IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 June 2024, accepted 23 July 2024, date of publication 29 July 2024, date of current version 7 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3434708

RESEARCH ARTICLE

Enhancing Gas Turbine Fault Diagnosis Using a Multi-Scale Dilated Graph Variational Autoencoder Model

ZHANG KUN, LI HONGREN, WANG XIN, XIE DAXING, AND YANG SHUAI
Huadian Electric Power Research Institute Company Ltd., Hangzhou 310030, China

Corresponding author: Zhang Kun (3240569595@qq.com)

This work was supported by China Postdoctoral Science Foundation under Grant 282205.

ABSTRACT This paper proposes a Multi-scale Dilated Variational Graph Convolutional Autoencoder (MG-VAE) model for gas turbine fault diagnosis. The model integrates a multi-scale dilated convolutional attention mechanism to extract features across different scales, enhancing its ability to represent complex data and improving robustness in noisy environments. Additionally, a graph convolution module captures correlations between sensors, further enhancing diagnostic accuracy. Experimental results demonstrate the model's effectiveness, achieving high diagnostic accuracy in both gear fault simulation and real gas turbine fault datasets. Ablation experiments show that the integration of the graph convolutional network and the multi-scale dilated convolutional attention mechanism significantly improves accuracy, highlighting the model's potential for practical industrial applications in gas turbine fault diagnosis.

INDEX TERMS Gas turbine, fault diagnosis, MG-VAE, noise robustness.

I. INTRODUCTION

Since the 20th century, gas turbines, recognized for their outstanding characteristics as thermal machinery, have found extensive applications in various fields such as shipping, power generation, and aerospace, gradually becoming pivotal core power equipment. With the continuous increase in the number and scale of gas turbine installations, their security has become a subject of widespread concern. During the operational service of these devices, gas turbine failures occasionally occur, posing threats not only to personnel safety but also risking serious consequences such as production interruptions, energy wastage, and environmental pollution. In the practical maintenance and management of gas turbine equipment, the monitoring process is influenced comprehensively by complex structural units such as the intake section, low-pressure compressor, high-pressure system, combustion chamber, and turbine, leading to the intertwining of different fault types in critical components and posing challenges in feature extraction. Even though sensors can be independently deployed for key elements in the component monitoring process, the collected gas path or vibration signals still contain components from different sources and environmental noise. Additionally, correlations among sensors may cause interference in the fault state recognition of intelligent diagnostic models. Gas turbine fault diagnosis faces the following specific challenges: the complex structure of gas turbines, interference in component monitoring, overlapping fault modes with redundant noise, making fault feature extraction difficult; and the correlated nature of gas turbine sensor data, which traditional variational autoencoders cannot fully exploit, thereby limiting fault diagnosis performance.

Currently, fault diagnosis methods are divided into two main categories: traditional feature extraction and artificial intelligence network learning. In conventional feature extraction methods, fault types are identified based on the physical characteristics reflected in the monitoring signals.

Li et al. proposed a method integrating adaptive multi-scale filtering and hierarchical permutation entropy. They refined the extracted entropy features using the Laplacian scoring method and employed support vector machines to identify various degrees of faults in the gearbox [1]. In 2020, Mao et al. introduced an improved singular spectrum decomposition method, incorporating mode mixing indices and over-decomposition indices to select the

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

optimal embedding dimension. Additionally, they utilized the frequency-weighted energy slice bispectrum technique to effectively extract weak fault symptoms in rolling bearings [2]. Song et al. presented a support vector machine classification method based on a sparse scaling convex hull. They conducted dimensionality reduction of sample set features using the random forest algorithm and employed sparse techniques to adjust the convex hull scaling coefficients of the sample set, achieving high-performance fault recognition in rolling bearings [3].

While traditional fault recognition based solely on conventional features suitable for diagnosing simple rotating component faults, it becomes challenging when dealing with monitoring data from complex gas turbine equipment. Recognition results are susceptible to monitoring signal noise and reliance on human expertise. Combining artificial intelligence network learning effectively addresses these challenges. In 2019, Zhong et al. investigated a transfer learning method based on convolutional neural networks and support vector machines for gas turbine fault diagnosis, demonstrating excellent diagnostic performance under small sample conditions [4]. Zhao et al. developed a novel deep learning method named the Deep Residual Contraction Network. They achieved high-precision fault diagnosis by extracting features from highly noisy vibration signals by inserting soft thresholding as a nonlinear transformation layer into the deep structure [5]. Castro-Cros et al. compared and validated the superiority of different artificial intelligence network learning methods in their application to gas turbine diagnostics, showcasing their significant potential in this field [6]. Guo et al. used envelope spectrum analysis to extract fault information, followed by support vector machines for gear fault recognition [7]. Cheng et al. evaluated and selected features based on the stability and relevance of decision objectives, followed by gear damage identification using grey relational analysis [8]. Zuo et al. designed an ordered classification learning method for gear pitting fault recognition based on the recognition and ordinal relationships between fault severity and categories [9]. Wang et al. extracted time-frequency domain features and wavelet packet energy features from plunger pump monitoring data, achieving health status recognition using the K-nearest neighbor classifier [10]. Asgari et al. proposed an offline gas turbine diagnostic method based on artificial neural networks, which effectively predicted overall performance [11]. Li et al. used deep statistical feature learning to assess the health status of gearboxes and bearing systems, significantly improving the fault classification performance of the model [12]. Fentaye et al. described the development of fault diagnosis technology for modified gas turbines, demonstrating the huge potential application of artificial intelligence network learning methods in fault diagnosis [13].

The Variational Autoencoder (VAE), recognized as a powerful model for data feature learning and sample generation, acquires deep state representations of input signals and maps them to latent variable space [14], [15], [16], [17], [18], [19]. It can learn the latent distribution of fault features from complex data, providing new insights and methods for gas turbine fault diagnosis. In 2021, Dewangan et al. proposed an intelligent fault diagnosis solution based on a deep convolutional variational autoencoder. They achieved high accuracy on a rotating machinery dataset by extracting discriminative features and combining data reduction and random sampling techniques [20]. Yan et al. introduced an optimized stacked variational denoising autoencoder method for intelligent fault diagnosis in rolling bearings. Through integrating variational autoencoders and denoising autoencoders, robust fault features were extracted. Experimental results demonstrated that this method exhibited higher recognition accuracy in rolling bearing fault diagnosis than other deep learning methods [21]. However, applying variational autoencoders to gas turbine fault diagnosis presents specific challenges. Gas turbine structures are complex, and the monitoring process of components is subject to interference from self or other components. Internal fault modes overlap, and a considerable amount of redundant noise complicates the extraction of fault features. Additionally, sensor data in gas turbines typically exhibit correlations, and traditional fault diagnosis methods using variational autoencoders often fail to fully leverage this correlated information. This limitation may result in the model lacking awareness of the dependencies among sensors, thereby restricting the performance of fault diagnosis.

It is evident that artificial intelligence network learning methods have become the trend in recent years. This is largely due to the fact that the monitoring systems of major equipment often accumulate vast amounts of data, making it difficult to analyze using traditional feature extraction methods. In contrast, deep network learning methods, such as deep variational autoencoder models, can achieve autonomous feature acquisition based on data-driven approaches. The U.S. military research center has already applied deep network learning methods as an important end-to-end task solution and has achieved positive results in various aspects.

To address the practical challenges of gas turbine fault diagnosis, this paper proposes a method called the Multi-Scale Dilated Variational Graph Convolutional Network (MG-VAE). This approach introduces a multi-scale dilated convolutional attention mechanism and a graph convolutional network based on the variational autoencoder model. The multi-scale dilated convolutional attention mechanism allows the model to extract features and retain detailed information at various scales. Through this mechanism, the model can better represent complex data and enhance the robustness and accuracy of noisy monitoring signal recognition through denoising capabilities. Additionally, the incorporation of the graph convolutional module addresses sensor correlations. By constructing a graph structure among sensors and leveraging the capabilities of graph neural networks to enhance feature representation, the model can more effectively capture correlated information in gas turbine data.

This further improves the accuracy and robustness of fault diagnosis. Finally, validation through simulated gear fault tests and real gas turbine fault data demonstrates the effectiveness of the proposed method. Experimental results indicate that the MG-VAE significantly advances the original variational autoencoder model's ability to express complex data and achieves notable improvements in robustness and recognition accuracy. Section II lays the theoretical foundation, detailing the concepts of VAEs, GCNs, and the Multi-Scale Dilated Convolutional Attention Mechanism. Section III introduces the MG-VAE model, elaborating on its structure and integrating of these key concepts to enhance fault diagnosis capabilities. Section IV provides experimental verification using both simulated gear fault data and real gas turbine failure data, demonstrating the superior performance of our proposed model in various noise conditions. The paper concludes with a discussion of the results and their implications for gas turbine fault diagnosis in Section V.

II. THEORETICAL FOUNDATION

A. TRADITIONAL METHODS OVERVIEW

Autoencoders (AE) are unsupervised learning models used for data dimensionality reduction and feature representation. They consist of an encoder network and a decoder network with multiple layers of neural networks. The encoder learns latent features of the input signal and compresses the signal's dimensions, while the decoder reconstructs the original input signal using the learned features. By comparing the input and output differences, autoencoders obtain robust feature representations. They can extract features from signals and achieve dimensionality reduction. The features extracted from the encoding layer can capture important data features. By extracting features from the encoding layer and using supervised learning to classify them, autoencoders can be used for model classification tasks.

Convolutional Neural Networks (CNN) is a deep neural network with convolutional structures, widely used in deep learning models. It has significant advantages over traditional multilayer perceptrons. Through convolution operations, CNN can achieve local receptive fields and weight sharing, effectively reducing the number of parameters in deep networks, reducing memory usage, and alleviating overfitting. CNN generally consists of an input layer, convolutional layers, activation functions, pooling layers, fully connected layers, and an output layer. By stacking and training convolutional and pooling layers repeatedly, CNN can learn a complex non-linear mapping in the neural network, extracting features from complex input signals that can be used for classification. Fully connected layers can use the features extracted from the previous layers to classify input signals, finally outputting probabilities for each class.

Long Short-Term Memory (LSTM): is a variant of recurrent neural networks (RNNs) used for processing sequence data. It introduces gate mechanisms on the basis of traditional RNNs, effectively solving problems such as vanishing and exploding gradients in traditional RNNs. The core idea of LSTM is to introduce a memory cell to store and update information. The memory cell controls the flow of information through three main gates.

1. Input Gate, whether to add the current input to the memory cell.

2. Forget Gate, whether to delete part of the information from the memory cell.

3. Output Gate, how much information to output from the memory cell.

LSTM also has a step for calculating a new memory candidate value, which combines the current input and the previous hidden state to generate a new candidate memory value. The input gate then decides whether to add the new memory candidate value to the memory cell, while the forget gate decides whether to retain the old memory. LSTM's memory cell has the ability to store long-term memory, effectively handling and capturing long-term dependencies, and performing well in processing long sequence data. It alleviates the gradient problem in traditional RNNs to some extent, improving the model's performance and training effects.

B. VARIATIONAL AUTOENCODER

VAE is a generative deep neural network structure based on variational Bayesian inference, introduced by Kingma et al. in 2013 [22]. Deviating from conventional autoencoders, VAE describes the latent space in a probabilistic fashion. This unique approach enables the model to learn latent attribute features from the probability distribution within the latent variable space, subsequently facilitating the generation of new data. VAE stands out as a powerful deep generative model with broad applications in data generation.

Traditional autoencoders (AEs) are data compression models. During the compression process, the original data X is encoded through an encoding layer to output a low-dimensional variable Z, which is then decoded through a decoding layer to output the reconstructed data X′. During this process, discrete points are generated in the hidden space of the AE, which can lead to the generation of invalid content when decoding specific points in the hidden space. This, in turn, results in deficiencies in the interpretability and usability of the AE's hidden space. In contrast, variational autoencoders (VAEs) encode input data into probability distributions and constrain their hidden space distributions to closely approximate a standard normal distribution, thereby regularizing the structure of the hidden space. Consequently, the continuous encoding of the hidden space enhances its ability to encapsulate the information and noise presented in the input signal, thereby improving its effectiveness in fault diagnosis.

VAEs are categorized as approximate inference model that fit the true distribution by optimizing the evidence lower bound. Given a set of data samples $\{x_1, \cdots x_n\}$, collectively described by X, the goal is to infer the distribution $p(x)$ of X based on $\{x_1, \cdots x_n\}$. If achievable, sampling from $p(x)$ would yield all possible instances of X, including those beyond $\{x_1, \cdots x_n\}$, constituting an ideal generative model.

However, this process is often challenging, hence utilizing latent variables Z.

The structure of the variational autoencoder is illustrated in Figure 1.

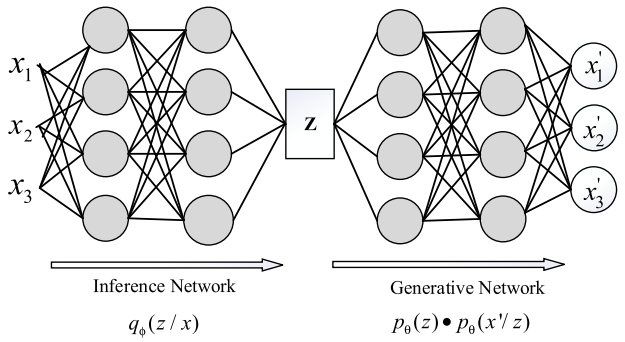


FIGURE 1. Variational autoencoder structure diagram.

As shown above, the VAE is similar to an autoencoder, comprising an encoder and a decoder. VAE uses neural networks to establish two probabilistic models: one that generates a probabilistic variational distribution of latent variables from the input data using variational inference, called the inference network (encoder), and another that generates an approximate probability distribution of the original data from the probabilistic variational distribution of the latent variables, called the generative network (decoder).

Assuming that the data can be generated according to a certain distribution containing unobservable continuous variables z, new data can be generated through the following two steps: First, find a value z_i from the prior distribution $p_\theta*(z)$; then, use the conditional probability distribution $p_\theta*(x/z)$ to generate x_i. VAEs introduce a recognition model $q_\phi(z/x)$ and a generative model $p_\theta(x/z)$, using neural networks to learn the parameters ϕ of the recognition model and the parameters θ of the generative model, ultimately approximating the true distribution. The recognition model $q_\phi(z/x)$ can be regarded as a probabilistic encoder, and the generative model $p_\theta(x/z)$, can be considered a probabilistic decoder. Although VAEs and AEs have a similar overall structure, their operational principles are entirely different. In VAEs, the outputs of the encoder and decoder are probability density distributions, and the latent space probability distribution is constrained to approach a standard Gaussian distribution, instead of generating specific codes as in AEs. The VAE's working process involves encoding the input into a probability distribution in the latent space using the encoder, sampling a point from this distribution, and then using the decoder to reconstruct the signal from the sampled point and calculate the reconstruction error.

Meanwhile, a constraint is imposed on the distribution of the latent space z, forcing it to closely approximate a standard normal distribution, ensuring both local and global regularization of the latent space. Local regularization is achieved through variance control, while global regularization is managed by mean control. This approach prevents the model's encodings in the latent space from diverging and encourages as much distribution ''overlap'' as possible, satisfying the conditions of continuity and completeness in the latent space.

To make the probability distributions $q_\phi(z/x)$ and $p_\theta(x/z)$ as similar as possible, one can minimize the Kullback-Leibler (KL) divergence between the two distributions. KL divergence measures the distance between two distributions; a smaller value indicates greater similarity, while a larger value indicates a greater discrepancy. The operational process of a Variational Autoencoder is as follows: First, the encoder encodes the input into a probability distribution in the latent space. Then, a point in the latent space is sampled from this distribution. Finally, the decoder to reconstruct the signal and calculate the reconstruction error. The loss function of a VAE is as follows:

$$L_{AVE} = L_{recon} + L_{KL} \tag{1}$$

In the equation above, the first term on the right side represents the reconstruction loss, and the second term represents the KL divergence. Regarding the reconstruction error term, the random sampling of the latent variable z can complicate backpropagation through the network. This issue is addressed using the reparameterization trick. Assuming the prior distribution $p_\theta(z)$ follows a certain normal distribution, one can sample ε from the standard normal distribution $N(0,1)$. In contrast, the encoder determines the parameters μ and δ of the generative distribution. Thus, when obtaining the latent variable, $z=\mu+\varepsilon \times \delta$ can be used as the latent variable for reconstruction. The reparameterization method allows the VAE network to efficiently perform sampling, as well as conduct derivative calculations and error backpropagation.

C. GRAPH CONVOLUTIONAL NETWORKS

The data generated by multiple sensors in gas turbines often exhibit complex spatial and temporal correlations, which traditional data processing methods may struggle to capture. Graph Convolutional Networks (GCNs), an emerging deep learning approach, provide a powerful tool for handling such multi-channel signals [23], [24], [25], [26], [27]. Multi-sensor data usually show spatial correlations, where the measurements from one sensor may be affected by those from adjacent sensors. GCNs can effectively capture this spatial correlation through graph convolution operations, thereby enhancing the accuracy of signal processing. Moreover, multi-channel signals often have time-series characteristics. GCNs can aggregate information in the temporal dimension through graph convolution operations. By incorporating graph convolution modules into the model, the signal processing performance can be significantly improved, providing more accurate results for fault diagnosis tasks.

GCNs are a type of deep learning model designed for processing graph-structured data. Unlike traditional Convolutional Neural Networks (CNNs) that are suited for handling regular grid-structured data such as images, GCNs can effectively manage the relationships and local structures

between nodes. In GCNs, the emphasis is on a graph composed of nodes and their interconnections. Each node has a feature vector representing its attributes or characteristics. The goal of GCNs is to learn representations for each node that capture the relationships between nodes and the global structure of the graph. The core idea of GCNs is to update the representation of each node based on information from its neighboring nodes. Specifically, for each node, GCNs aggregate the features of the node with those of its neighbors and merge the information of neighboring nodes through learned weights. In this way, each node can consider the features of related nodes and generate an updated representation. By stacking multiple graph convolutional layers, GCNs can gradually expand the receptive field of nodes and extract higher-level feature representations. Each graph convolutional layer refreshes nodes' representations and transfers these updated representations to the subsequent layer, allowing information to propagation through the network progressively enriching and refining the nodes' representations.

The computation process of GCNs can be summarized in the following steps. First, initialize the feature vectors of nodes by assigning an initial feature vector to each node in a directed or undirected graph, representing the attributes or features of the nodes. Graph neural networks extract spatial features of graph data through graph convolution. Assuming X is the input signal and Y is the output signal, the feature extraction method f of the graph convolutional neural network is defined as:

$$f(X, A) = Y \tag{2}$$

where A represents the adjacency matrix of the graph.

For a graph $G = (V, E, A)$, where V is the number of nodes in the graph $V = \{v_i\}^{\mathrm{N}}\,\mathrm{i} = 1$, E is the set of edges, and A is the adjacency matrix of the graph, with $A \in R^{N\times N}$ The forward propagation formula for graph convolution is

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{3}$$

$$\tilde{A} = A + I_n \tag{4}$$

$$\tilde{D} = \sum_{j}\tilde{A}_{ij} \tag{5}$$

where I_{n} represents identity matrix, $\tilde{D}$ denotes degree matrix, W is trainable parameters, H represents node parameters in the network. When I=0, i.e., for the first graph convolutional layer, $H^{(0)}$ is the original graph data X.

Traditional convolutional neural networks have achieved great success in mechanical fault diagnosis, but they also have some limitations. The self-learning process used for feature representation cannot explicitly explore the relationships between signals. In mechanical fault diagnosis, the relationships between monitoring signals change significantly with the health status of the machine. Therefore, modeling and learning the relationships between signals is very effective for machine fault diagnosis. The relationships that are missing between different sensors can be compensated for by constructing the sensor signals in the form of a graph. Representing different sensor data as nodes in graph convolution, each node is related to every other node through some complex edges. Edges are used to represent the relationships between sensors. Therefore, graph convolution can effectively utilize the graph structure to model the correlations between sensors, thereby better capturing the complex dependencies between signals. Additionally, graph convolutional networks propagate the feature information of nodes to their neighboring nodes and aggregate information through convolution operations on the graph structure. This ability for feature propagation and information aggregation enables graph convolutional networks to effectively utilize the relevant information between sensors and extract more representative feature representations.

D. THE MULTI-SCALE DILATED CONVOLUTIONAL ATTENTION MECHANISM

The monitoring signals of gas turbines contain high-frequency carrier components and various low-frequency modulation signal components, with fault information often residing in the low-frequency modulation signals. However, due to environmental and structural influences, signal redundancy and noise are severe, causing useful features to be obscured. When utilizing network models for fault diagnosis, it is necessary to focus on different types of signal components simultaneously, i.e., to acquire signal features from a multi-scale perspective. Therefore, a multi-scale dilated convolution attention mechanism is introduced to obtain feature information from different scales, further enhancing the model's robustness to noise.

Dilated convolution, also known as atrous convolution, is a common method that involves inserting spaces (zeros) between the elements of the convolution kernel to expand the kernel [28]. For larger objects that occupy more pixels in the original data, there may be discontinuities in the convolution kernel, which could result in not all points in the feature map participating in the computation, leading to a loss of information continuity. Introducing multi-scale dilated convolution and forming a cyclic structure can ensure that all pixels participate in computing in each cycle. The receptive field is one of the most fundamental concepts in convolutional neural networks, referring to the range of the input space corresponding to a pixel on the output feature map. The formula for calculating the size of the receptive field is as follows:

$$RF_{l+1} = RF_l + (f_{l+1} - 1) \times \prod_{i=1}^{l} s_i \tag{6}$$

where, RF_{l+1} represents the size of the receptive field corresponding to the current feature map, which is the target receptive field to be calculated. RF_l denotes the size of the receptive field corresponding to the feature map of the previous layer, f_{l+1} represents the size of the convolution kernel

of the preceding convolution layer, and the final product term represents the product of the strides of the preceding convolution layers. When the stride is greater than 1, the size of the receptive field will increase exponentially. In dilated convolution, the stride size can be adjusted by tuning the dilation rate, thereby achieving the adjustment of the receptive field.

Attention models enhance the model's ability to focus on important features and avoid interference from irrelevant features in downstream tasks, thereby improving model performance [29]. They are commonly applied in deep network models for tasks such as text translation and image processing and have become widely adopted network mechanisms or concepts.

In convolutional networks, high-level features are obtained from low-level features through operations such as stacking convolutions or pooling. The number of low-level features affecting a high-level feature element is limited, and this influence scale is described by the receptive field. Modulation components in mechanical vibration signals often contain fault information, such as gear meshing frequencies modulated by fault shaft rotational frequencies. To simultaneously focus on low-frequency modulation components and high-frequency carrier components for fault detection, a one-dimensional convolutional network needs a large and sparse receptive field to extract low-frequency modulation features and a small and dense receptive field to extract high-frequency carrier components, thereby extracting features at different scales. This paper designs a multi-scale dilated convolutional attention mechanism based on this consideration.

The receptive field of a certain layer in a convolutional network can be defined as the size of the region in the original input space that influences the features learned by that layer. An intuitive description is shown in Figure 2

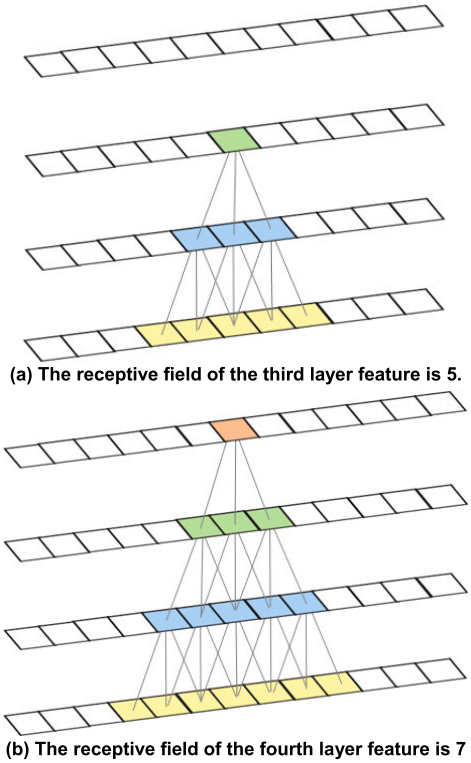
The calculation of the receptive field can be represented by the following recursive formula:

$$r^{(i)} = r^{(i\text{-}1)} + (k^{(i)} - 1) \cdot j^{(i-1)} \tag{7}$$

$$j^{(i\text{-}1)} = j^{(i-2)} \cdot s^{(i-1)} \tag{8}$$

where $r^{(i)}$ represents the receptive field of the i^{th} layer, $k^{(i)}$ denotes the size of the convolutional kernel for i^{th} layer, $j^{(i-1)}$ represents cumulative stride up to the i-1^{th} layer, $s^{(i-1)}$ represents stride of the i-1^{th} layer.

Analyzing the formula for calculating the receptive field, it is not difficult to find that convolutional networks can rapidly increase the receptive field by using multiple convolutions or pooling operations with strides. This enables the top-level features to respond to a certain range of inputs, corresponding to the local high-frequency components in the signal. To construct features that focus on low-frequency components, dilated convolution is introduced into one-dimensional convolutional networks. The core idea of dilated convolution is to insert holes between the elements of a regular convolutional kernel, and the size of these holes



(a) The receptive field of the third layer feature is 5.

(b) The receptive field of the fourth layer feature is 7

FIGURE 2. Convolutional receptive field diagram.

is represented by the dilation rate. When considering the dilation rate, the formula for calculating the receptive field of a convolutional network is modified as follows:

$$r^{(i)} = r^{(i\text{-}1)} + (k^{(i)} - 1) \cdot j^{(i-1)} \cdot d^{(i)} \tag{9}$$

$$j^{(i\text{-}1)} = j^{(i-2)} \cdot s^{(i-1)} \tag{10}$$

where $d^{(i)}$ represents dilation rate of the convolution in the i^{th} layer.

The introduction of the dilation rate can increase the receptive field, and at this point, the receptive field is "sparse". A "large and sparse" receptive field corresponds to low-frequency components in the features. A multi-scale convolution module formed by dilated convolutions with different dilation rates can gradually extract high-frequency and low-frequency features based on the features of the previous layer and fuse them. This module has the potential to extract useful features at multiple scales within a larger frequency band.

During the process of monitoring mechanical equipment in gas turbines, there is a large amount of noise in the monitored vibration signals due to external environmental and internal component interferences. This can adversely affect the classification task of data-driven pattern recognition models. The original VAE model can use a sample generation process and multi-layer convolutional networks for feature extraction to reduce noise interference and to some extent suppress noise. However, under strong noise signal inputs, further improvement is needed to enhance the model's robustness to noise.

The human visual system has a tendency to focus on information relevant to its own judgments when receiving visual information, while ignoring irrelevant information. Attention models work in a similar way by enhancing the model's focus on important features, avoiding interference from irrelevant features on downstream tasks, and thereby improving model performance. Attention mechanisms are widely used in deep network models for tasks such as text translation and image processing, and have become a widely adopted network mechanism or concept. According to their role and the region of interest in the network model, attention mechanisms can be classified into spatial attention, channel attention, and hybrid attention mechanisms. Spatial attention converts spatial information to another space during model training while focusing on important spatial information. Channel attention divides the input into different levels of importance from the channel perspective, thereby achieving a balance of information weights in the channel. The hybrid attention mechanism simultaneously uses both spatial and channel attention mechanisms to improve model performance.

Introducing an attention mechanism based on the multi-scale dilated convolution module allocates channel-level weights to features at different scales, especially adaptively assigning lower weights to channels where noise-related features are located. This suppresses their negative impact on the overall performance of composite features, thus avoiding interference from signal redundancy and noise.

In response to the challenge of severe noise in complex mechanical equipment monitoring signals, which makes pattern recognition difficult during fault diagnosis, this paper combines multi-scale dilated convolution with attention mechanisms to establish a multi-scale dilated convolution attention mechanism. This approach aims to avoid interference from signal redundancy noise, with the specific structure shown in Figure 3.

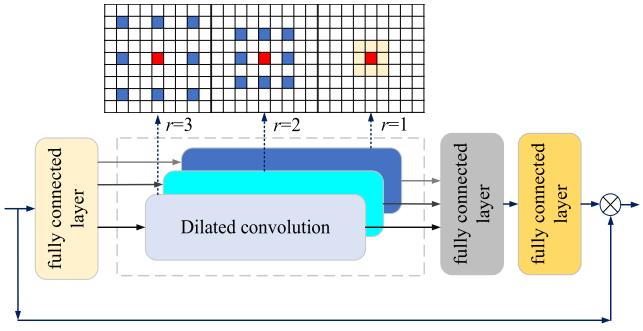


FIGURE 3. Schematic diagram of a multi-scale dilated convolution attention mechanism.

III. MULTI-SCALE DILATED GRAPH VARIATIONAL AUTOENCODER

To diagnose faults, an MG-VAE model is constructed based on the VAE model as follows:

- Encoder with Graph Convolution and Multi-scale Dilated Convolution Attention Mechanism. In the encoder part of the original variational autoencoder, graph convolution, along with the multi-scale dilated convolution attention mechanisms are introduced to encode the input vibration signals. This step achieves feature extraction and dimensionality reduction, resulting in a compressed latent variable Z.
- Decoder with Deconvolution Network: A decoder composed of a deconvolution network is used to reconstruct the signals from the latent variable Z. During this process, Z learns to contain features related to faults.
- Classifier with Separate Neural Network: The latent variable z is then input into a classifier constructed by another neural network pathway to perform fault classification tasks.

The final model is the multiscale dilated variational graph convolutional network, as illustrated in Figure 4.

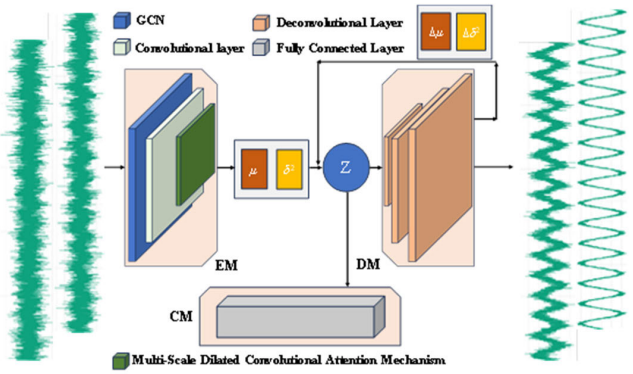


FIGURE 4. Multiscale dilated variational graph convolutional global model.

In the global model graph, the original input data x is encoded by the encoding network EM to obtain the hidden layer distribution information μ and δ^2. Through the resampling operation, the hidden layer variable z is obtained, which is then input into the decoding network DM for sample generation and outputting the correction parameters $\Delta\mu$ and $\Delta\delta^2$, for updating the posterior distribution of the hidden variables. Simultaneously, the multiscale information of the decoding network is obtained for the classification and recognition network CM, thereby unfolding the sample recognition analysis. The multiscale dilated attention mechanism can further improve the feature extraction ability and reduce noise interference, while the graph convolutional structure can effectively capture the correlation between different sensor information, thus enhancing the model's diagnostic accuracy and robustness to noise. Finally, the loss function of the i^{th} sample during the training of the network model can be expressed as:

$$L^{(i)} = \sum_{j}^{m} -y_j^{(i)} \log \tilde{y}_j^{(i)} + \left\| x^{(i)} - \tilde{x}^{(i)} \right\|_2^2 + \sum_{d}^{D} \left(1 + \log \left(\sigma_d^{(i)} \right)^2 - \left(\mu_d^{(i)} \right) - \left(\sigma_d^{(i)} \right)^2 \right) \quad (11)$$

where the loss function $L^{(i)}$ for the i^{th} sample is represented as follows: $\sum_{j}^{m} -y_j^{(i)} \log \tilde{y}_j^{(i)}$ denotes the cross-entropy loss; $\left\| x^{(i)} - \tilde{x}^{(i)} \right\|_2^2$ denotes the mean squared loss for the reconstructed signal and $\sum_{d}^{D} \left(1 + \log\left(\sigma_d^{(i)}\right)^2 - \left(\mu_d^{(i)}\right) - \left(\sigma_d^{(i)}\right)^2\right)$ represents the *KL* divergence loss between the latent variables and the independent multivariate standard normal distribution.

The multi-scale dilated variational graph convolutional network model proposed in this paper encompasses three main innovations:

1. Introduction of a multi-scale dilated convolutional attention mechanism. This paper integrates multi-scale dilated convolution with an attention mechanism to form a multi-scale dilated convolutional attention mechanism. This mechanism enables the model to extract features at different scales while enhancing focus on important features through the attention mechanism. Consequently, it effectively handles noise and redundant information in gas turbine monitoring signals, thereby improving the model's ability to recognize fault features.

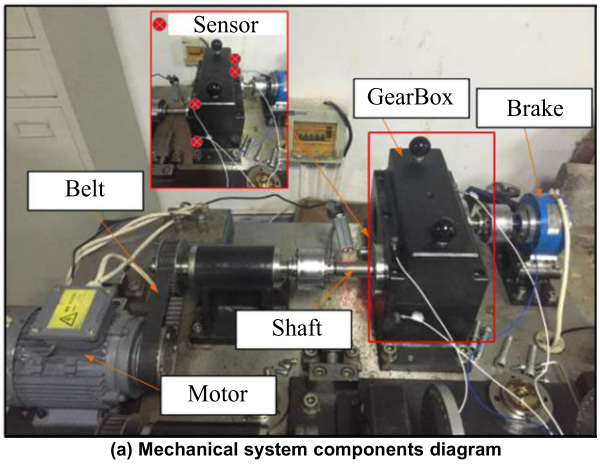
2. Application of graph convolutional networks. To better address the complex spatial and temporal correlations in multi-sensor data from gas turbines, this paper introduces GCNs. GCNs effectively capture the correlations between sensors, and enhance the expressive power of feature representation by constructing the graph structure of sensors and utilizing the capabilities of graph neural networks, Consequently, this improves the accuracy of fault diagnosis.

3. MG-VAE model. This paper constructs a novel multi-scale dilated variational graph convolutional network model. This fusion not only fully leverages the capabilities of VAE in feature learning and data generation but also enhances the model's ability to capture signal details through multi-scale dilated convolution and handle the complex correlations between sensor data through GCN. This integrated approach significantly improves the model's performance in gas turbine fault diagnosis tasks.

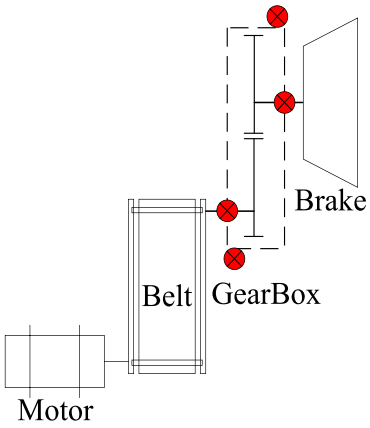
IV. EXPERIMENTAL VERIFICATION

A. GEAR FAULT SIMULATION EXPERIMENT

Under high-speed and heavy-load conditions, the gearbox is a component of gas turbines prone to failure. Failures in the gearbox transmission system often lead to safety incidents, resulting in significant human and economic losses [30], [31], [32], [33], [34], [35], [36], [37]. In the operation and maintenance management of gas turbine units, the gearbox is considered a critical component to monitor, and its fault diagnosis is crucial for the safe operation and maintenance of gas turbines. To validate the effectiveness of the MG-VAE model in the operation and maintenance management of essential components in gas turbines, a gearbox transmission system fault simulation experiment was conducted.



(a) Mechanical system components diagram



(b) Drive system schematic

FIGURE 5. Gear fault simulation experiment.

1) DATASET INTRODUCTION

A predetermined fault simulation experiment for the gearbox was conducted. The test bench comprised components such as a driving motor, transmission belt, transmission shaft, gearbox, and brake, as depicted in Figure 5. Gear failures in the gearbox transmission system were simulated by pre-setting gear cracks and wear faults.

During the preset fault simulation experiment on the gearbox, gear parameters were established as detailed in Table 1. Four vibration acceleration sensors were mounted on the gearbox near the input and output shafts. With one sensor installed in both the vertical and horizontal directions at each location. The arrangement of the sensors is depicted in Figure 5. The experiment included eight variations of gear failure scenarios: a normal condition, single-tooth cracks ranging from 1mm to 4mm, a 2mm double-tooth crack, 0.5mm single-tooth wear, and 0.5mm double-tooth wear. The cracks, spanning the width of the tooth, were positioned at the tooth root, while the wear occurred on the tooth surface. Some of these fault modes are presented in Figure 6. The rotational speed was set to 1000 RPM, and for each fault mode, a dataset with a duration of approximately 115 seconds was collected at a sampling rate of 10,240 Hz.

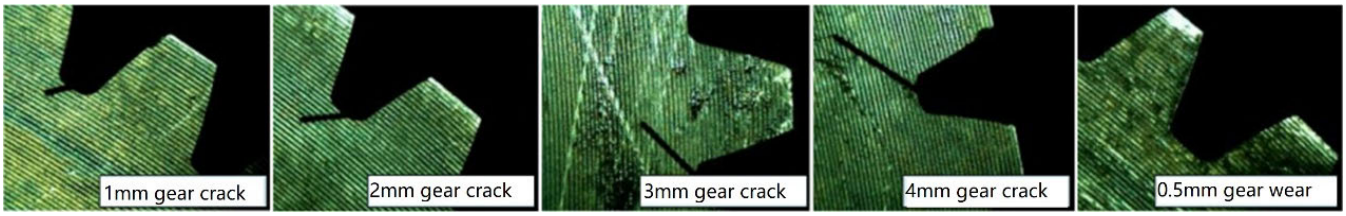


FIGURE 6. The gearbox is pre-configured with a malfunction.

TABLE 1. Test bench gear parameters.

Gear type	number of teeth	module /mm	width/mm	Helix angle /°
Driving	53	2	20	10.0633
Driven	75	2	20	10.0633

The raw vibration signals collected from the gearbox under normal conditions and different fault modes are shown in Figure 7. Upon analyzing the time-domain vibration signals, it becomes evident that the waveform patterns vary. However, distinguishing between different fault modes based solely on these variations is not straightforward or intuitive

To verify the effectiveness of the proposed MG-VAE model, the collected preset fault monitoring data were used as model inputs to conduct gearbox fault identification. In the specific implementation process, the input data sequence was divided into a training set and a test set at a ratio of 4:1. Additionally, a sliding window with a length of 128 was employed to perform interval-based non-overlapping sliding sampling on the 4-channel data, with the interval length set to 220, thereby constructing samples with dimensions of 4×128. Based on this, the total number of samples and the number of test samples obtained are shown in Table 2. The samples were normalized to establish the model input sample set, thereby validating the target model.

To simulate the various noise interferences present during the actual operation of a gas turbine gearbox and to verify the effectiveness of the model diagnosis under strong noise environments, the experiment set three noise levels from high to low. Gaussian white noise with a mean of 0 and different standard deviations was added to the original signal. The noise level settings are shown in Table 3.

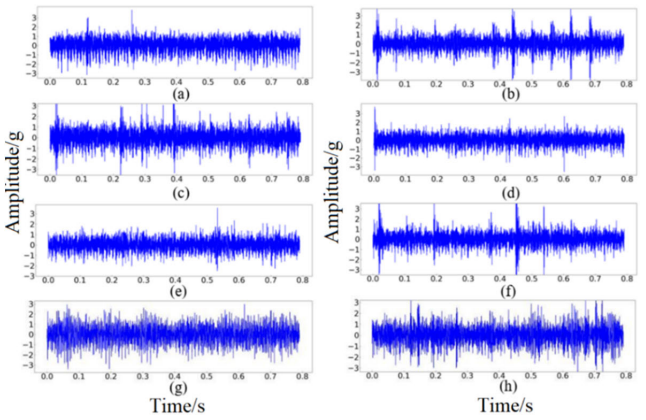


FIGURE 7. Time-domain waveforms of gear vibration signals under eight different conditions: (a) normal condition, (b) 1mm crack fault, (c) 2mm crack fault, (d) 2mm double-tooth crack fault, (e) 3mm crack fault, (f) 4mm crack fault, (g) 0.5mm wear fault, (h) 0.5mm double-tooth wear fault.

2) ANALYSIS OF EXPERIMENTAL RESULTS

To assess the performance of the MG-VAE model at different noise levels, the experimental results presented in Figure 8 illustrate the accuracy and loss metrics during the training process, demonstrating the model's convergence and effectiveness in fault diagnosis.

As illustrated in Figure 6, the MG-VAE model achieves its highest accuracy after 5 training epochs at noise level I, whereas, it requires 100 epochs to attain convergence accuracy at noise level III. The challenge of training the model significantly escalates with the increase in noise level thereby exerting a considerable impact on model accuracy. To further validate the improvement in accuracy of the MG -VAE model at different noise levels, we selected other mainstream fault diagnosis models for comparison, including AE, CNN, and LSTM. These three models possess the same network layers and hyperparameter settings as MG -VAE and are trained with gearbox fault data. The final results are presented in Table 4.

It is evident from the recognition results in Table 4, it can be seen that as the noise level increases, the MG-VAE model consistently outperforms other network models, maintaining high accuracy even under strong noise conditions. This demonstrates the superior performance of the model in complex signal environments. Thus, the effectiveness of the MG-VAE model in enhancing fault diagnosis accuracy under strong noise conditions is validated. The confusion matrices of the MG-VAE model's test results at noise levels I and III are depicted in Figure 9, with the horizontal axis representing predicted labels and the vertical axis representing the true labels of the samples.

The GM-VAE model uses feature reduction to 2 dimensions for visualization at noise levels I and III, with the results shown in Figure 10.

TABLE 2. Total number of samples for different fault modes.

Sample set	Normal	1mm Crack	2mm Crack		3mm Crack	4mm Crack	0.5mm wear	
			Single tooth	Double tooth			Single tooth	Double tooth
Total number of samples	5520	5230	5160	5195	5160	5195	5160	5160
Number of test samples	1104	1046	1032	1039	1032	1039	1032	1032

TABLE 3. Settings for different noise levels.

noise levels	Mean	Standard deviation
I	0	0.18
II	0	0.32
III	0	1.00

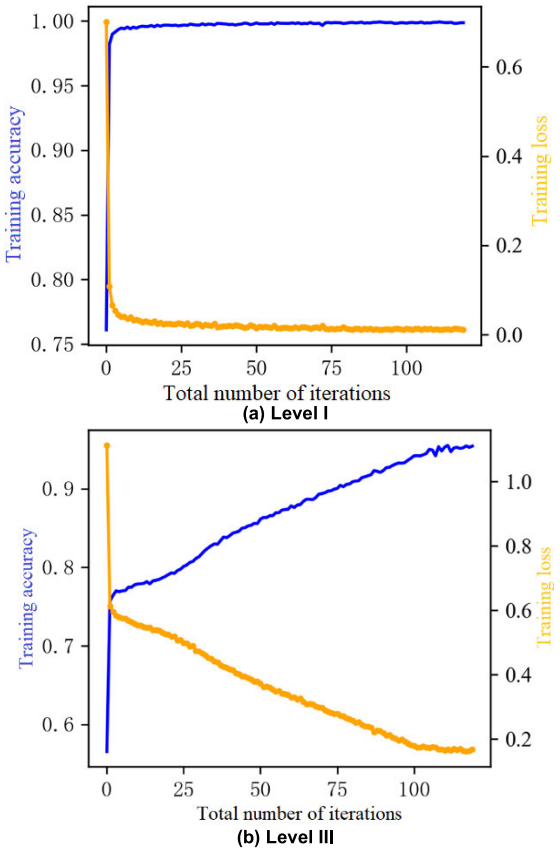


FIGURE 8. Changes in accuracy and loss during the training process.

B. GEAR FAULT SIMULATION EXPERIMENT

Gas turbine sets have high reliability requirements and a long lifespan. The working conditions of critical components can change under different loads or environmental states, leading to various potential faults. Fault-type identification methods, can establish a connection between critical components and the working environment, providing essential information for maintenance decisions. Therefore, conducting fault identification for multiple gas turbine units holds significant importance. This section validates the fault identification model using monitoring data from multiple gas turbine sets.

TABLE 4. Fault diagnosis results of different models for noisy monitoring signals.

Noise level	Model	Training set accuracy (%)	Test set accuracy (%)
level I	AE	96.28	92.33
	CNN	100	97.87
	LSTM	98.07	94.28
	GM-VAE	99.87	98.41
level II	AE	97.78	76.51
	CNN	99.04	82.38
	LSTM	86.34	80.73
	GM-VAE	99.16	93.65
level III	AE	55.33	35.24
	CNN	81.21	38.36
	LSTM	72.17	64.65
	GM-VAE	94.56	73.52

1) MONITORING DATA ANALYSIS

Table 5 presents field monitoring data from gas turbine units under different equipment conditions. The data primarily focus on monitoring the rotational speed of the units. After the monitoring period or the occurrence of an abnormal event, disassembly and bore probing reveal the types of faults involved in each unit, including normal condition, blade detachment, turbine damage, blade block loss, and nozzle abnormalities.

The normalized rotational speed of the collected units is partially shown in the time domain waveform in Figure 11. It is challenging to intuitively distinguish different fault states from the original monitoring time domain signals. Data analysis is required to find information related to fault types within the internal structure of the data. The multiscale dilated variational graph convolutional network model is used for data analysis and processing, and model validation is conducted.

Due to the significant presence of invalid values and outliers in the original data, data cleaning is performed on the monitoring data before constructing model input samples. This process primarily involves removing anomalies, null values, and invalid monitoring data. The data is normalized to reduce the influence of dimensions by scaling the data range to the interval [0, 1]. Subsequently, random sampling is carried out on the dataset using a sliding window of size

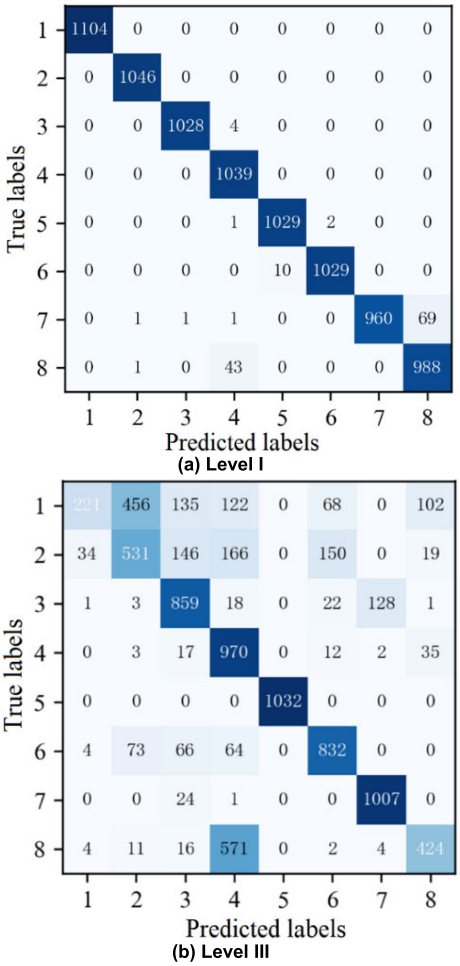


FIGURE 9. Confusion matrices of the GM-VAE model at different noise levels.

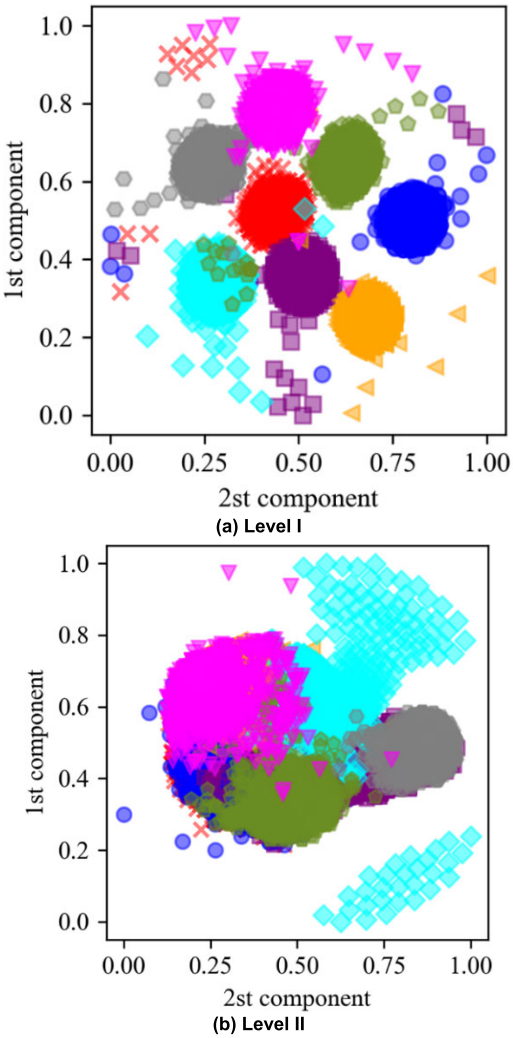


FIGURE 10. Visualization results of the MG-VAE model's features at different noise levels.

128 to establish a sample set, wherein the training and test set samples are divided in a 4:1 ratio. The test set does not overlap with the training set, thereby validating the model's generalization ability. As a result, we have acquired a total of 3936 samples for the training set and 984 samples for the test set.

2) FIELD MONITORING DATA ANALYSIS

During the training process, the relationship between training accuracy and training loss with the number of iterations is depicted in Figure 12. It can be observed that the training process is stable, and the accuracy rises to above 95% after 18 iterations of training.

To validate the effectiveness of the MG-VAE model, we constructed several models for comparison, including AE, CNN, and LSTM. The network layer settings of these models are the same as those in the gearbox simulation experiment described in the previous section. The models are trained using real gas turbine data, and the final results are presented in Table 6.

On the test set, the MG-VAE model performs the best, achieving an accuracy of 99.39%, followed by the CNN model with an accuracy of 97.87%. The LSTM model's test set accuracy is 94.28%, slightly lower than CNN and MG-VAE. The autoencoder (AE) has the lowest accuracy on the test set, at 92.33%. Considering the performance on both the training and test sets, the MG-VAE model demonstrates the highest performance in fault diagnosis tasks for real gas turbine monitoring signals, indicating that MG-VAE can more effectively capture the complex patterns and underlying structures of such time series data.

The test set classification confusion matrix and the visualization results of the sample's hidden layer feature reduction are shown in Figure 13.

TABLE 5. Monitoring data of gas turbine sets faults.

No.	Fault types	Fault labels
F1	Normal operation	0
F2	Blade detachment	1
F3	Turbine damage	2
F4	Blade block loss	3
F5	Nozzle anomaly	4

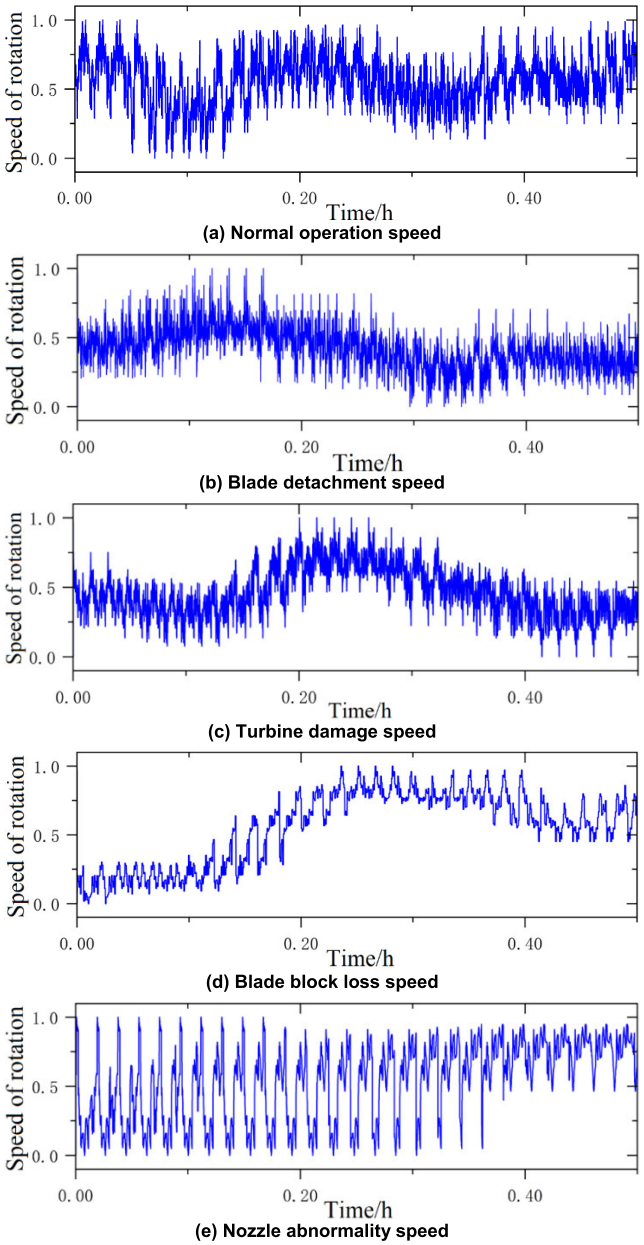


FIGURE 11. Time domain signals under different conditions.

As depicted in the figure, the model accurately classified the bulk of samples in the test set. The hidden layer distribution in the model revealed distinct separation among abnormal data caused by different fault types within the feature space, with data of the same fault type gradually clustering together. The model has effectively learned the internal fault features of the collected data and assigned them to respective fault types. These results validate the model's applicability of the model to actual field monitoring data of gas turbine sets.

C. ABLATION EXPERIMENT

To verify the effectiveness of the improved method, an ablation experiment was conducted as presented in Table 7, utilizing the dataset from the gearbox simulation fault

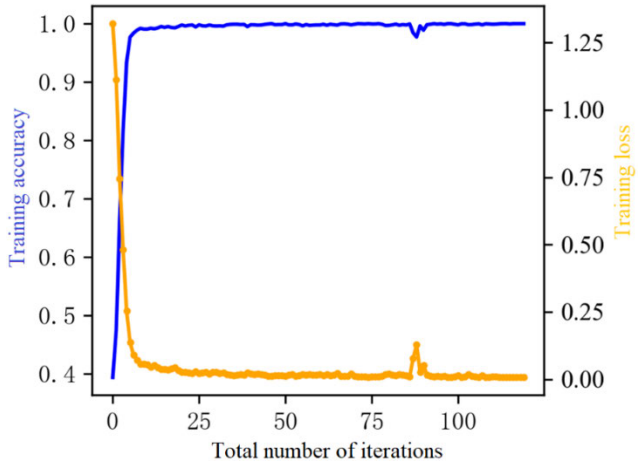


FIGURE 12. Accuracy and loss changes during the training process.

TABLE 6. Fault diagnosis results of different models for real gas turbine monitoring signals.

Data samples	Model	Training set accuracy (%)	Test set accuracy (%)
gas turbine data	AE	96.28	92.33
	CNN	100	97.87
	LSTM	98.07	94.28
	GM-VAE	99.87	99.39

experiment at noise level III. In the table, G represents the graph convolutional network structure, and M represents the multi-scale dilated convolutional attention mechanism. The ablation experiment entailed a quantitative analysis of the contributions of the multi-scale dilated convolutional attention mechanism (M) and the graph convolutional network (G) structure to the gas turbine fault diagnosis model. By comparing the training and testing accuracies of different model configurations, we assessed the performance improvement of each structure relative to the baseline model (using only the VAE). Incorporating the graph convolutional network structure (G) into the baseline model increased the training accuracy from 84.58% to 89.21%, indicating a 4.63% improvement. Regarding testing accuracy, there was an increase from 70.12% to 71.66%, representing a 1.54% improvement. This suggests that the graph convolutional network effectively utilizes spatial correlations between sensor data, enhancing the model's ability to recognize fault features. After introducing the multi-scale dilated convolutional attention mechanism (M), the training accuracy further increased to 88.51%, showing a 3.93% improvement compared to the baseline model. Testing accuracy also improved, from 70.12% to 72.16%, an increase of 2.04%. This confirms the advantages of multi-scale dilated convolution in handling noise and extracting signal details. By employing both the graph convolutional network and the multi-scale dilated convolutional attention mechanism (i.e., the MG-VAE model), the training accuracy reached 94.56%, showing a significant improvement of 10.98% compared to

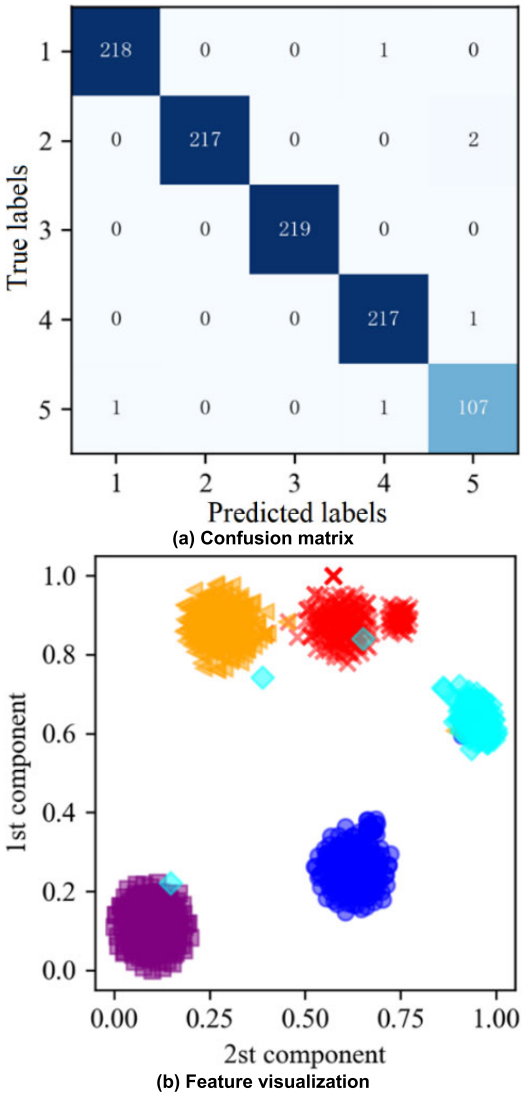


FIGURE 13. Test set confusion matrix and hidden layer feature visualization.

TABLE 7. Comparison of ablation experiment results.

Model	Gear data (Noise level III)	
	Training accuracy (%)	Test accuracy (%)
VAE	84.58	70.12
VAE+G	89.21	71.66
VAE+M	88.51	72.16
MG-VAE	94.56	73.52

the baseline model. In terms of testing accuracy, it increased from 70.12% to 73.52%, representing a 3.4% improvement. This result highlights the complementarity of the two structures and underscores the importance of combining multiple advanced technologies in fault diagnosis of complex systems.

The results of the ablation experiment not only validate the effectiveness of each structure but also highlight their synergistic effect in improving the accuracy of gas turbine fault diagnosis. These findings offer valuable insights for future fault diagnosis in similar complex systems, particularly in designing diagnostic models capable of handling high noise and complex data.

The models were built on the PyCharm platform using the PyTorch framework. The experiments were conducted on a computer equipped with an RTX 2060 GPU and an i5-8400 CPU. The Adam optimizer was used for model training, with an initial learning rate of 0.001, a batch size of 64 per training round, and 130 training iterations.

V. CONCLUSION

The MG-VAE model proposed in this paper has demonstrated outstanding performance in gas turbine fault diagnosis. By integrating a multi-scale dilated convolutional attention mechanism, the model efficiently extracts data features across various scales, enhancing its ability to represent complex data and improving its robustness and accuracy in recognizing for noisy monitoring signals. Additionally, the introduced graph convolution module effectively handles the correlations between sensors, capturing associative information in gas turbine data and further enhancing fault diagnosis accuracy. Across both the gear fault simulation experiment and validation with real gas turbine fault data, the model demonstrated high diagnostic accuracy, particularly in challenging noise environments. In the gear fault simulation experiment at noise level III, the test accuracy attained 73.52%, while in the real gas turbine fault dataset, it reached 99.39%, markedly surpassing other mainstream fault diagnosis models. Ablation experiments revealed that compared to the baseline model using only the VAE, the incorporation of the graph convolutional network structure improved test accuracy by 1.54%, while the integration of the multi-scale dilated convolutional attention mechanism increased it by 2.04%. With both structures operating simultaneously, i.e., the MG-VAE model, test accuracy rose to 73.52%, marking a 3.4% improvement over the baseline model. This underscores the synergistic effect of the graph convolutional network and the multi-scale dilated convolutional attention mechanism in handling complex and noisy signals, as well as their pivotal role in enhancing gas turbine fault diagnosis. These findings underscore the efficacy of the proposed model as a solution for gas turbine fault diagnosis, poised to make significant contributions to practical industrial applications.

REFERENCES

[1] Y. Li, G. Li, Y. Yang, X. Liang, and M. Xu, "A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy," *Mech. Syst. Signal Process.*, vol. 105, pp. 319–337, May 2018.

[2] Y. Mao, M. Jia, and X. Yan, "A new bearing weak fault diagnosis method based on improved singular spectrum decomposition and frequency-weighted energy slice bispectrum," *Measurement*, vol. 166, Dec. 2020, Art. no. 108235.

[3] S. Renwang, Y. Baiqian, S. Hui, Y. Lei, and D. Zengshou, "Support vector machine fault diagnosis based on sparse scaling convex hull," *Meas. Sci. Technol.*, vol. 34, no. 3, Mar. 2023, Art. no. 035101.

[4] S.-S. Zhong, S. Fu, and L. Lin, "A novel gas turbine fault diagnosis method based on transfer learning with CNN," *Measurement*, vol. 137, pp. 435–453, Apr. 2019.

[5] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.

[6] M. de Castro-Cros, M. Velasco, and C. Angulo, "Machine-learning-based condition assessment of gas turbines—A review," *Energies*, vol. 14, no. 24, p. 8468, Dec. 2021.

[7] L. Guo, J. Chen, and X. Li, "Rolling bearing fault classification based on envelope spectrum and support vector machine," *J. Vibrat. Control*, vol. 15, no. 9, pp. 1349–1363, Sep. 2009.

[8] Z. Cheng, N. Hu, M. Zuo, and B. Fan, "Crack level estimation approach for planetary gear sets based on simulation signal and GRA," *J. Phys. Conf. Ser.*, vol. 364, May 2012, Art. no. 012076.

[9] X. Zhao, M. J. Zuo, Z. Liu, and M. R. Hoseini, "Diagnosis of artificially created surface damage levels of planet gear teeth using ordinal ranking," *Measurement*, vol. 46, no. 1, pp. 132–144, Jan. 2013.

[10] W. Haoren, H. Yixiang, and Z. Shuai, "Health assessment for a piston pump based on WPD and LE," *J. Vib. Shock*, vol. 36, no. 22, pp. 45–50, 2017.

[11] H. Asgari, X. Chen, M. B. Menhaj, and R. Sainudiin, "Artificial neural network–based system identification for a single-shaft gas turbine," *J. Eng. Gas Turbines Power*, vol. 135, no. 9, Sep. 2013, Art. no. 92601.

[12] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis," *Neurocomputing*, vol. 168, pp. 119–127, Nov. 2015.

[13] A. D. Fentaye, A. T. Baheta, S. I. Gilani, and K. G. Kyprianidis, "A review on gas turbine gas-path diagnostics: State-of-the-Art methods, challenges and opportunities," *Aerospace*, vol. 6, no. 7, p. 83, Jul. 2019.

[14] D. Dehaene and R. Brossard, "Re-parameterizing VAEs for stability," 2021, *arXiv:2106.13739*.

[15] A. Asperti, D. Evangelista, and E. Loli Piccolomini, "A survey on variational autoencoders from a green AI perspective," *Social Netw. Comput. Sci.*, vol. 2, no. 4, p. 301, Jul. 2021.

[16] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," 2019, *arXiv:1906.02691*.

[17] S. Chen and W. Guo, "Auto-encoders in deep learning—A review with new perspectives," *Mathematics*, vol. 11, no. 8, p. 1777, Apr. 2023.

[18] L. Girin, S. Leglaive, X. Bie, J. Diard, T. Hueber, and X. Alameda-Pineda, "Dynamical variational autoencoders: A comprehensive review," 2020, *arXiv:2008.12595*.

[19] A. Asperti, D. Evangelista, and E. Loli Piccolomini, "A survey on variational autoencoders from a GreenAI perspective," 2021, *arXiv:2103.01071*.

[20] G. Dewangan and S. Maurya, "Fault diagnosis of machines using deep convolutional beta-variational autoencoder," *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 287–296, Apr. 2022.

[21] X. Yan, Y. Xu, D. She, and W. Zhang, "Reliable fault diagnosis of bearings using an optimized stacked variational denoising auto-encoder," *Entropy*, vol. 24, no. 1, p. 36, Dec. 2021.

[22] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[23] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.

[24] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Comput. Social Netw.*, vol. 6, no. 1, p. 11, Dec. 2019.

[25] U. A. Bhatti, H. Tang, G. Wu, S. Marjan, and A. Hussain, "Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–28, Feb. 2023.

[26] F. Wu, T. Zhang, A. Holanda de Souza, C. Fifty, T. Yu, and K. Q. Weinberger, "Simplifying graph convolutional networks," 2019, *arXiv:1902.07153*.

[27] J. Zhang, W. Ma, E. Zhang, and X. Xia, "Time-aware dual LSTM neural network with similarity graph learning for remote sensing service recommendation," *Sensors*, vol. 24, no. 4, p. 1185, Feb. 2024.

[28] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[29] Z. Shen, X. Kong, L. Cheng, R. Wang, and Y. Zhu, "Fault diagnosis of the rolling bearing by a multi-task deep learning method based on a classifier generative adversarial network," *Sensors*, vol. 24, no. 4, p. 1290, Feb. 2024.

[30] I. Khalfaoui-Hassani, T. Pellegrini, and T. Masquelier, "Dilated convolution with learnable spacings," 2021, *arXiv:2112.03740*.

[31] L. Chen, L. Gu, and Y. Fu, "Frequency-adaptive dilated convolution for semantic segmentation," 2024, *arXiv:2403.05369*.

[32] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[33] Y. Shanqin, W. Jinhua, and C. Jie, "Gearbox fault diagnosis method based on PSM-BN," *IEEE Access*, vol. 12, pp. 39876–39886, 2024.

[34] G. Istenes and J. Polák, "Investigating the effect of gear ratio in the case of joint multi-objective optimization of electric motor and gearbox," *Energies*, vol. 17, no. 5, p. 1203, Mar. 2024.

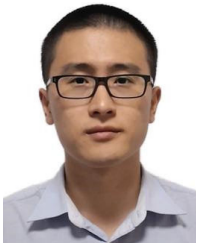
[35] L. Xu, T. Wang, J. Xie, J. Yang, and G. Gao, "A mechanism-based automatic fault diagnosis method for gearboxes," *Sensors*, vol. 22, no. 23, p. 9150, Nov. 2022.

[36] M. S. Raghav and R. B. Sharma, "A review on fault diagnosis and condition monitoring of gearboxes by using AE technique," *Arch. Comput. Methods Eng.*, vol. 28, no. 4, pp. 2845–2859, Jun. 2021.

[37] M. Sohaib, S. Munir, M. M. M. Islam, J. Shin, F. Tariq, S. M. M. A. Rashid, and J.-M. Kim, "Gearbox fault diagnosis using improved feature representation and multitask learning," *Frontiers Energy Res.*, vol. 10, pp. 1–23, Sep. 2022.

ZHANG KUN received the B.S. and M.S. degrees in electronic science from Northeast Petroleum University, Daqing, China, in 2013, and the Ph.D. degree in electrical machines and appliances from Harbin Institute of Technology, Harbin, China. He is currently pursuing the Postdoctoral degree in electrical engineering with Zhejiang University, Hangzhou, China.

From 2021 to 2023, he was a Research Assistant with the Huadian Electric Power Research Institute, Hangzhou. His research interests include fault diagnosis and data-driven modeling of gas turbine.



LI HONGREN received the B.S. degree in thermal engineering from Jilin University, Changchun, in 2008, and the M.S. degree in aerospace engineering from Huazhong University of Science and Technology, Wuhan, China, in 2012.

From 2012 to 2023, he was a Research Assistant with the Huadian Electric Power Research Institute, Hangzhou, China. His research interest includes fault warning and diagnosis of gas turbine systems.

WANG XIN received the B.S. degree in electrical engineering and automation and the M.S. degree in aerospace engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2021.

From 2021 to 2023, she was a Research Assistant with the Huadian Electric Power Research Institute, Hangzhou, China. Her research interests include fault warning and diagnosis of gas turbine systems.

XIE DAXING received the B.S. and M.S. degrees in power thermodynamics and engineering from Southeast University, Nanjing, China, in 2009.

From 2009 to 2023, he was the Director of the Gas Turbine Institute and the Huadian Electric Power Research Institute, Hangzhou, China. His research interest includes combined cycle performance of gas turbines.

YANG SHUAI received the B.S. degree in energy and power engineering from Huazhong University of Science and Technology, Wuhan, in 2015, and the M.S. degree in mechanical engineering from National Tsing Hua University, Taiwan, in 2020.

From 2020 to 2023, he was a Research Assistant with the Huadian Electric Power Research Institute, Hangzhou, China. His research interests include new energy and mechanical engineering.