bias subgraph. Subsequently, two independent GNN modules are trained for encoding these subgraphs into their respective representations, which are then disentangled. Finally, these representations are trained to extract causal features, eliminating the spurious correlations. The overall process is shown in Fig. 3.

1) Causal Estimate: Given a graph $G = (\mathbf{A},\mathbf{X})$, where $\mathbf{A}$ is adjacency matrix, and $\mathbf{X}$ is node features matrix. We estimate the importance of $edge(i,j)$ in the causal subgraphs using a multi-layer perception (MLP) connecting feature matrix x_i of node v_i and feature matrix x_j of node v_j.

$$\begin{aligned} \alpha_{ij} &= MLP([x_i, x_j]) \\ c_{ij} &= ReLU(\alpha_{ij}) \in (0,1) \end{aligned} \tag{6}$$

α_{ij} is the importance of $edge(i,j)$, and c_{ij} indicates the probability that $edge(i,j)$ belongs to a causal subgraph.

Naturally, the probability of $edge(i,j)$ belongs to a bias subgraph is: $b_{ij} = 1 - c_{ij}$. The causal and bias masks can be denoted as $\boldsymbol{M_c} = [c_{ij}]$ and $\boldsymbol{M_b} = [b_{ij}]$, respectively. Finally, we disentangle G to obtain the causal subgraph $G_c = \{\boldsymbol{M_c} \odot \boldsymbol{A}, \boldsymbol{X}\}$ and bias subgraph $G_b = \{\boldsymbol{M_b} \odot \boldsymbol{A}, \boldsymbol{X}\}$. In summary, mask allows GNN to distinguish different structural information of graphs, enabling GNNs trained on different subgraphs can extract features from different parts of graphs.

2) Causal Disentanglement: How to ensure G_c and G_b are causal and bias subgraph, respectively? Uncertainty properties of bias graphs have diverse representations and are easier to extract relevant features [13]. Inspired by this, a pair of GNNs (g_c, g_b) with linear classifiers (C_c, C_b) are simultaneously trained in our approach. (1) Based on the fact that bias substructures are easily to grasp [11], we employ a bias-aware loss function to train a bias GNN model (g_b, C_b); (2) Correspondingly, a causal GNN model (g_c, C_c) on graphs that are difficult to learn for the bias GNN.

As illustrated in Fig. 3, causal and bias subgraphs are embedded into causal and bias representations $z_c = g_c(G_c;\gamma_c)$ and $z_b = g_b(G_b;\gamma_b)$ by the GNN g_c and g_b, respectively. γ are parameters of GNNs. The classifiers C_c and C_b leverage representation vectors z_c and z_b to predict the target label y, respectively. To train g_b and C_b to extract bias features, we employ generalized cross-entropy (GCE) [21] to enhance the bias of bias GNN and classifier.

$$GCE(C_b(z_b;\alpha_b), y) = \frac{1 - C_b^y(z_b;\alpha_b)^q}{q} \quad q \in (0,1] \tag{7}$$

where $C_b(z_b;\alpha_b)$ and $C_b^y(z_b;\alpha_b)$ represent the $softmax$ output of bias classifier and the probability that belongs to y, respectively. α are parameters of C_b. The hyperparameter q is used to regulate degree of bias.

Bias features are normally easier to learn, so C_b^y of the biased graph is higher than unbiased graph. Thus, the models g_b and C_b trained with GCE can pay attention to bias features and obtain bias subgraphs. For high confidence samples, the standard cross entropy (CE) [22] will yields lower loss. Thus,

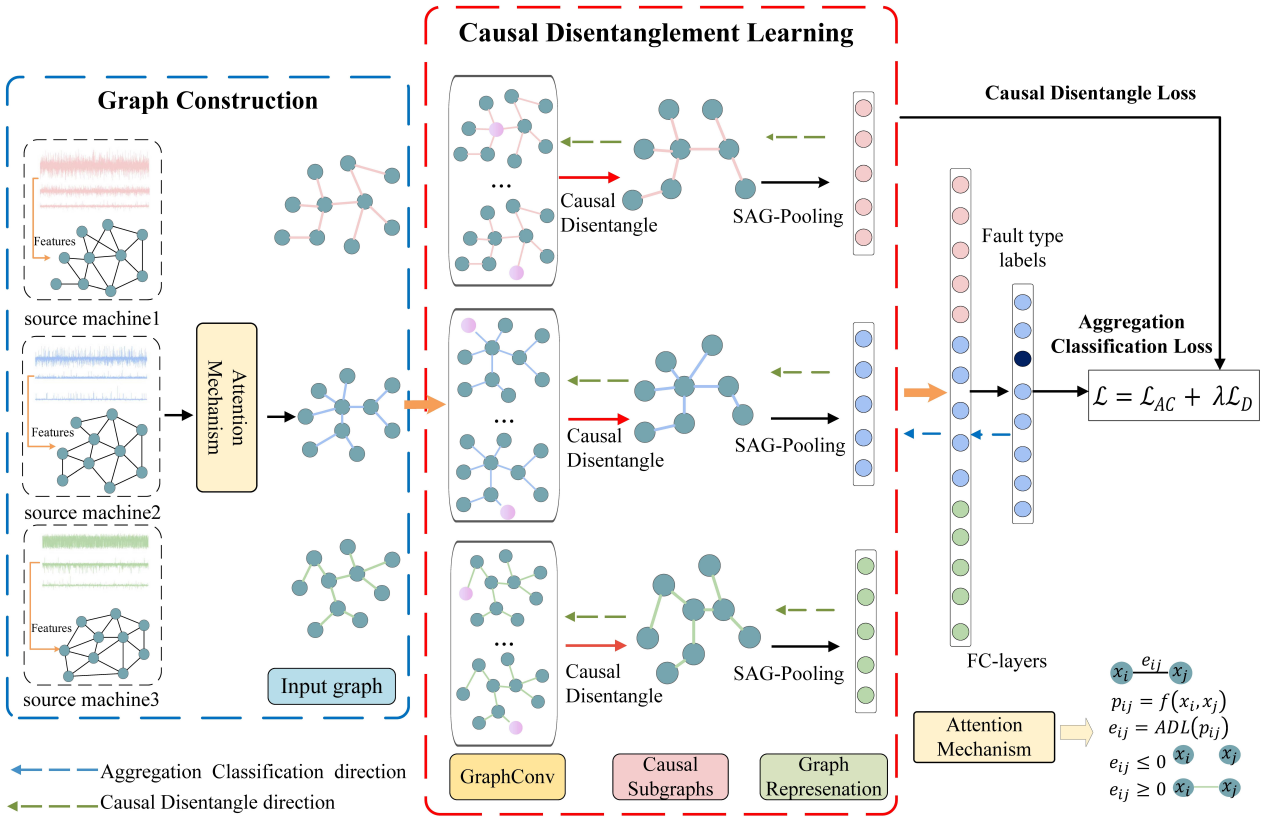

Fig. 2. The framework of proposed Causal Disentangled Graph Neural Network (CDGNN).

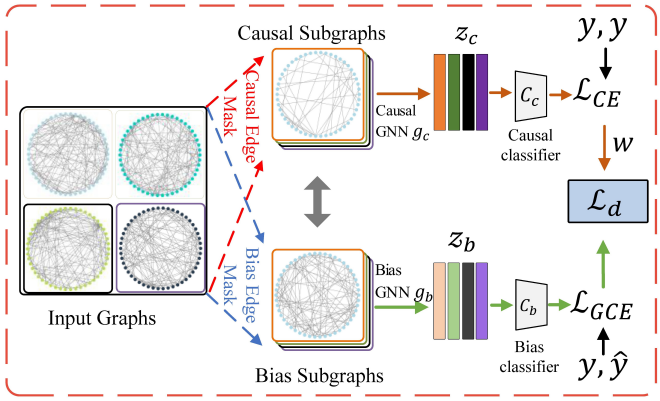

Fig. 3. Causal Disentanglement Learning in GNN.

we train g_c and C_c as causal extractor by CE loss. In this way, we can get the causality scores for each graph:

$$W(z) = \frac{CE(C_b(z_b), y)}{CE(C_c(z_c), y) + CE(C_b(z_b), y)} \tag{8}$$

A large value of W indicates that the subgraph is a causal subgraph. Based on the above analysis, we reweight the input graphs to achieve inverse probability weighting on them.

$$\mathcal{L}_d = W(z)CE(C_c(z), y) + GCE(C_b(z), y) \tag{9}$$

where Eq. 9 is defined as causal disentanglement learning loss. It reweights the training losses to train g_c and C_c towards learning causal information and push causal subgraphs to be stable and invariant in prediction.

Algorithm 1 CDGNN

Input: Dataset $\mathcal{G}$, GNN modules (g_c, g_b) and parameters (γ_c, γ_b), classifiers (C_c, C_b), hyper-parameter λ
Output: The trained parameters

for sampled K graphs $\{G^k = (A^k, X^k)\}_{k=1}^K$ **do**
  compute the importance α_{ij} and probability of causal subgraph $c_{ij} \leftarrow$ Eq. 6 for all edges of G^k
  get causal edge mask M_c^k based on c_{ij}
  get causal subgraph $G_c^k = \{M_c^k \odot A^k, X^k\}$
  get causal embedding $z_c^k = g_c\left(G_c^k; \gamma_c\right)$
  get non-causal embedding $z_b^k = g_b\left(G_b^k; \gamma_b\right)$
  train bias extractor $GCE(C_b(z^k;\alpha_b), y^k) \leftarrow$ Eq. 7
  train causal extractor $CE(C_c(z^k), y^k)$
  compute causality score $W(z^k) \leftarrow$ Eq. 8
  causal disentanglement loss $\mathcal{L}_d^k \leftarrow$ Eq. 9
  total causal disentanglement loss $\mathcal{L}_D = \sum_{k=1}^K \mathcal{L}_d^k$
  get final graph representation $R \leftarrow$ Eq. 10-12
  aggregation classification loss $\mathcal{L}_{AC} \leftarrow$ Eq. 13
  total loss $\mathcal{L} = \mathcal{L}_{AC} + \lambda\mathcal{L}_D$
end for

E. Aggregation Classification Training

Sensors are susceptible to noise data, which varies depending on the working conditions. Thus, even the same fault behaves differently on different equipment. To mitigate

the interference of noise data, we aggregate causal subgraph representations from different working conditions into a final global graph representation. First, we employ SAG-Pooling [23] as a readout function $f_{readout}$ to obtain representation of the $k-$th causal subgraph G_c^k.

$$r_k = f_{readout}(G_c^k) \tag{10}$$

There are differences in causal subgraph representations from different working conditions $d \in D_M$, so we propose a weighting function to computes the weights β_k of subgraphs.

$$\begin{aligned} w_k &= \sum_{k \in K} \overrightarrow{\mathbf{p}} \cdot \mathrm{ReLu}\left(f_a\left(r_k\right)\right) \\ \beta_k &= \frac{\exp(w_k)}{\sum_{k \in K} \exp(w_k)} \end{aligned} \tag{11}$$

where w_k is the importance coefficient of representation are measured by an attention vector $\overrightarrow{\mathbf{p}}$. β_k is the normalized importance coefficient obtained from the $softmax$ function.

Aggregating subgraph representations under different working conditions yields a final global graph representation $\mathbf{R}$.

$$\mathbf{R} = \sum_{k=1}^{K} r_k \beta_k \tag{12}$$

The GNN-based fault diagnosis on multiple sensor signals is a graph classification task. We use multiple fully connected (FC) layers to process the graph feature information extracted by CDGNN, and output the probabilities of different fault categories using a $softmax$ classifier in the last layer. Therefore, the objective function for aggregation classification can be written as:

$$\mathcal{L}_{AC} = -\sum_{j=1}^{N} y_j \; log \; q_j(w|g) \tag{13}$$

where w are parameters of the model and q_j represents the probability that the predicted target belongs to fault type j. Thus, the total loss of CDGNN can be described as the sum of losses:

$$\mathcal{L} = \mathcal{L}_{AC} + \lambda \mathcal{L}_D \tag{14}$$

The hyper-parameter λ determines the strength of causal disentanglement learning, $\mathcal{L}_D = \sum_{k=1}^{K} \mathcal{L}_d^k$. A comprehensive explication of the implementation of CDGNN in Algorithm 1.

IV. CASE STUDY

This section presents an evaluation of the effectiveness of CDGNN in addressing fault diagnosis in complex industrial process. The experimental results and subsequent analysis demonstrate the superiority and efficacy of CDGNN in this domain compared to current state-of-the-art methods.

A. Data Description

1) Three-phase flow facility (TFF): **TFF** [24] is a simulated pressurized system to provide controllable and measurable flows of air, water and oil, refer as Fig. 4. The system comprises 24 sensors, which are used to measure temperature, pressure, and flow at various points within system. The sensors

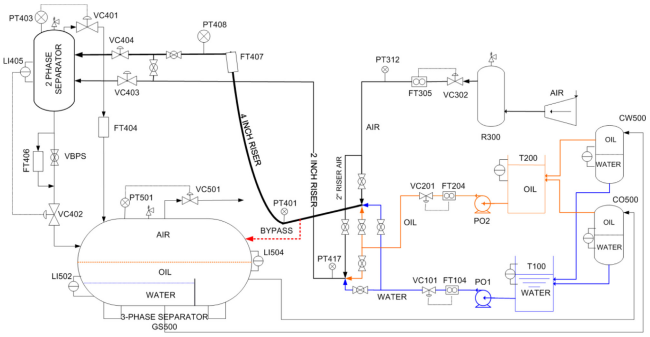

Fig. 4. Sketch of the three-phase flow facility.

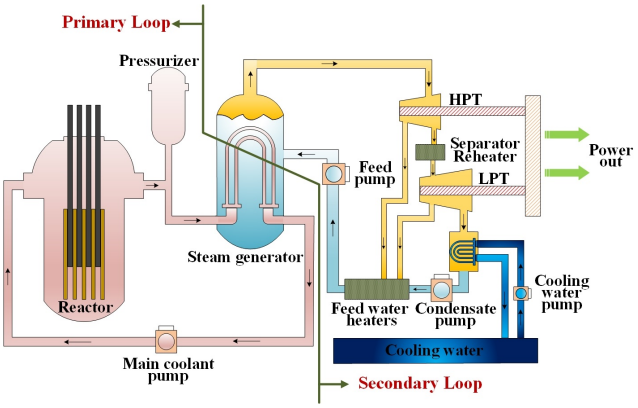

Fig. 5. Simplified schematic of nuclear power system

are configured to sample data at a frequency of 1 Hz. TFF simulates three datasets under conditions of normal and six faults that may potentially arise during the real-world deployment. It should be pointed out that the 24th variable is used solely for fault case 6. The dataset is available at TFF DATA. To evaluate the impact of faults on health indicators, we introduce faults into the simulation following a period of standard operation. When a system fault escalates to a predefined threshold, the fault condition is automatically resolved, and TFF system returns to normal.

2) Nuclear power system (NPS): **NPS** [10] is the nuclear power system data collected by a large number of sensors. It constitutes power plant generation modules, including pumps, steam generators, turbines and other related components, refer to Fig. 5. This system has 121 sensors to monitor and record component pressures, temperatures, and valve operations etc. The sampling frequency is 4 Hz. Each simulation of the NPS starts from a standard operating state and then simulates a fault during operation, thus creating a dataset with alternating normal and fault states. Consistent with real practice monitoring, where numerous sensors are often positioned in same place, leading to the overlap of data, we strategically select 64 monitored variables to eliminate redundancy.

To extract fault data more efficiently, we slice the sensor signals based on sampling frequency and data complexity. Specifically, the TFF and NPS dataset are sliced into segments of 50 and 80 data points, respectively. We preprocess two datasets with max-min normalization and data split is detailed in Table I.

TABLE I
SAMPLES OF THE DATASETS

Dataset	Fault types	Training set	Testing set
TFF	Fault type(1-6)	874	379
	Normal	478	190
NPS	Fault type(1-53)	8084	3751
	Normal	2698	1175

TABLE II
ACCURACY AND F1-SCORE (%) COMPARISON BETWEEN DIFFERENT FAULT DIAGNOSIS METHODS

Method \ Dataset	TFF		NPS	
	Accuracy	F1-score	Accuracy	F1-score
PCA+LDA	73.73	72.11	72.78	73.04
WPD-MSCNN	76.70	75.89	75.53	76.42
GCN	83.68	81.35	71.39	74.31
IAGNN	92.29	91.67	87.34	87.67
CCN	91.47	91.11	85.13	85.81
CAL	93.35	93.04	86.88	87.03
CDGNN	**95.23**	**94.92**	**89.48**	**89.91**

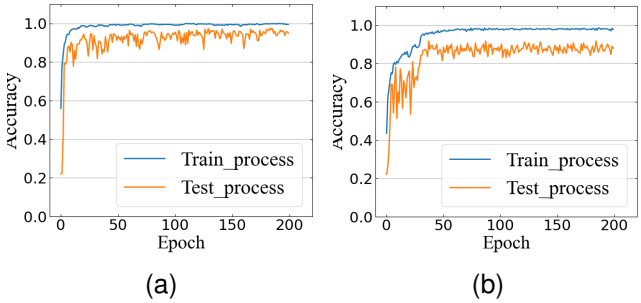

Fig. 6. Convergence of CDGNN on (a) TFF dataset. (b) NPS dataset.

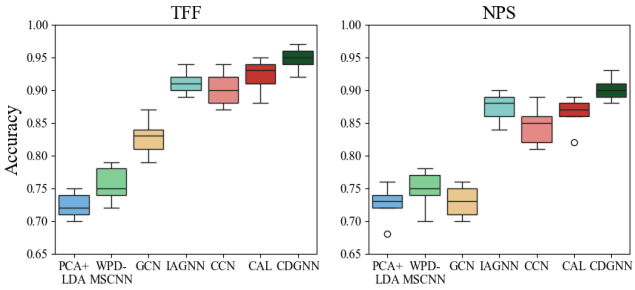

Fig. 7. Box plots for different methods on TFF and NPS.

B. Experimental Setup

1) Baseline methods: Comparison methods include traditional methods, deep learning-based methods, GNN-based methods and causal learning methods for evaluating the superiority and effectiveness of CDGNN.

1) PCA+LDA [25]: A common fault diagnosis method based on data dimension reduction.
2) WPD-MSCNN [26]: A multi-feature scale CNN based on wavelet packet decomposition to adaptively acquire components at multiple feature scales, which is robust in dealing with the scale uncertainty of vibration signals.
3) GCN [27]: A method for extracting features from graph data for processing data from non-Euclidean space.
4) IAGNN [6]: It designs an interaction-aware module for extracting sensor interaction relationships and creating feature subgraphs. Connect features of multiple subgraphs by employing a weighted sum function.
5) CCN [15]: It designs a causal consistency loss that describes the causal consistency of fault signals, mining invariant causal information in in sensor signals.
6) CAL [14]: It designs an attention module to distinguish between causal and shortcut features of the input graphs, and employs backdoor adjustment to reduce confounding effects caused by shortcut features.

2) Implementation details and evaluation metrics: To identify the optimal hyperparameters for achieving the best results, we conduct extensive experiments on CDGNN and baseline methods. Learning rate of CDGNN is chosen in $\{0.0005, 0.001, 0.005\}$. When applied to TFF, the input graph consists of 24 nodes, corresponding to 24 sensors and the initial feature size of the nodes is 50. As for NPS, there are 64 nodes with the initial features size of 80. These experiments are carried out on a machine with the Windows 11 OS, with 16G RAM, an Intel(R) Xeon(R) E-2224 CPU and a NVIDIA RTX 3090 GPU. We analyze the performance of CDGNN in fault diagnosis task using metrics such as accuracy, F1-score, and confusion matrix and compared it with baseline methods.

C. Fault Classification Performance

In practice, mechanical equipment often ceases to function before a fault can be identified, resulting in a paucity of fault data. According to this principle, extensive experiments have been conducted, and results are presented in Table II. Fig. 6 illustrates the accuracy trajectory of CDGNN over the first 200 rounds of training and testing process. Despite fluctuations in training and testing curves due to imbalanced sampling, they eventually converge on stable values, indicating that CDGNN is less susceptible to overfitting.

1) TFF Experiment Results Analysis: Obviously, CDGNN achieves optimal results on TFF compared with baseline methods. This indicates that the fault features of sensor signals learned by CDGNN are effective in revealing the nature of faults in complex industrial process, which can be attributed to three reasons: 1) CDGNN constructs graphs using attention mechanism, which is able to reasonably represent the interactions of the multiple sensor measurements. 2) CDGNN causally disentangle input graphs, which weakens the confounding effects of noise and irrelevant sensor signals. The prediction based on causal subgraphs significantly improve the accuracy of fault diagnosis. 3) CDGNN aggregates causal subgraphs from various working conditions to obtain the final complete graph representation, eliminating the data bias caused by individualized machines and improving its generalization ability.

It can be observed that GCN outperforms PCA+LDA and WPD-MSCNN, indicating that it is essential to take into ac-

count the interactions between multiple sensors and correlation between fault features. IAGNN designs an interaction-aware module that learns the complex interactions between different components and is able to fuse information from multiple sensor signals more efficiently. However, IAGNN is unable to eliminate interference caused by irrelevant sensor signals and noise data, and therefore cannot accurately extract fault features, which are critical for fault diagnosis.

CCN reduces the data bias generated by changes in operating conditions by mining the invariant causal information in fault data. However, CCN does not take into account the interactions between multiple sensors, rendering it incapable of identifying the propagation of fault signals. Therefore, CCN has limited effectiveness in fault diagnosis on complex industrial datasets. CAL analyzes the causal effects in graph classification tasks, which utilizes a backdoor adjustment method to assist model in eliminating spurious correlations. Notice that CAL is built on image recognition task, where causal information in images is transparent, consistent and easy to extract. However, in fault diagnosis, fault signals, irrelevant sensor signals, and noise data are highly coupled. CAL does not take into account checking the correctness of causal disentanglement and the interference of noise data.

2) NPS Experiment Results Analysis: To verify the effectiveness of CDGNN on large-scale imbalance data, we applied it to the NPS data, which contains 53 fault categories with large sample differences between different fault categories. According to the results in Table II, CDGNN again achieves superior performance on NPS. We further compare the fault diagnosis results of CDGNN with baseline methods using confusion matrices as shown in Fig. 8, where darker colors indicate more accurate fault diagnosis. It can be observed that CDGNN demonstrates commendable performance across various faults. The competitive results on a large-scale imbalance dataset further validate the effectiveness of CDGNN. Additionally, PCA+LDA and WPD-MSCNN perform better than GCN when applied to NPS because NPS has much more sensor signal sequences than TFF. Therefore, the complexity of the structure is high when constructing them to graphs, which prevents GCN from effectively learning information about neighbouring nodes.

CCN can only mine fault information in each sensor signal sequence, when a fault causes multiple sensor signal readings to be abnormal, CCN is unable to identify the correlation relationships between them. IAGNN, CAL, and CDGNN are all GNN-based methods, where fault information and correlation relationships are represented by node features and edge attributes, respectively, and the effective information is learnt in message passing. It is worth to note that CAL performed worse than IAGNN on NPS data, in contrast to their performance on TFF data. The more complex an industrial system is, the more important it is for fault diagnosis to take into account the topology and interactions between equipment. NPS contains 64 sensors, so constructing a reasonable graph structure is crucial for distinguishing faults. Obviously, the graph structure of NPS data is much more complex than that of TFF data. CAL can only disentangle obvious causal information in images, but it cannot ensure correct causal disentanglement when dealing with complex sensor signals.

CDGNN employs a causal disentanglement learning module to extract key substructures in large scale graphs and thus learn effective information for fault prediction. This suggests that CDGNN with causal learning capability can handle unbalanced data more effectively. Combined with the box plots in Fig. 7, it is easy to find that CDGNN is more effective and stable in fault diagnosis of complex industrial process.

D. Analysis on Causal Disentanglement Learning

CCN, CAL and CDGNN are all classification models based on causal learning. To visually analyse the causal disentanglement capability of CDGNN, the features of TFF data are visualized using t-distributed stochastic neighbourhood embedding (t-SNE) [28], including raw features, features from the last layer of CCN and CAL, and features of causal subgraphs extracted by CDGNN. As shown in Fig. 9(a), the features of TFF fault samples are distributed depending on the distinct machine statuses and working conditions because the sample features contain noise and data biases. As a result, even for the same fault, the feature distribution has a large deviation. It can be seen that the feature space distributions of the raw data of the faults are very close to each other except for Fault 1. Because each fault causes only a few sensor readings to deviate from normal, a large number of irrelevant sensor signals with time-series features dominate the feature distribution, making different faults highly similar in the feature space.

According to existing studies [6] and experimental analysis in Section IV-C, mining interrelationships among multiple sensor measurements and fusing the information is crucial for fault diagnosis of complex industrial process. CCN aims to mine the invariant causal information in machines and resolve the data bias caused by different machines. However, CCN can only process each sensor signal individually, and cannot guarantee the reliability of the extracted signals without considering the influence of other sensors. According to the visualisation results in Fig. 9(b) and Fig. 9(c), the feature clustering performance extracted by CCN is mediocre and CAL is better than CCN. The clustering performance of the causal subgraph features extracted by CDGNN is excellent, as shown in Fig. 9(d). There are three reasons: 1) CDGNN uses the attention mechanism to construct the graph, which takes into account the interactions between multi-sensor measurements. 2) CAL cannot ensure the credibility of the causal features extracted from sensor signals, whereas CDGNN trains two independent GNN modules using a causal/bias-aware loss function to ensure the correctness of causal disentanglement. 3) CDGNN aggregates causal subgraphs from different machines to obtain a complete graph representation to learn the bias information caused by noise and working conditions.

E. Ablation Study

To assess the efficacy and contribution of individual modules within CDGNN, ablation studies were conducted. It is clear that each module plays an important part in the overall performance of CDGNN, as shown in Table III. The employment of an attention mechanism for the construction of graphs

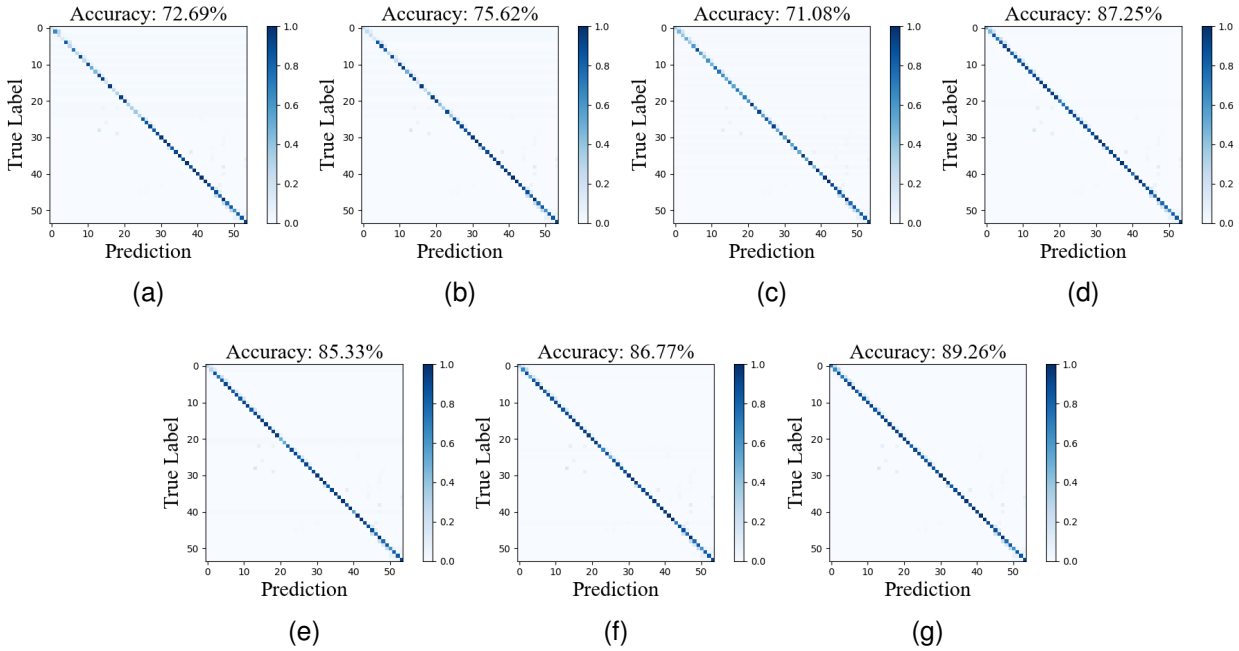

Fig. 8. Confusion matrices of the NPS. (a) PCA+LDA, (b)WPD-MSCNN, (c)GCN, (d)IAGNN, (e)CCN, (f)CAL, and (g)CDGNN

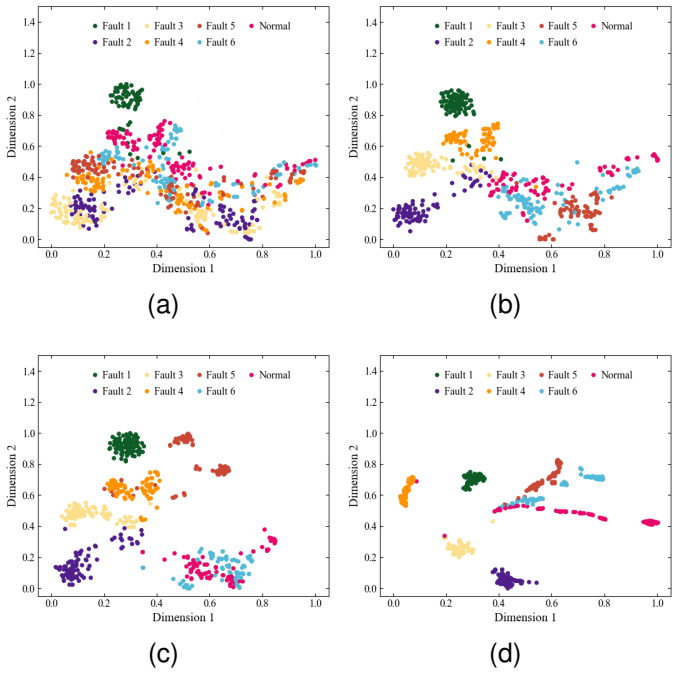

Fig. 9. Feature visualization via t-SNE of TFF. (a) Raw data space, (b) CCN learning space, (c) CAL learning space, (d) Causal subgraph features of CDGNN learning space.

TABLE III
ABLATION STUDY RESULTS

Attention	Disentanglement	Aggregation	TFF	NPS
—	—	—	85.72	72.37
✓	—	—	87.33	75.28
—	✓	✓	92.37	85.11
✓	✓	—	93.25	85.68
✓	—	✓	89.26	79.97
✓	✓	✓	**95.23**	**89.48**

1) Attention: attention mechanism for graph construction, 2) Disentanglement: causal disentanglement learning, 3) Aggregation: subgraph features aggregation.

proves to be beneficial for the GNN to learn the topology in complex industrial systems. By aggregating causal subgraphs from multiple sources, CDGNN effectively eliminates data bias generated by machines in different working conditions, thus improving its generalization. Obviously, the causal disentanglement learning module exerts the most significant influence on fault diagnosis performance, underscoring its pivotal role. It disentangles the input graph from the perspective of causality to extract causal subgraphs. The prediction based on causal subgraphs indeed mitigates the confounding effect of irrelevant sensor signals and noise. After removing these three modules, CDGNN reverts to a classic GNN model, exhibiting a classification performance analogous to that of GCN. The above analysis proves the rationality and necessity of CDGNN.

F. Hyper-parameter and Computational Cost

The sensitivity of λ is illustrated as shown in Fig. 10. For λ, a specific range is identified that optimizes model performance over two datasets, and CDGNN achieves the best accuracy when λ is 0.5. It is notable that NPS data exhibits greater sensitivity to variations of λ may due to the NPS dataset is more unbalanced and complex.

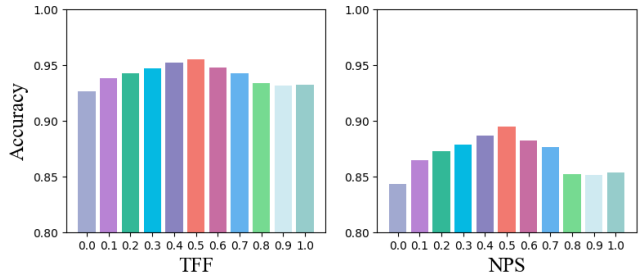

Fig. 10. Accuracy comparison under different λ.

TABLE IV
COMPUTATIONAL TIME COMPARISON (EPOCH/SECONDS)

Method \ Dataset	TFF		NPS	
	Training	Testing	Training	Testing
PCA+LDA	19.03	5.76	322.34	76.45
WPD-MSCNN	14.78	3.65	170.14	57.33
GCN	**1.76**	**0.33**	**17.04**	**4.56**
IAGNN	5.87	1.25	63.85	18.77
CCN	10.43	2.89	152.87	39.11
CAL	3.63	0.65	43.76	8.92
CDGNN	4.71	0.97	51.76	14.33

Computational time is an important metric to evaluate the efficiency of models. To this end, we measured the computational costs of CDGNN and baseline methods, as shown in Table IV. GCN has a simple structure and naturally requires the least computation time, but it offers poor fault diagnosis capabilities. Although CAL requires less computation time than CDGNN, it performs mediocrely on NPS with higher complexity. In contrast, CDGNN achieves excellent fault diagnosis results with only a minor increase in computational load. CDGNN has the lowest computation time compared to the other baseline methods, which proves that it can provide fault diagnosis results faster.

V. CONCLUSION

In this study, we excavate the causal relationships inherent to GNN-based fault diagnosis according to the causal theory and introduce the novel causal disentangled graph neural network (CDGNN). According to causal theory, irrelevant sensor signals and noise data are regarded as bias variables, which exert confounding effect in fault diagnosis and mislead GNN to learn spurious correlations. CDGNN solves the problem by causally disentangling input graph into causal/bias subgraphs and aggregating causal subgraphs. Experimental results on the TFF and NPS datasets indicate that predictions obtained from causal subgraphs are closer to real results. CDGNN has better stability and generalization ability by disentangling causal and bias features, and can be more effectively applied to complex industrial process. This study is conducted in the closed-world assumption that all faults in testing data have already appeared in training data. Further research will focus on applying CDGNN to open-set recognition, distinguishing and identifying unknown faults based on causal features, and detecting hidden unknown faults in time.

REFERENCES

[1] R. Alfredo Osornio-Rios, J. A. Antonino-Daviu, and R. de Jesus Romero-Troncoso, "Recent industrial applications of infrared thermography: A review," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 615–625, 2019.
[2] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020.
[3] P. Gangsar and R. Tiwari, "Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review," *Mechanical Systems and Signal Processing*, vol. 144, p. 106908, 2020.
[4] Z. Wu, S. Pan, F. Chen, G. Long, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.
[5] T. Li, Z. Zhou, S. Li, C. Sun, R. Yan, and X. Chen, "The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study," *Mechanical Systems and Signal Processing*, vol. 168, p. 108653, 2022.
[6] D. Chen, R. Liu, Q. Hu, and S. X. Ding, "Interaction-aware graph neural networks for fault diagnosis of complex industrial processes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 6015–6028, 2023.
[7] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, "On explainability of graph neural networks via subgraph explorations," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139, 2021, pp. 12 241–12 252.
[8] D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, "Parameterized explainer for graph neural network," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020, pp. 19 620–19 631.
[9] J. Pearl, M. Glymour, and N. P. Jewell, "Causal inference in statistics: A primer," *John Wiley & Sons*, 2016.
[10] Y. Wang, R. Liu, D. Lin, D. Chen, P. Li, Q. Hu, and C. L. P. Chen, "Coarse-to-fine: Progressive knowledge transfer-based multitask convolutional neural network for intelligent large-scale fault diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 2, pp. 761–774, 2023.
[11] S. Fan, X. Wang, Y. Mo, C. Shi, and J. Tang, "Debiasing graph neural networks via learning disentangled causal substructure," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2024, pp. 24 934–24 946.
[12] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang, "A survey on causal inference," *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 5, pp. 1–46, 2021.
[13] J. Lee, E. Kim, J. Lee, and J. Lee, "Learning debiased representation via disentangled feature augmentation," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 25 123–25 133.
[14] Y. Sui, X. Wang, J. Wu, M. Lin, X. He, and T.-S. Chua, "Causal attention for interpretable and generalizable graph classification," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, p. 1696–1705.
[15] J. Li, Y. Wang, Y. Zi, H. Zhang, and C. Li, "Causal consistency network: A collaborative multimachine generalization method for bearing fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5915–5924, 2023.
[16] P. Schober, C. Boer, and L. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, pp. 1763–1768, 2018.
[17] J. Choe, S. Lee, and H. Shim, "Attention-based dropout layer for weakly supervised single object localization and semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4256–4271, 2021.
[18] J. Pearl, "Interpretation and identification of causal mediation," *ERN: Other Econometrics: Econometric Model Construction*, 2013.
[19] J. Pearl and H. White, "Causality," *Cambridge University Press*, 2009.
[20] P. R. ROSENBAUM and D. B. RUBIN, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.
[21] Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, p. 8792–8802.
[22] A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: theoretical analysis and applications," in *Proceedings of the 40th International Conference on Machine Learning*, 2023, pp. 23 803–23 828.

[23] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, 2019, pp. 3734–3743.
[24] C. Ruiz-Cárcel, Y. Cao, D. Mba, L. Lao, and R. Samuel, "Statistical process monitoring of a multiphase flow facility," *Control Engineering Practice*, vol. 42, pp. 74–88, 2015.
[25] P. Belhumeur and J. Hespanha, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
[26] D. Huang, W.-A. Zhang, and F. Guo, "Wavelet packet decomposition-based multiscale cnn for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 443–453, 2023.
[27] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations*, 2017, pp. 1–14.
[28] L. van der Maaten and G. E. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

Ruonan Liu (Senior Member, IEEE) received the B.S., M.S. and PhD degrees from Xi'an Jiaotong University, Xi'an, China, in 2013, 2015 and 2019, respectively. She was a postdoctoral researcher with the School of Computer Science, Carnegie Mellon University in 2019. She currently is an associate professor in the Department of Automation, Shanghai Jiao Tong University. And she is also an Alexander von Humboldt Fellow with the University of Duisburg-Essen, Germany.

She has been awarded the 2021 Outstanding Paper Award by IEEE Transactions on Industrial Informatics, Best Paper Award Finalist by IEEE ICARM 2024, Runner-up Paper Award of GLOW workshop in IJCAI 2024. She has been recognized as one of the World's Top 2% Scientists by Stanford University consecutively from 2021 to now and selected in the Young Elite Scientist Sponsorship Program by CAST in 2022. Now she serves as the associate editor or leading guest editor for IEEE Transactions on Industrial Cyber-Physical Systems, Sustainable Energy Technologies and Assessments and so on. Her research interests include machine learning, intelligent manufacturing and intelligent unmanned system.

Quanhu Zhang received his B.S. degree in computer science and technology from Nanjing Forestry University, Nanjing, China, in 2022. He is currently pursuing an M.S. degree in computer technology at the College of intelligence and computing, Tianjin University, Tianjin, China. His research interests include deep learning, graph neural networks, time series modeling, and industrial cyber-physical systems.

Di Lin (M'17) received the bachelor degree in software engineering from Sun Yat-sen University in 2012, and the PhD degree from the Chinese University of Hong Kong in 2016.

He is currently an associate professor with the College of Intelligence and Computing, Tianjin University, China. His research interests are computer vision and machine learning.

Weidong Zhang (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control science and engineering from Zhejiang University, China, in 1990, 1993, and 1996, respectively. He is currently the Director of the Engineering Research Center of Marine Automation Shanghai Municipal Education Commission, China. He was as a Postdoctoral Fellow with Shanghai Jiaotong University, Shanghai China, where he joined as an Associate Professor in 1998, and has been a Full Professor, since 1999. From 2003 to 2004, he was an Alexander von Humboldt Fellow with the University of Stuttgart, Stuttgart, Germany. From 2013 to 2017, he was Deputy Dean with the Department of Automation. He has authored or coauthored more than 300 papers and one book, and has been recognized as an Elsevier most cited Researcher. His research interests include control theory, machine learning theory, and their applications in industry and autonomous systems.

Dr. Zhang is a recipient of National Science Fund for Distinguished Young Scholars, China.

Steven X. Ding received the Ph.D. degree in electrical engineering from Gerhard–Mercator University, Duisburg, Germany, in 1992.

From 1992 to 1994, he was a Research and Development Engineer with Rheinmetall GmbH, Germany. From 1995 to 2001, he was a Professor of Control Engineering with the University of Applied Science Lausitz, Senftenberg, Germany, and the Vice President from 1998 to 2000. He is currently a Full Professor of Control Engineering and the Head of the Institute for Automatic Control and Complex Systems (AKS) with the University of Duisburg-Essen, Germany. His research interests include model-based and data-driven fault diagnosis, fault tolerant systems, real-time control, and their application in industry with a focus on automotive systems, chemical process, and renewable energy systems.