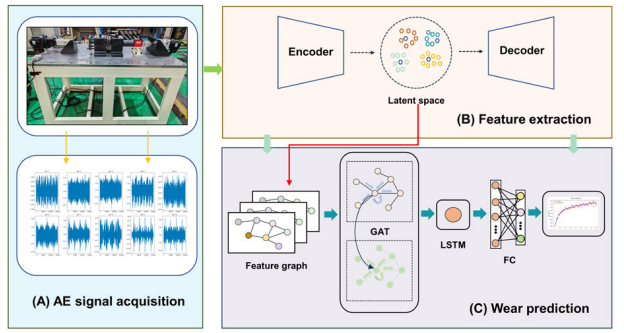
# An Improved Variational Autoencoder and Graph Attention Network Method for Wear Prediction of Aerospace Self-Lubricating Bearing Using Acoustic Emission Signal

Danyue Shen, Shichang Du, *Member, IEEE*, Shuo Wang, Liang Yan, Shanshan Li, and Xianmin Chen

***Abstract*—Aerospace self-lubricating bearings are critical components in aircraft transmission systems, where wear-induced degradation under high-load and dynamic conditions poses significant challenges to operational safety and system longevity. In recent years, deep-learning methods have shown promise in wear prediction by leveraging abundant monitoring data from sensor networks. However, these methods often struggle to detect early-stage degradation and rely on labor-intensive feature engineering, limiting their effectiveness in handling noisy, high-dimensional data. To overcome these issues, this article proposes an improved variational autoencoder (VAE) and graph attention network method for wear prediction based on acoustic emission (AE) signals. First, a clustering-guided contrastive VAE (CGC-VAE) model is proposed to process noisy, high-dimensional AE signals. The CGC-VAE employs *K*-means clustering to segment wear stages, combined with Gaussian mixture model (GMM) regularization and contrastive learning, to extract low-dimensional, discriminative latent features. Subsequently, a temporal graph attention network (T-GAT) is proposed to construct a dynamic graph based on temporal proximity and feature similarity, which can effectively model the spatiotemporal relationships of latent features. It employs graph attention mechanism and a long short-term memory (LSTM) layer for accurate wear prediction. Finally, experimental validation on aerospace self-lubricating bearing datasets, covering full lifecycle and partial wear scenarios, demonstrates the superior accuracy and adaptability of the proposed method.**




***Index Terms*— Acoustic emission (AE), graph attention network (GAT), self-lubricating bearing, variational autoencoder (VAE), wear prediction.**

## I. Introduction

SELF-LUBRICATING bearings are essential for ensuring the safe operation of aerospace systems [1]. However, extreme operating conditions, such as high loads, oscillatory motions, and elevated temperatures, accelerate wear degradation in these bearings [2]. Such degradation increases friction, reduces efficiency, and risks mechanical failures that threaten flight safety, requiring accurate wear prediction to ensure system reliability [3]. Nevertheless, current traditional methods, often relying on manual inspections or low-sensitivity sensors, frequently fail to accurately capture wear degradation [4]. Consequently, effective wear prediction is critical to enhancing system reliability, enabling predictive maintenance, and preventing catastrophic failures in aerospace applications.

To address the shortcomings of traditional techniques, acoustic emission (AE) has emerged as a valuable nondestructive testing technique for monitoring wear in mechanical systems, including sliding bearings used in aerospace applications. AE technology captures high-frequency stress waves from material deformation, enabling precise detection of wear-related changes [5]. Its high sensitivity enables early detection of wear-related phenomena, which is essential for preventive

Received 25 November 2025; accepted 30 December 2025. Date of publication 12 January 2026; date of current version 13 February 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 92467101 and Grant 52275499 and in part by the Open Fund of National Key Laboratory of Strength and Structural Integrity under Grant LSSIKFJJ202402003. The associate editor coordinating the review of this article and approving it for publication was Dr. Dandan Peng. *(Corresponding author: Shichang Du.)*

Danyue Shen, Shichang Du, Shuo Wang, and Liang Yan are with the Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: sdy3218@sjtu.edu.cn; lovbin@sjtu.edu.cn; SJD8013@sjtu.edu.cn; Yan-Liang@sjtu.edu.cn).

Shanshan Li and Xianmin Chen are with the National Key Laboratory of Strength and the Aircraft Strength Research Institute of China (ASRI), Xi'an 310014, China (e-mail: shanshanmail2010@163.com; chenxm007@avic.com).

Digital Object Identifier 10.1109/JSEN.2025.3650493

maintenance in critical applications [6], [7]. For instance, Revill et al. [8] demonstrated the sensitivity of AE in tracking wear behavior in aerospace self-lubricating bearing liner materials, noting its potential to detect subtle changes in material composition. Similarly, König et al. [9] highlighted AE's potential to identify wear-critical conditions in sliding bearings, especially under mixed friction regimes. However, in complex aerospace environments, challenges persist as wear-related signals can be obscured, such as mechanical vibrations and environmental noise [10]. These challenges emphasize the need for advanced data-driven methods to effectively extract and interpret noisy, high-dimensional, and nonstationary AE signals.

In recent years, deep learning has transformed the analysis of such complex AE signals and become the mainstream solution [11], [12], [13], [14]. Unlike traditional machine-learning methods, which depend on manual feature engineering, deep learning uses multilayer neural networks to automatically extract latent features from high-dimensional, nonlinear data [15], [16], [17]. This method significantly reduces reliance on prior knowledge [18]. Its strength lies in decoupling noise from meaningful information, making it well-suited for processing AE signals characterized by high dimensionality and nonstationarity [19]. For example, Nashed et al. [20] developed a convolutional neural network (CNN) combined with a continuous wavelet transform (CWT) to classify gas turbine faults using AE signals. Likewise, Moradi et al. [21] integrated complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and long short-term memory (LSTM) networks. They constructed health indicators for single-stiffened composite panels under fatigue loading, leveraging AE signals to capture nonstationary damage behavior in aerospace structures. Additionally, Han et al. [22] advanced AE-based monitoring by proposing a deep CNN (DCNN) with CWT. This method identified initial damage in aeroengines using a single sensor, which can achieve feature extraction from time-frequency images in complex environments. Other studies have also shown deep learning's versatility in AE-based monitoring, such as convolutional autoencoders for crack evaluation in welded joints [23] and stacked autoencoders for AE source localization in aerospace structures [24]. These advances highlight deep learning's potential to address AE signal analysis challenges, offering improved feature representation for wear prediction in aerospace self-lubricating bearings compared to traditional methods.

In spite of these advancements, existing deep-learning methods still face significant limitations in fully capturing the complex dynamics of AE signals for wear-stage modeling. For example, Han et al. [25] noted that while LSTM networks excel in modeling temporal dependencies, they often struggle with high computational costs and gradient vanishing issues when processing high-dimensional signal sequences. Similarly, CNNs are widely used for local feature extraction. However, Cheng et al. [26] highlighted that their constrained receptive fields make it challenging to model long-term dependencies in complex time-series data. Standard variational autoencoders (VAEs) also have drawbacks. They rely on simple prior distributions that may fail to capture the intricate features in AE signals. This limits their ability to extract discriminative features for degradation states [27].

Recently, graph neural networks (GNNs) have emerged as a promising paradigm to overcome the above limitations by explicitly modeling spatial dependencies and spatiotemporal relationships [28], [29]. For instance, Zhao et al. [30] proposed a heterogeneous temporal GNN (HTGNN). This method explicitly models diverse temporal dynamics in bearing load prediction by using graph structures to integrate various signal characteristics and operating conditions for reliable virtual sensing. Similarly, Wang et al. [31] used a graph attention network (GAT) to aggregate features from related signals based on their similarity, improving the representation of complex relationships. However, many GNN-based methods adopt sequential architectures, extracting spatial features before temporal modeling, which can lead to prolonged information propagation paths and potential loss of spatiotemporal interactions, as noted by He et al. [32]. These limitations are particularly evident with AE signals, which are noisy and have complex spatiotemporal dependencies. This highlights the need for more robust feature extraction and modeling techniques adapted to the challenges of AE-based wear prediction.

The limitations of existing methods reveal a critical research gap in AE-based wear prediction for self-lubricating bearings: the absence of a unified framework that integrates high-dimensional feature extraction with robust spatiotemporal dependency modeling. Current deep-learning methods, while powerful, often struggle to balance noise robustness with comprehensive dependency capture, particularly for nonstationary AE signals [33]. Similarly, GNN-based methods, in spite of their spatial modeling capabilities, require enhanced integration of spatial and temporal features to capture the wear degradation in self-lubricating bearings [34]. To tackle this gap, this article proposes a new framework. It focuses on three critical questions: 1) How can effective wear-related features be extracted from high-dimensional, nonstationary AE signals? 2) How can a robust model be developed to predict self-lubricating bearing wear using AE signals? and 3) How can the proposed method be applied across diverse operating conditions in aerospace self-lubricating bearings? These questions are critical, as diverse aerospace operating conditions demand a robust, adaptable method. Addressing these questions is essential for ensuring reliable wear prediction under complex and dynamic environments.

To address the above challenges, this article proposes a novel wear prediction method integrating a clustering-guided contrastive VAE (CGC-VAE) and a temporal graph attention network (T-GAT). Unlike traditional VAEs, CGC-VAE innovatively combines $K$-means clustering with contrastive learning and Gaussian mixture model (GMM) regularization to extract discriminative latent features from noisy, high-dimensional AE signals. This integration enhances the model's ability to distinguish wear stages by leveraging pseudo labels and structured priors, a significant advancement over the standard VAE's uniform prior assumption. Complementing this, T-GAT constructs a dynamic graph based on temporal proximity and feature similarity, employing a multihead attention mechanism and an LSTM layer to effectively capture spatiotemporal

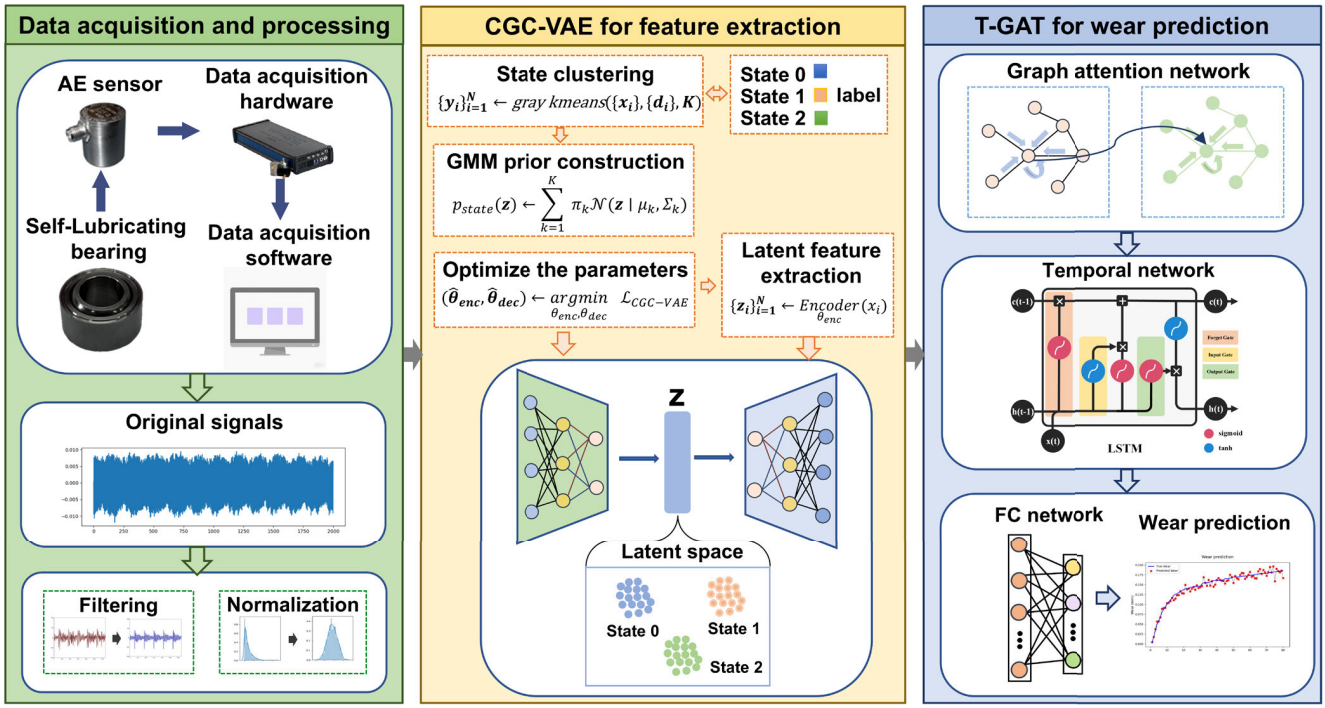



Fig. 1. Framework of the proposed method.

dependencies. This approach overcomes the limitations of sequential GNNs by reducing information propagation delays and enhancing feature aggregation.

The main contributions are summarized as follows.

1) A CGC-VAE model is developed to handle high-dimensional, noisy AE signals. It uses $K$-means clustering to segment different wear stages, and applies contrastive learning with GMM regularization to extract low-dimensional, discriminative latent features that effectively capture wear degradation patterns.
2) A T-GAT model is proposed to construct a dynamic graph based on temporal proximity and latent feature similarity. The model uses a graph attention mechanism and an LSTM unit to robustly model spatial and temporal dependencies for wear prediction.
3) The proposed method is validated on experimental datasets of aerospace self-lubricating bearings. The results demonstrate superior performance across full lifecycle and partial wear scenarios through transfer learning, outperforming baseline methods in wear prediction accuracy.

The structure of the remaining sections is as follows. Section II details the proposed wear prediction method, including AE signal preprocessing, CGC-VAE, and T-GAT models. Section III describes the experimental design for aerospace self-lubricating bearing wear testing and presents the validation results, including full lifecycle and partial wear predictions.

## II. Proposed Method

The framework of the proposed method for self-lubricating bearing wear prediction is illustrated in Fig. 1. Initially, raw AE signals are preprocessed through filtering and normalization to mitigate noise and ensure data consistency. Subsequently, the CGC-VAE extracts low-dimensional latent features, enhanced by GMM regularization and contrastive learning guided by $K$-means clustering. These features are then processed by T-GAT, which constructs a dynamic graph to capture spatiotemporal dependencies, leveraging a graph attention mechanism and an LSTM layer for wear prediction. Sections II-A–II-C elaborate on each component of this framework.

### A. Data Preprocessing

Raw AE signals, characterized by high-frequency content and susceptibility to environmental noise, undergo a preprocessing pipeline to enhance their suitability for analysis. This pipeline consists of two primary stages: filtering and normalization, addressing the nonstationary and noisy nature of the signals.

Filtering eliminates low-frequency noise components that obscure wear-related information. A high-pass filter with a cutoff frequency $f_c$ is applied, where $f_c$ is selected to preserve the frequency band associated with AE phenomena, such as friction and crack propagation, while suppressing mechanical vibrations from external sources [35]. As recommended by [8], $f_c$ is set to 100 kHz to effectively capture AE characteristics. This filtering process enhances the signal-to-noise ratio, ensuring that the filtered signal $x_f(t)$ retains critical characteristics for subsequent processing.

Following filtering, normalization standardizes the amplitude of the filtered signals across samples, facilitating subsequent model training. Min-max normalization is adopted, scaling the signal to a range [0, 1] according to the

transformation

$$x_{\text{norm}}(t) = \frac{x_f(t) - \min\left(x_f(t)\right)}{\max\left(x_f(t)\right) - \min\left(x_f(t)\right)} \tag{1}$$

where $x_{\text{norm}}(t)$ denotes the normalized signal, $x_f(t)$ indicates the filtered AE signal, and $\max(x_f(t))$ and $\min(x_f(t))$ represent the maximum and minimum values of the filtered AE signal, respectively.

### B. Clustering-Guided Contrastive VAE

Self-lubricating bearings in aerospace applications undergo a distinct wear degradation, typically divided into three stages: the running-in stage, steady stage, and severe stage [36]. The running-in stage involves initial surface adaptation with high AE activity due to material transfer and smoothing. The steady stage exhibits stable wear behavior with consistent AE signals, while the severe stage is marked by accelerated degradation and intensified AE events due to material failure. Capturing these stages is crucial for accurate wear prediction, as they reflect the evolving degradation dynamics under high-load and oscillatory conditions.

To address this, a clustering-based method is employed to segment the preprocessed AE signals into three distinct wear states. The $K$-means clustering algorithm is adopted due to its stability and efficacy in unsupervised learning scenarios, as it iteratively minimizes intracluster distances while maximizing intercluster distances. This segmentation leverages the inherent similarities in AE signal features, enabling the model to distinguish degradation states without manual intervention.

The selection of $K = 3$ as the number of wear stages is supported by domain knowledge of bearing wear mechanisms. According to [8], the wear progression in self-lubricating bearings under aerospace operating conditions is systematically divided into three phases. The running-in stage is defined by the initial contact of bearing surfaces, where micro-asperity interactions lead to elevated friction, material transfer, and plastic deformation, generating pronounced AE signals. The steady wear stage follows as surface adaptation occurs, resulting in a consistent wear rate with moderate AE activity, reflecting the effectiveness of the lubricating layer. The severe wear stage is marked by the depletion of lubrication or the onset of material fatigue, leading to rapid wear, crack propagation, and heightened AE emissions due to subsurface damage and debris formation. This tripartite classification, as proposed in [8], serves as a theoretical foundation for setting $K = 3$ in the clustering process.

VAE is a generative model that learns latent representations by optimizing the evidence lower bound (ELBO), defined as

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] - D_{\text{KL}}\left(q_\phi(z|x) \parallel p(z)\right) \tag{2}$$

where $q_\phi(z|x)$ is the encoder's approximate posterior, $p_\theta(x|z)$ is the decoder's likelihood, and $p(z) \sim \mathcal{N}(0, I)$ is the prior distribution. While VAEs effectively capture general data distributions, their ability to extract discriminative features from complex data patterns with varying degradation states is limited [37]. Building on the $K$-means segmentation, CGC-VAE is proposed to extract low-dimensional, discriminative latent features from high-dimensional AE signals. The CGC-VAE integrates a GMM prior and a contrastive learning loss, both guided by the $K$-means clustering results, to enhance feature discriminability. The structure of the CGC-VAE is illustrated in Fig. 2. The encoder, composed of 1-D convolutional layers, maps the preprocessed AE signals to latent representations $z$, outputting the mean $\mu$ and variance $\sigma$. Latent features are sampled using the reparameterization trick

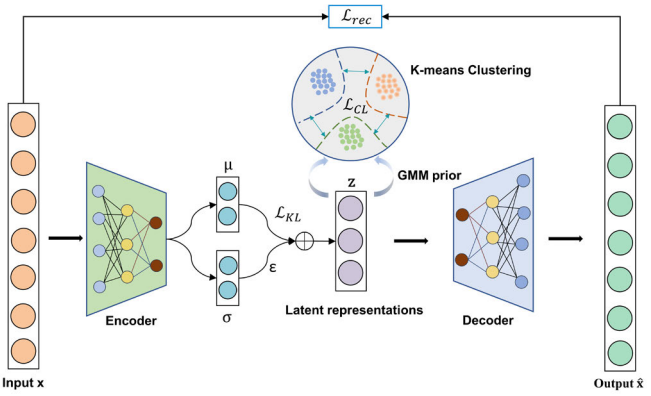



Fig. 2. Structure of CGC-VAE.

$$z = \mu + \sigma \odot \epsilon, \epsilon \sim \mathcal{N}(0, 1) \tag{3}$$

where $z_i \in \mathbb{R}^d$ represents the latent feature vector, and $d$ denotes the latent dimension. The decoder reconstructs the input as $\hat{x}$, while the latent representations $z$ are regularized by the GMM prior and contrastive learning loss to capture wear degradation patterns. The training objective balances reconstruction quality ($\mathcal{L}_{\text{rec}}$), regularization ($\mathcal{L}_{\text{KL}}$), and feature discriminability ($\mathcal{L}_{\text{CL}}$), as detailed in Sections II-B.1–II-B.3 .

*1) GMM Prior Construction:* Standard VAEs rely on a single gaussian prior, which is inadequate for capturing the complex distributions of AE signals across different wear stages. To overcome this limitation, a GMM prior is incorporated into the CGC-VAE, inspired by clustering-based VAEs [38]. The GMM prior is defined as

$$p_{\text{state}}(z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(z \mid \mu_k, \Sigma_k), \tag{4}$$

where $K = 3$ corresponds to the number of wear stages, $\pi_k$ denotes the mixing coefficient, $\mu_k$ is the mean vector, and $\Sigma_k$ is the covariance matrix of the $k$th Gaussian component. The subscript "state" reflects the prior's alignment with the wear stages.

The GMM prior is initialized using the $K$-means clustering results from the preprocessed AE signals. The clustering assigns pseudo labels $\{y_i\}_{i=1}^N$, where $y_i \in 1, 2, 3$ denotes the wear stage of sample $i$. The GMM parameters are computed as follows.

1) $\pi_k = N_k/N$, where $N_k = \sum_{i=1}^N 1(y_i = k)$ is the number of samples in the $K$th cluster, and $N$ is the total number of samples.
2) $\mu_k = (1/N_k) \cdot \sum_{i:y_i=k} z_i$, the mean of latent features in the $K$th cluster.

3) $\Sigma_k = (1/N_k)\cdot\sum_{i:y_i-k}(z_i-\mu_k)(z_i-\mu_k)^\top$, the covariance of latent features in the $K$th cluster.

The latent features $z_i$ are obtained from an initial encoder pass. This GMM prior captures the complex distribution of AE signals across the wear stages, facilitating subsequent contrastive learning.

*2) Contrastive Learning:* Contrastive learning is a self-supervised learning technique that learns representations by contrasting positive pairs (similar examples) against negative pairs (dissimilar examples), showing great promise in various domains [39]. In the context of unsupervised feature learning for complex sequential data, contrastive learning enhances the discriminability of latent features by leveraging the pseudo labels from $K$-means clustering. In this framework, contrastive learning encourages features within the same wear stage to be similar while separating those from different stages, capturing wear degradation patterns effectively. Inspired by InfoNCE [40], the contrastive learning loss is defined as

$$\mathcal{L}_{\mathrm{CL}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\sum_{j:y_i=y_j}\exp\left(\mathrm{sim}\left(z_i, z_j\right)/\tau\right)}{\sum_{k=1}^{N}\exp\left(\mathrm{sim}\left(z_i, z_k\right)/\tau\right)} \tag{5}$$

where $\mathrm{sim}(z_i, z_j)$ is the cosine similarity, $\tau$ is a temperature parameter controlling the softness of the distribution, and $y_i, y_j \in 1, 2, 3$ are the cluster labels corresponding to the wear stages. The value of $\tau$ is chosen based on contrastive learning theory, which indicates that a lower $\tau$ sharpens the similarity distribution, aiding in the separation of wear stages. $L_{\mathrm{CL}}$ maximizes the similarity of the same cluster while minimizing the similarity of different clusters, enhancing the robustness of the extracted features.

*3) Loss Function and VAE Framework:* The CGC-VAE integrates the GMM prior and contrastive learning within a variational framework, optimizing a combined loss function. Three loss components are defined to balance reconstruction, regularization, and feature discriminability.

First, the reconstruction loss measures the difference between the input data $x$ and its reconstruction $\hat{x}$, ensuring the decoder captures the input's structural patterns

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{N\cdot L}\sum_{i=1}^{N}\sum_{j=1}^{L}\left(x_{i,j}-\hat{x}_{i,j}\right)^2 \tag{6}$$

where $x_{i,j}$ and $\hat{x}_{i,j}$ are the original and reconstructed signal values for sample $i$ at time step $j$, with $N$ samples and signal length $L$.

Next, the KL divergence regularizes the encoder's approximate posterior $q_\phi(z|x)$ to align with the GMM prior $p_{\mathrm{state}}(z)$

$$\mathcal{L}_{\mathrm{KL}} = D_{\mathrm{KL}}\left(q_\phi\left(z|x\right)\left\|\, p_{\mathrm{state}}\left(z\right)\right.\right) \tag{7}$$

where $q_\phi(z|x) = \mathcal{N}(z|\mu_\phi(x), \Sigma_\phi(x))$ is the encoder's output distribution. The contrastive learning loss $\mathcal{L}_{\mathrm{CL}}$ is as defined in (5). The total loss is

$$\mathcal{L}_{\mathrm{CGC-VAE}} = \mathcal{L}_{\mathrm{rec}} + \beta\cdot\mathcal{L}_{\mathrm{KL}} + \gamma\cdot\mathcal{L}_{\mathrm{CL}} \tag{8}$$

where $\beta$ and $\gamma$ are hyperparameters balancing the trade-off between reconstruction, regularization, and feature discriminability.

The CGC-VAE is trained by minimizing the total loss $L_{\mathrm{CGC-VAE}}$, as outlined in Algorithm 1. The GMM prior parameters are updated iteratively using the expectation-maximization (EM) algorithm, ensuring the prior adapts to the evolving latent representations. During inference, the encoder outputs discriminative latent features $\{z_i\}_{i=1}^N$, which are used for downstream wear prediction tasks.

**Algorithm 1** Training Procedure of CGC-VAE

**Input**: AE signals: $\{x_i\}_{i=1}^N, x_i \in \mathbb{R}^N$, wear depth data: $\{d_i\}_{i=1}^N$
**Output**: Latent features: $\{z_i\}_{i=1}^N, z_i \in \mathbb{R}^d$
**Parameter**: number of wear states $K$, temperature coefficient $\tau$, latent dimension $d$, number of iterations $T$, hyperparameters $\beta$, $\gamma$
1: Preprocess wear states: $\{y_i\}_{i=1}^N \leftarrow K\_\mathrm{means}(\{x_i\}, \{d_i\}, K)$
2: Construct GMM prior according to (4)
3: Initialize VAE parameters: initialize encoder $\theta_{enc}$ and decoder $\theta_{dec}$
4: Train CGC-VAE:
**for** $t = 1$ to $T$ **do**
5: Calculate $\mathcal{L}_{\mathrm{CL}}, \mathcal{L}_{\mathrm{rec}}, \mathcal{L}_{\mathrm{KL}}$ according to (5),(6) and (7).
6: Update parameter: $\hat{\theta}_{enc}, \hat{\theta}_{dec}$
$$(\hat{\theta}_{enc}, \hat{\theta}_{dec}) \leftarrow \underset{\theta_{enc},\theta_{dec}}{argmin}\ \mathcal{L}_{\mathrm{rec}} + \beta\,\mathcal{L}_{\mathrm{KL}} + \gamma\,\mathcal{L}_{\mathrm{CL}}$$
**end**
7: Extract latent features: $\{z_i\}_{i=1}^N \leftarrow \underset{\theta_{enc}}{Encoder}(x_i)$
8: Return $\{z_i\}_{i=1}^N$

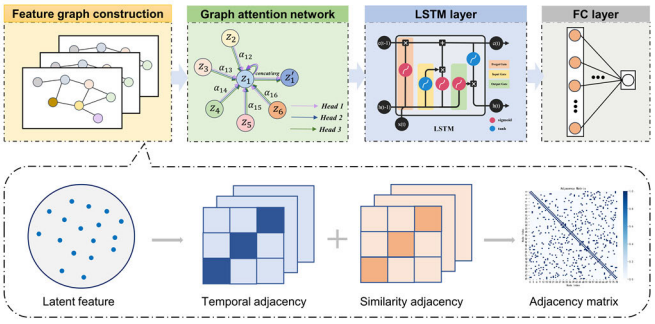



Fig. 3. Structure of T-GAT.

## C. Temporal Graph Attention Network

In this section, the T-GAT is developed to model the spatiotemporal dependencies of the latent features extracted by the CGC-VAE. As shown in Fig. 3, the latent features are first used to construct a dynamic graph capturing spatial and temporal relationships. The graph attention network then aggregates spatial information, followed by an LSTM layer to model temporal dynamics, and a fully connected layer to produce the final prediction.Sections II-C.1–II-C.4 elaborate on each component.

*1) Feature Graph Construction:* The latent features extracted by CGC-VAE, denoted as $\{z_i\}_{i=1}^N$, where $z_i \in \mathbb{R}^d$ and $N$ is the number of samples, are used to construct a graph $\mathcal{G} = (V, E, A, X)$. Here, $V$ represents the set of nodes with $|V| = N$, $E$ denotes the edge set, $A \in \mathbb{R}^{N\times N}$ is the adjacency matrix encoding node connectivity, and $X \in \mathbb{R}^{N\times D}$ contains the node features. The adjacency matrix $A$ is designed to integrate temporal proximity and feature similarity, capturing the spatiotemporal dynamics of the data.

Temporal adjacency, denoted as $A_{\mathrm{time}}$, is established by connecting consecutive samples in the sequence to capture

temporal evolution

$$A_{\text{time}}(i, i+1) = A_{\text{time}}(i+1, i) = 1 \tag{9}$$

where $A_{\text{time}}(i, j) = 0$ otherwise. This configuration ensures that the graph preserves the sequential nature of the degradation process.

Similarity adjacency, denoted as $A_{\text{sim}}$, is constructed using the $K$-nearest neighbors (KNNs) method based on cosine similarity between latent features, quantifying the structural resemblance between samples. The cosine similarity is computed as

$$\text{sim}\left(z_i, z_j\right) = \frac{z_i \cdot z_j}{\|z_i\| \|z_j\|} \tag{10}$$

where $\|\cdot\|$ represents the Euclidean norm. For each node $i$, the cosine similarity $\text{sim}(z_i, z_j)$ is calculated for all nodes $j$, sorted in descending order, and the $K$ most similar neighbors are selected to form the similarity neighbor set $S_i(K)$. This set is defined as

$$S_i(K) = \left\{ j \,|\, j \in \text{Top-}K\left(\{\text{sim}(z_i, z_k)\}_{k=1}^{N-1}\right)\right\} \tag{11}$$

where Top-$K(\{\text{sim}(z_i, z_k)\}_{k=1}^{N-1})$ denotes the set of indices corresponding to the $K$ largest values of $\text{sim}(z_i, z_k)$ after sorting in descending order. The similarity adjacency matrix is then defined as

$$A_{\text{sim}}(i, j) = \begin{cases} 1, & \text{if } j \in S_i(K) \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

The final adjacency matrix integrates both temporal and similarity relationships

$$A(i, j) = \begin{cases} 1, & \text{if } A_{\text{time}}(i, j) = 1 \text{ or } A_{\text{sim}}(i, j) = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

*2) Graph Attention Network:* Graphs provide an efficient representation of latent features as nodes, with edges capturing their spatial and temporal relationships. To effectively process this graph-structured data and fuse information, the GAT is employed. Introduced by Veličković et al. [41], GAT leverages self-attention mechanisms to capture spatial dependencies within the graph, assigning attention weights to neighboring nodes to emphasize the most salient features.

The core of GAT is the graph attention layer (GAL), which computes node feature representations by focusing on influential neighbors. The neighborhood set $\mathcal{N}_i = \{j | A(i, j) = 1\}$ is defined based on the adjacency matrix $A$. The input features for node $i$ are denoted as $z_i \in \mathbb{R}^d$. These features are linearly transformed using a shared weight matrix $W \in \mathbb{R}^{d' \times d}$, where $d'$ is the reduced dimension. To enhance the stability of the attention learning process, a multihead attention mechanism with $H$ heads is adopted. Each head independently computes the unnormalized attention score $e_{ij}^h$ between nodes $i$ and $j$ as

$$e_{ij}^h = \text{LeakyReLU}\left(\left(\alpha_h^\top \cdot \left[W z_i \| W z_j\right]\right)\right) \tag{14}$$

where $\alpha_h \in \mathbb{R}^{2d'}$ is a learnable attention vector for the $h$th head, and $\|$ denotes concatenation. A mask ensures attention is applied only to connected nodes

$$\tilde{e}_{ij}^h = \begin{cases} e_{ij}^h, & \text{if } A(i, j) = 1 \\ -\infty, & \text{otherwise} \end{cases} \tag{15}$$

and the normalized attention coefficients are obtained via the Softmax function

$$\alpha_{ij} = \text{Softmax}_j\left(e_{ij}\right) = \frac{\exp\left(\tilde{e}_{ij}\right)}{\sum_{k \in \text{N}_i} \exp\left(\tilde{e}_{ik}\right)} \tag{16}$$

where $\mathcal{N}_i = \{j \mid A(i, j) = 1\}$ is the neighborhood of node $i$. The updated feature for node at the $h$th head is computed as

$$z_i^h = \sigma\left(\sum\nolimits_{j \in \mathcal{N}_i} \alpha_{ij}^h W z_j\right) \tag{17}$$

where $\sigma(\cdot)$ is a nonlinear activation function. The outputs from all $H$ heads are aggregated to produce the final output

$$\text{z}_i' = \frac{1}{H} \sum\nolimits_{h=1}^{H} \text{z}_i^h. \tag{18}$$

*3) Temporal Modeling Layer:* After the above processing, the latent features are transformed into time-series representations. Effectively capturing the temporal dynamics of these sequential data is essential for accurate wear prediction. To address this, an LSTM layer is integrated into the temporal modeling block of the T-GAT. LSTM is widely recognized as one of the most effective variants of recurrent neural networks for time-series analysis, designed to address the vanishing gradient problem in long-term prediction [42], [43]. The LSTM unit, as illustrated in Fig. 3, consists of a memory cell and a set of gates, including an input gate, a forget gate, and an output gate. The memory cell stores and passes memory information across different time steps, while the gates regulate the addition, removal, or output of this information, enabling the model to capture long-term dependencies in the wear degradation.

The output of the GAL, $z_i' \in \mathbb{R}^{d'}$, forms a sequence $\{z_i'\}_{i=1}^N$ that encodes spatially aggregated features over time. This sequence is fed into the LSTM layer to model the temporal evolution of wear degradation patterns in the AE signals. The detailed computations of the LSTM unit are given by

$$\begin{cases} i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \\ f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \\ g_t = \tan h\left(W_g x_t + U_g h_{t-1} + b_g\right) \\ c_t = f_t \odot c_{t-1} + i_t \odot g_t \\ h_t = o_t \odot \tan h\left(c_t\right) \end{cases} \tag{19}$$

where $i_t$, $f_t$, and $o_t$ are the activation vectors of the input, forget, and output gates, respectively, $g_t$ is the cell input activation vector, $c_t$ is the cell state vector, and $h_t$ is the hidden state vector at time step $t$. Here, $x_t = z_t' \in \mathbb{R}^{d'}$ is the input feature, $W_i, W_f, W_o, W_g$ and $U_i, U_f, U_o, U_g$ are the weight matrices for input and recurrent connections, respectively, $b_i, b_f, b_o, b_g$ are bias vectors. The LSTM processes the sequence $\{z_i'\}_{i=1}^N$ over $N$ time steps, producing a sequence of hidden states $\{h_t\}_{t=1}^N$. The final hidden state $h_N$, encapsulates the temporal

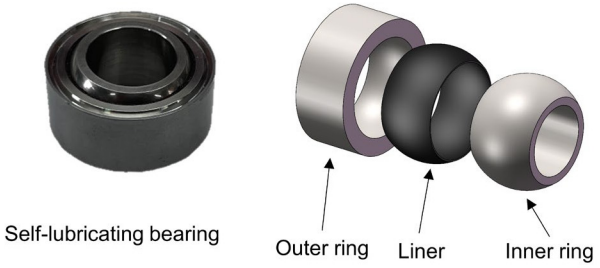



Fig. 4. Self-lubricating bearing structure.

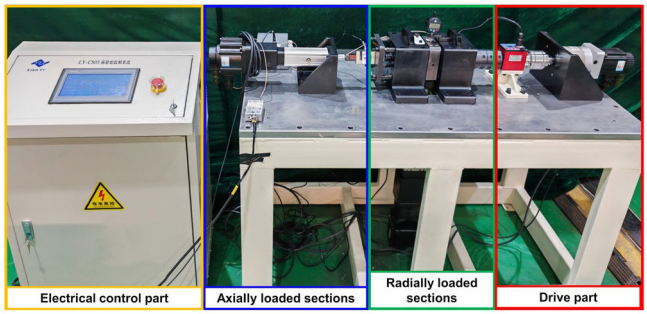



Fig. 5. Self-lubricating bearing experimental platform.

dynamics of the wear process, integrating both past and current degradation patterns.

*4) Prediction Layer:* Following the temporal modeling, a fully connected layer is employed as the output layer to perform a linear transformation on $h_N$. The predicted wear value is computed as

$$\hat{y} = W_p \cdot h_N + b_p \tag{20}$$

where $W_p$ and $b_p$ are the weight and bias of the fully connected layer, and $\hat{y}$ represents the predicted wear value.

## III. Experimental Verification

### A. Experimental Setup and Dataset

The experimental study focuses on a self-lubricating bearing designed with precise geometric specifications, tailored for monitoring wear under simulated operational conditions. Fig. 4 shows the self-lubricating bearing structure, highlighting its key components, including the outer ring, inner ring, and liner. The bearing consists of an outer ring with an outer diameter of 36.512 mm and a width of 19.05 mm, paired with an inner ring having an inner diameter of 19.05 mm, an outer diameter of 23.37 mm, and a width of 15.06 mm. The bearing features a spherical outer raceway with a curvature of 0.254 mm and supports an angular misalignment capability of up to 8°, ensuring reliable performance under varying loads and oscillatory motions.

The experimental setup for evaluating the wear performance of the self-lubricating bearing is a custom-designed test rig, as depicted in Fig. 5. The test platform integrates four primary systems: a driving system, a loading system, a control system, and a measurement system. The driving system employs a servo motor (Delta ECMA-J11010RS) coupled with a 35:1 reduction gearbox to enable reciprocating oscillation of the

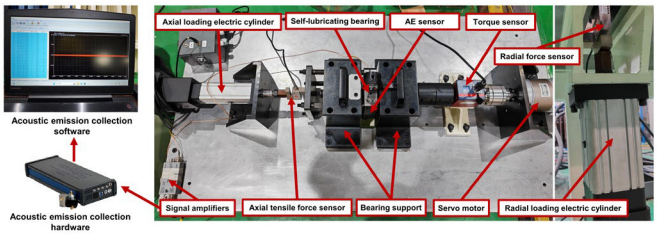



Fig. 6. Experimental setup for wear testing with AE signals.

bearing, simulating real-world operational dynamics. The loading system facilitates both radial and axial force application through servo-electric cylinders. In the axial direction, loading is achieved by connecting a loading plate to the bearing housing via a long screw, with tension and pressure transmitted through an electric cylinder, and a force transducer measures the axial loading force. In the radial direction, the loading module is linked to a servo-electric cylinder via a loading actuator, with the cylinder mounted on a bracket secured to a cast iron base using connecting rods and bolts. The control system, managed by an industrial computer and servo drivers integrated into a unified control cabinet, ensures accurate regulation of motion and load parameters via a programmable logic controller (PLC). The measurement system is equipped with multiple sensors to capture critical parameters: a torque sensor monitors torque variations, a temperature sensor tracks thermal changes, and AE, axial, and radial force sensors collect mechanical responses for bearing health evaluation. Specifically, the AE sensor is directly mounted on the outer ring of the self-lubricating bearing to capture the AE signals generated during wear degradation with high sensitivity.

Fig. 6 presents the detailed schematic of the experimental setup for wear testing, highlighting the integration of key components. The experimental procedure begins with mounting the bearing onto the bearing housing, ensuring proper alignment for axial and radial measurements. Predefined radial and axial loads, set to replicate real-world bearing stresses, are applied using servo-electric cylinders. The driving system induces reciprocating oscillation at a specified frequency, controlled by the servo motor and gearbox, to simulate cyclic motion. Throughout the experiment, the measurement system continuously records essential parameters, including torque, axial and radial forces, temperature, and AE signals. The AE signals, which capture bearing mechanical responses under loading and motion, are collected at a 2 MHz sampling frequency. Among these, the AE signals are acquired by an AE sensor using 16-bit AD sampling, amplified with a 40 dB gain by a preamplifier, and digitized by AE hardware before being processed using AE software for subsequent analysis.

In this section, the proposed method is validated using experimental data from three wear prediction tasks, as detailed in Table I. Task 1 involves a complete wear cycle of bearing 1–1 and bearing 1–2, each with 150 000 wear cycles, where bearing 1–1 serves as the labeled training set with wear progression until failure, and bearing 1–2 acts as the unlabeled test set to validate the model's performance under identical experimental conditions. Task 2 and task 3, involving

TABLE I
DETAILED INFORMATION OF THE DATASET

| | Task 1 | Task 2 | Task 3 |
|---|---|---|---|
| Bearing type | Self-lubricating bearing | | |
| Bearing | Bearing 1-1, Bearing 1-2 | Bearing 2 | Bearing 3 |
| Training set | Labeled Bearing 1-1 | | |
| Test set | Unlabeled Bearing 1-2 | Labeled Bearing 2 | Labeled Bearing 3 |
| Wear Cycles | 150000 | 50000 | 60000 |

TABLE II
ARCHITECTURE OF T-GAT

| Layer | Input dimension | Output dimension | Number of heads | Activation function |
|---|---|---|---|---|
| GAT 1 | 256 | 128 | 8 | ELU |
| GAT 2 | 128 | 128 | 8 | ELU |
| LSTM | 128 | 64 | – | Tanh/Sigmoid |
| Fully connected | 64 | 1 | – | Sigmoid |

bearing 2 and bearing 3, respectively, contain partial wear data with 50 000 and 60 000 wear cycles, and are used to evaluate the model's ability to predict the wear failure time through transfer learning. The validation process includes training the model on task 1 data, transferring the trained model to task 2 and task 3 with limited data.

### B. Model Training and Prediction on Task 1

*1) Model Training:* The full wear cycle data of task 1 is divided into training data (bearing 1–1) and testing data (bearing 1–2). The proposed method is first trained on bearing 1–1 and validated on bearing 1–2 to evaluate its performance in wear prediction under a complete lifecycle scenario. The training process involves several key steps.

First, the raw AE signals are preprocessed through high-pass filtering and normalization to mitigate noise interference, as described in Section II. The preprocessed signals are then input into the CGC-VAE, which employs $K$-means clustering to segment wear stages, enhancing feature discriminability via contrastive learning.

Next, the latent features extracted by CGC-VAE are directly fed into the T-GAT model, which constructs a dynamic graph structure to capture the spatiotemporal dependencies of the AE signals. To enhance the model's capability, a multihead attention mechanism with $H$ heads is adopted for the GALs. The architecture of T-GAT consists of two GAT layers to model spatial dependencies, one LSTM layer to capture temporal dynamics, and one fully connected layer for final wear prediction. The GALs compute attention coefficients to weigh the importance of neighboring nodes, while the LSTM layer models the temporal evolution of the wear degradation, enabling the model to effectively integrate both spatial and temporal information. The detailed structure of the T-GAT model is summarized in Table II.

The selection of $H = 8$, as reflected in Table II, is chosen based on its ability to process diverse feature subspaces in parallel, capturing complex spatiotemporal dependencies in AE signals effectively. Finally, a fully connected layer transforms the aggregated temporal features from the LSTM into the predicted wear value. The model is optimized using the Adam optimizer with an initial learning rate of 0.001 over 200 iterations, employing the mean absolute error (MAE) as the loss function to minimize the discrepancy between predicted and actual wear values. The model training results are shown in Fig. 7.

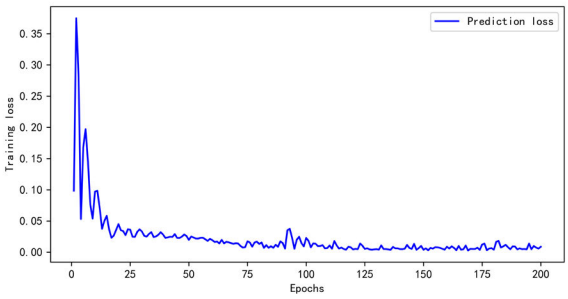



Fig. 7. Convergence of the training loss on task 1.

*2) Compared Methods and Evaluation Metrics:* To evaluate the effectiveness of the proposed method, five baseline methods are introduced for comparative experiments.

*a) SVR:* This method uses support vector regression on statistical features of preprocessed AE signals, employing an RBF kernel, with parameters $C$ and gamma optimized via grid search for wear prediction.

*b) CNN + LSTM:* This method uses CNN to extract features from the AE signals, followed by a two-layer LSTM network for wear prediction. The CNN consists of six convolutional layers and a fully connected layer to map the features into the latent space. The LSTM model captures temporal dependencies but ignores spatial relationships.

*c) CNN + T-GAT:* This method also employs a CNN to extract features from the AE signals, which are then fed into the T-GAT model for wear prediction. The T-GAT model leverages both spatial and temporal dependencies through its graph attention and LSTM layers.

*d) VAE + T-GAT:* This method replaces CGC-VAE with a standard VAE to extract latent features, which are then fed into the same T-GAT model for wear prediction.

*e) HTGNN:* This method adopts the HTGNN framework [30], which models preprocessed AE signals as distinct node types within a graph structure to capture their spatiotemporal relationships.

The performance of all methods is assessed using two widely adopted metrics: MAE and root mean squared error (RMSE). These are defined as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{21}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{22}$$

where $N$ denotes the total number of samples, $\hat{y}_i$ is the predicted wear value, and $y_i$ is the actual wear value. Notably, a smaller value for both MAE and RMSE indicates superior model performance.

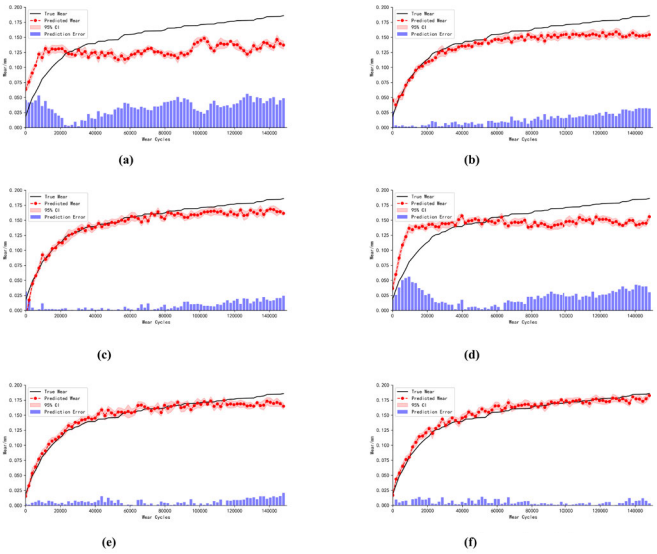


Fig. 8. Wear prediction results of task 1. (a) SVR. (b) CNN + LSTM. (c) CNN + T-GAT. (d) VAE + T-GAT. (e) HTGNN. (f) Proposed method.

TABLE III
WEAR PREDICTION RESULTS ON TASK1 TEST DATASET

| Method | MAE | RMSE |
|---|---|---|
| SVR | 0.03854 | 0.04738 |
| CNN + LSTM | 0.01403 | 0.01672 |
| CNN + T-GAT | 0.00758 | 0.00965 |
| VAE + T-GAT | 0.02993 | 0.03446 |
| HTGNN | 0.00685 | 0.00921 |
| **Proposed method** | **0.00423** | **0.00527** |

*3) Results and Analysis:* The wear prediction results for all methods over the 150 000 wear cycles of task 1 are shown in Fig. 8. Each subplot compares the predicted wear amount with the true wear amount, along with a 95% confidence interval (CI) for the predictions and a histogram of prediction errors.

The quantitative results on the test dataset of task 1 are summarized in Table III, showing the MAE and RMSE values for wear prediction. The proposed method achieves the lowest MAE and RMSE, outperforming all baseline methods. From Table III and Fig. 8, several key observations can be made.

*a) SVR [Fig. 8(a)]:* This method demonstrates the highest deviations in wear prediction. These deviations are attributed to its dependence on statistical features and the absence of a mechanism to model temporal dynamics, which restricts its capability to accurately represent the complex wear degradation patterns.

*b) CNN + LSTM [Fig. 8(b)]:* This method shows significant deviations between the predicted and true wear values, particularly in the later stages (after 80 000 cycles). The predicted wear consistently underestimates the true wear in the later stages, indicating that the CNN + LSTM method struggles to capture the complex wear degradation patterns, likely due to its inability to model spatial dependencies.

*c) CNN + T-GAT [Fig. 8(c)]:* The performance improves compared to CNN + LSTM, with the predicted wear values generally closer to the true wear across the wear cycles. However, the method still exhibits noticeable deviations in the later stages, where the predicted wear slightly underestimates the true wear. This suggests that while T-GAT's ability to model spatial dependencies helps, the CNN's feature extraction lacks the robustness of CGC-VAE, limiting overall accuracy.

*d) VAE + T-GAT [Fig. 8(d)]:* This method shows larger deviations compared to CNN + T-GAT, with the predicted wear often underestimating the true wear, especially in the early and later stages (before 30 000 and after 80 000 cycles). The VAE's lack of contrastive learning and GMM regularization likely results in less discriminative features, leading to poorer wear prediction performance in spite of the use of T-GAT for spatiotemporal modeling.

*e) HTGNN [Fig. 8(e)]:* This method exhibits predicted wear values with a strong correlation to the true wear. Nevertheless, minor deviations persist, suggesting that while HTGNN effectively captures spatiotemporal relationships, its performance is marginally surpassed by the proposed method, potentially due to differences in modeling AE signals.

Compared to the above models, the proposed method [Fig. 8(f)] demonstrates the highest prediction accuracy, with the predicted wear values closely matching the true wear across the entire wear cycle. It excels particularly in the stable wear growth region (60 000–130 000 cycles), where the predicted wear aligns almost perfectly with the true wear trend. The combination of CGC-VAE's contrastive learning and GMM regularization enables effective feature extraction from AE signals. Additionally, T-GAT's ability to model both spatial and temporal dependencies allows the model to capture the complex wear degradation patterns more accurately than the baseline methods.

## C. Wear Prediction for Tasks 2 and 3 Using Transfer Learning

This section applies the model validated in task 1 to predict the wear failure cycles of bearings 2 and 3 in tasks 2 and 3, respectively, through transfer learning. Task 1 encompasses a complete wear cycle, including running-in, steady, and severe stages, whereas tasks 2 and 3 contain partial wear data from early operational stages (50 000 and 60 000 cycles, respectively, as shown in Table I), covering only the running-in and steady stages without reaching the failure threshold. The objective is to adapt the pretrained model to these new bearings and estimate the cycles at which wear failure occurs, simulating an online prediction scenario with limited data.

*1) Transfer Learning Strategy:* The transfer learning method uses the CGC-VAE and T-GAT models pretrained on task 1. Given that tasks 2 and 3 lack the severe wear stage, the feature distribution of their AE signals differs from that of task 1. To address this discrepancy, the CGC-VAE is fine-tuned to adapt to the partial wear data. Specifically, the encoder and decoder of the CGC-VAE, pretrained on task 1, are partially frozen to preserve general feature extraction capabilities, while $K$-means clustering is re-applied to the preprocessed AE signals of tasks 2 and 3, setting $K = 2$ to segment the running-in and steady stages. Subsequently, the GMM prior is updated with two gaussian components based on the new clustering results, ensuring the latent feature space aligns with the data distribution of tasks 2 and 3.

TABLE IV
WEAR PREDICTION RESULTS FOR FINE-TUNING STRATEGIES ON TASKS 2 AND 3

| Strategy | Task 2 | | | Task 3 | | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | Failure cycles | MAE | RMSE | Failure cycles |
| FC | 0.01552 | 0.01981 | 158300 | 0.01895 | 0.02124 | 159500 |
| FC+1G | 0.01215 | 0.01526 | 152600 | 0.01156 | 0.01448 | 151500 |
| FC+2G | 0.01357 | 0.01615 | 155000 | 0.01491 | 0.01916 | 156500 |
| Full | 0.02123 | 0.02656 | 161400 | 0.02867 | 0.03583 | 165200 |

The T-GAT model is then fine-tuned using all available data from tasks 2 and 3. To optimize adaptation while mitigating overfitting, four fine-tuning strategies are evaluated.

1) FC, where only the final fully connected layer is fine-tuned.
2) FC + 1G, where the FC layer and the last GAT layer are fine-tuned.
3) FC + 2G, where the FC layer and both GAT layers are fine-tuned.
4) Full, where the entire T-GAT model, including GAT, LSTM, and FC layer, is fine-tuned.

Fine-tuning is performed over 50 epochs using the Adam optimizer with a learning rate of 0.0005, adjusted from task 1 to account for the limited data. The loss function remains MAE, consistent with task 1.

*2) Wear Prediction Results:* Using the fine-tuned T-GAT model and features extracted by the adjusted CGC-VAE, wear degradation is predicted beyond the labeled data until the wear reaches the failure threshold, estimating the failure cycles for bearings 2 and 3. The performance of four fine-tuning strategies is assessed by computing MAE and RMSE on the labeled data and analyzing the predicted failure cycles, with results summarized in Table IV.

It can be seen from Table IV that the FC+1G strategy achieves the best performance, with an MAE of 0.01215 for task 2 and 0.01156 for task 3. This strategy effectively balances adaptability and generalization by fine-tuning only the final FC layer and the last GAT layer, preserving the pretrained spatial and temporal features while adapting to the new data. In contrast, the FC strategy, with limited adaptation, yields higher errors, while the FC+2G strategy performs slightly worse than FC+1G due to overfitting risks from fine-tuning both GAT layers. The full strategy results in the highest errors due to excessive fine-tuning across all layers, including the LSTM, leading to overfitting. Across all strategies, task 3 consistently outperforms task 2, which can be attributed to its larger dataset, providing more representative data for fine-tuning, as well as the adjusted CGC-VAE's improved feature alignment with the partial wear stages. Using the FC+1G strategy, the model predicts failure at 152 600 cycles for bearing 2 and 151 500 cycles for bearing 3, with task 3's lower MAE suggesting greater reliability in its failure cycle prediction. The predicted wear curves, along with 95% CIs, are presented in Fig. 9. The CIs indicate moderate uncertainty in the failure cycle predictions, primarily due to the limited data coverage, which lacks the severe wear stage, and the relatively small dataset sizes in tasks 2 and 3. In summary, the improved performance of task 3 reflects the effectiveness of the fine-tuning process, supported by the adjusted CGC-VAE and the larger dataset.

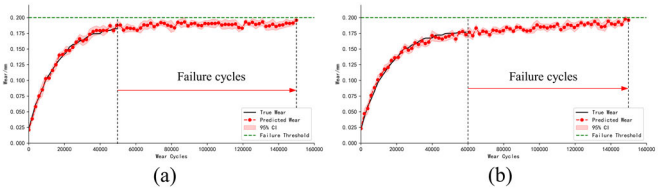



Fig. 9. Wear prediction results using the FC+1G fine-tuning strategy. (a) Task 2 and (b) Task 3.

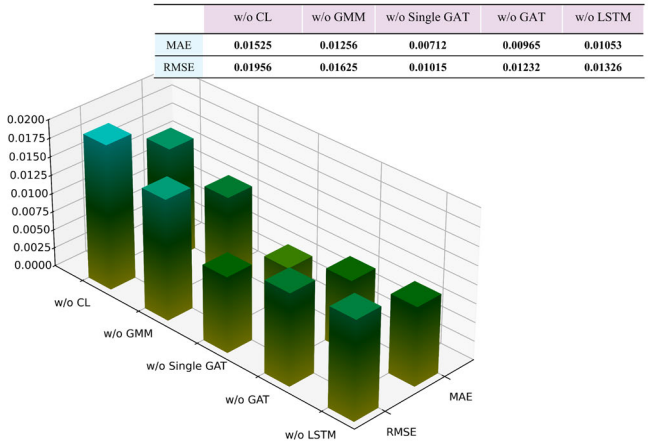



Fig. 10. Comparison of prediction performance in ablation experiments on task 1.

## D. Ablation Analysis

To further assess the effectiveness and importance of each module in the proposed method, several variant models are developed by removing key components and assessed using the wear cycle data of bearing 1. Specifically, the models are trained on the full wear cycle data of bearing 1–1 and tested on bearing 1–2, with performance evaluated using MAE and RMSE as metrics. The variant models are defined as follows.

1) *w/o CL:* The contrastive learning loss in CGC-VAE is removed ($\gamma = 0$).
2) *w/o GMM:* The GMM regularization in CGC-VAE is removed ($\beta = 0$), using standard KL divergence instead.
3) *w/o Single GAT:* The T-GAT model uses only one GAT layer instead of two.
4) *w/o GAT:* All GAT layers in T-GAT are removed.
5) *w/o LSTM:* The LSTM layer in T-GAT is removed.

The prediction performance of these variants is presented in Fig. 10, which illustrates the impact of removing each component on MAE and RMSE. The w/o CL variant exhibits the most pronounced increase in error metrics, indicating that contrastive learning plays a crucial role in enhancing the discriminability of latent features within the CGC-VAE. The w/o GMM variant shows a noticeable rise in errors, suggesting that GMM regularization is vital for maintaining a structured latent distribution. When comparing the w/o Single GAT and w/o GAT variants, both demonstrate a clear degradation in

TABLE V
SENSITIVITY ANALYSIS FOR PARAMETER $\tau$, $\beta$, AND $\gamma$ ON TASK1

| (a) The impact of parameter $\tau$ on model performance | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Metric | 0.03 | 0.05 | 0.07 | 0.10 | 0.15 | 0.20 | 0.30 | 0.50 | | |
| MAE | 0.01025 | 0.00726 | 0.00795 | 0.00526 | 0.00671 | 0.00667 | 0.01385 | 0.01648 | | |
| RMSE | 0.01413 | 0.00894 | 0.00983 | 0.00714 | 0.00916 | 0.00935 | 0.01876 | 0.02161 | | |
| (b) The impact of parameter $\beta$ on model performance | | | | | | | | | | |
| Metric | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| MAE | 0.01182 | 0.00976 | 0.00768 | 0.00493 | 0.00523 | 0.00665 | 0.00812 | 0.01167 | 0.01529 | 0.01498 |
| RMSE | 0.01495 | 0.01281 | 0.01029 | 0.00584 | 0.00627 | 0.00882 | 0.01138 | 0.01504 | 0.01976 | 0.01753 |
| (c) The impact of parameter $\gamma$ on model performance | | | | | | | | | | |
| Metric | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| MAE | 0.01235 | 0.01028 | 0.00819 | 0.00642 | 0.00623 | 0.00587 | 0.00636 | 0.00902 | 0.01275 | 0.01353 |
| RMSE | 0.01562 | 0.01347 | 0.01083 | 0.00856 | 0.00927 | 0.00713 | 0.00878 | 0.01245 | 0.01519 | 0.01502 |

performance, highlighting the importance of multiple GAT layers in capturing spatial dependencies among AE signals. Similarly, the w/o LSTM variant reveals a substantial decline in accuracy, underscoring the essential contribution of the LSTM layer in modeling temporal dynamics and ensuring precise tracking of wear degradation over time. In summary, these results demonstrate that each module significantly contributes to the overall predictive capability of the proposed method.

## E. Parameter Sensitivity Analysis

*1) Analysis of Hyperparameters $\tau$, $\beta$, and $\gamma$:* The CGC-VAE loss function contains three key hyperparameters: the GMM regularization weight $\beta$, the contrastive learning weight $\gamma$, and the temperature parameter $\tau$ in the contrastive loss. The value of $\tau$ was tested in the set {0.03, 0.05, 0.07, 0.10, 0.15, 0.20, 0.30, 0.50}, a range commonly used in contrastive representation learning for time-series tasks [39]. The weights $\beta$ and $\gamma$ were determined by grid search over [0.1, 1.0] with a step size of 0.1. The performance is evaluated on task 1 using MAE and RMSE, and the results are summarized in Table V.

Table V summarizes the sensitivity of the three hyperparameters. The lowest prediction errors are obtained at $\tau = 0.10$, $\beta = 0.4$, and $\gamma = 0.6$. For the temperature parameter $\tau$, the performance is best near 0.10 and degrades when $\tau$ is either significantly lower or higher than this value, which is consistent with the widely observed behavior in contrastive learning for time-series data. For $\beta$, errors decrease from 0.1 to 0.4 and increase beyond 0.4, indicating that moderate GMM regularization effectively aligns the latent posterior with the multimodal distribution of wear stages without over-constraining the representation. Similarly, for $\gamma$, errors decrease from 0.1 to 0.6 and rise thereafter, showing that an appropriate contrastive weight enhances intrastage cohesion and interstage separation while preserving reconstruction fidelity.

Theoretically, $\beta$ controls the strength of the KL regularization toward the GMM prior, balancing latent structure and flexibility, whereas $\gamma$ governs the contribution of the clustering-guided contrastive loss. The selected values $\tau = 0.10$, $\beta = 0.4$, and $\gamma = 0.6$ achieve the best trade-off among reconstruction accuracy, latent distribution alignment, and feature discriminability.

*2) Analysis of the Number of Adjacent Nodes K:* In the T-GAT model, the graph structure is constructed using a KNN method, where each sample node is connected to its $K$ most similar neighbors based on cosine similarity. The choice of $K$ significantly affects the graph's quality and the model's ability to capture spatial dependencies for wear prediction. To determine an optimal $K$ value, a grid search was conducted over a range of $K$ from 1 to 20, with performance evaluated on the wear cycle data of bearing 1 from task 1 using MAE and RMSE as the metrics. The search was performed by training the T-GAT model with each $K$ value while keeping other hyperparameters constant, and the results are summarized in Fig. 11.

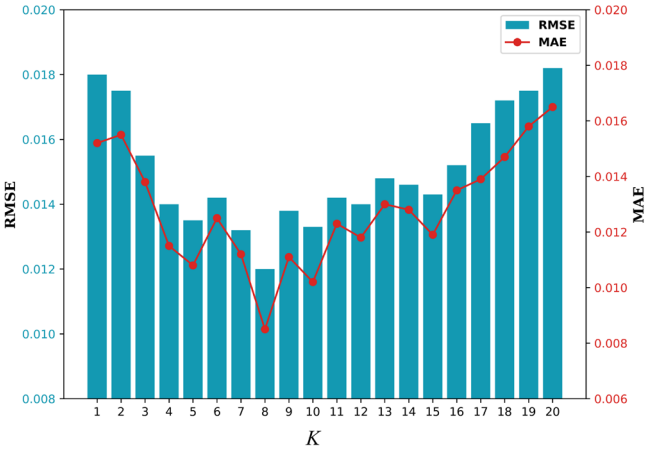



Fig. 11. Performance comparison with $K$ values change on task 1.

The grid search results reveal that $K = 1$ results in the highest prediction error due to an overly sparse graph, which fails to capture sufficient spatial relationships among AE signal segments. As $K$ increases from 1 to 8, the prediction error decreases progressively, indicating improved modeling of spatial dependencies with an increasing number of neighbors. The minimum prediction error is observed at $K = 8$, suggesting that this configuration achieves an optimal balance between capturing relevant local structures and avoiding excessive noise. Beyond $K = 8$, the prediction error increases with noticeable fluctuations, particularly as $K$ approaches 20, attributable to the inclusion of less relevant neighbors that degrade graph quality and predictive accuracy. This trend underscores the trade-off between connectivity and noise, with $K = 8$ identified as the point of optimal performance.

TABLE VI
SENSITIVITY AND STABILITY ANALYSIS OF $K$-MEANS WEAR-STAGE CLUSTERING

| K | SC | CHI | ARI |
|---|---|---|---|
| 2 | 0.492 | 19382 | / |
| 3 | 0.608 | 27641 | $0.967 \pm 0.019$ |
| 4 | 0.453 | 18107 | / |
| 5 | 0.387 | 16852 | / |
| 6 | 0.437 | 18054 | / |

Based on the grid search analysis, $K = 8$ is selected as the neighborhood size for KNN-based graph construction in the T-GAT model, providing a data-driven justification for its use in capturing spatiotemporal dependencies effectively. This value is recommended for practical applications in self-lubricating bearing wear prediction to ensure robust and reliable performance.

*3) Analysis of K-Means Clustering for Wear-Stage Segmentation:* Although the selection of three clusters ($K = 3$) is physically supported by the classical three-stage wear mechanism of self-lubricating bearings, the significant noise and nonstationarity of AE signals necessitate quantitative verification of clustering stability. A systematic sensitivity and stability analysis was, therefore, performed on the complete dataset of bearing 1–1. First, $K$-means clustering was performed for $K$ ranging from 2 to 6, and the resulting partitions were evaluated using the silhouette coefficient (SC) and the Calinski–Harabasz index (CHI). Second, to assess robustness against random initialization, the clustering with $K = 3$ was repeated 50 times using different random seeds. Consistency was measured by the adjusted rand index (ARI) relative to the first run. The results of both experiments are summarized in Table VI.

As shown in Table VI, both the SC and the CHI reach their highest values at $K = 3$, which agrees well with the three-stage wear mechanism of self-lubricating bearings. In addition, the ARI averaged over 50 random initializations is 0.969 with a standard deviation of 0.019. This indicates that the clustering result for $K = 3$ is highly consistent and insensitive to random initialization. The above results confirm that setting the number of wear stages to three is reasonable and that the $K$-means clustering applied to the latent features is sufficiently stable. Therefore, the pseudo labels generated by this clustering can be reliably used to guide the contrastive learning and GMM regularization in the CGC-VAE.

*4) Performance Under Multibearing Mixed Training:* To further verify the robustness of the proposed method to individual differences among bearings, an additional multibearing mixed training experiment was conducted. The experimental dataset was constructed by combining the data from bearings 1-1, 1-2, 2, and 3. The pooled data were split into training, validation, and test sets using a 7:2:1 ratio. The test set contains wear cycles from bearings 2 and 3, which were not included in the training process. The training set includes both partial and full life-cycle data to adequately capture individual variability among bearings. Model architecture and hyperparameters remain identical to previous experiments. Table VII compares the performance of single-bearing training versus multibearing mixed training.

TABLE VII
PERFORMANCE COMPARISON BETWEEN SINGLE-BEARING AND MULTIBEARING MIXED TRAINING

| Training strategy | MAE | RMSE |
|---|---|---|
| Single-bearing (Task 1) | 0.00423 | 0.00527 |
| Multi-bearing mixed | 0.00371 | 0.00403 |

As shown in Table VII, multibearing mixed training reduces MAE and RMSE by approximately 13%, demonstrating significantly enhanced robustness to individual bearing differences. The multibearing mixed training experiment shows that the proposed method maintains excellent performance even across different bearing individuals.

## IV. CONCLUSION

In this article, a novel method is proposed for wear prediction of aerospace self-lubricating bearings using AE signals, overcoming challenges associated with high-dimensional and noisy data in complex operational environments. Initially, CGC-VAE is employed to extract discriminative latent features from preprocessed AE signals, using $K$-means clustering to segment distinct wear stages, along with contrastive learning and GMM regularization to enhance feature discriminability. T-GAT is subsequently integrated to model the spatiotemporal dynamics of these features through a dynamic graph structure based on temporal adjacency and feature similarity, facilitating accurate wear prediction across diverse scenarios. Finally, experimental validation on aerospace bearing datasets demonstrates the method's effectiveness in both full lifecycle and partial wear scenarios, showcasing its robustness and adaptability through transfer learning. Ablation study and parameter analysis further validate the critical role of each component, ensuring reliable performance for predictive maintenance in aerospace applications.

In spite of these advancements, the proposed method has certain limitations that warrant consideration. First, the reliance on a limited dataset may restrict the model's generalization ability across diverse bearing types or operating conditions, as the current validation is based on specific experimental datasets. Second, the computational complexity of the T-GAT and CGC-VAE models poses challenges for real-time prediction, particularly in resource-constrained environments such as edge devices. Third, the method currently uses only AE signals, potentially missing complementary insights from multimodal data sources like torque or temperature signals, which could enhance overall prediction accuracy. These limitations provide valuable insights for future research.

For future research, integrating multimodal data, such as torque or temperature signals, could further enhance the analysis of wear degradation. Additionally, exploring advanced graph-based models to capture more intricate dependencies in AE signals may improve prediction accuracy across diverse operating conditions. Last, adapting the framework for real-time industrial applications, with a focus on computational efficiency and robustness, presents a promising direction for practical deployment.

## REFERENCES


[1] J. Lu, M. Qiu, and Y. Li, "Numerical analysis of self-lubricating radial spherical plain bearings and investigations on fatigue damage mechanisms of the liner," *Tribology Int.*, vol. 96, pp. 97–108, Apr. 2016.

[2] V. Pandiyan, M. Akeddar, J. Prost, G. Vorlaufer, M. Varga, and K. Wasmer, "Long short-term memory based semi-supervised encoder—Decoder for early prediction of failures in self-lubricating bearings," *Friction*, vol. 11, no. 1, pp. 109–124, Jan. 2023.

[3] N. Mokhtari, J. G. Pelham, S. Nowoisky, J.-L. Bote-Garcia, and C. Gühmann, "Friction and wear monitoring methods for journal bearings of geared turbofans based on acoustic emission signals and machine learning," *Lubricants*, vol. 8, no. 3, p. 29, Mar. 2020.

[4] F. König, C. Sous, A. O. Chaib, and G. Jacobs, "Machine learning based anomaly detection and classification of acoustic emission events for wear monitoring in sliding bearing systems," *Tribol. Int.*, vol. 155, Mar. 2020, Art. no. 106811.

[5] W. Caesarendra, B. Kosasih, A. K. Tieu, H. Zhu, C. A. Moodie, and Q. Zhu, "Acoustic emission-based condition monitoring methods: Review and application for low speed slew bearing," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 134–159, May 2015.

[6] A. Hase, "Early detection and identification of fatigue damage in thrust ball bearings by an acoustic emission technique," *Lubricants*, vol. 8, no. 3, p. 37, Mar. 2020.

[7] S.-C. Du, T. Liu, D.-L. Huang, and G.-L. Li, "An optimal ensemble empirical mode decomposition method for vibration signal decomposition," *J. Vib. Acoust.*, vol. 139, no. 3, Jun. 2017, Art. no. 031003.

[8] P. Revill, A. Clarke, R. Pullin, and G. Dennis, "Acoustic emission monitoring of wear in aerospace self-lubricating bearing liner materials," *Wear*, vols. 486–487, Dec. 2021, Art. no. 204102.

[9] F. König, J. Marheineke, G. Jacobs, C. Sous, M. J. Zuo, and Z. Tian, "Data-driven wear monitoring for sliding bearings using acoustic emission signals and long short-term memory neural networks," *Wear*, vol. 476, Jul. 2021, Art. no. 203616.

[10] F. Elasha, M. Greaves, D. Mba, and D. Fang, "A comparative study of the effectiveness of vibration and acoustic emission in diagnosing a defective bearing in a planetry gearbox," *Appl. Acoust.*, vol. 115, pp. 181–195, Jan. 2017.

[11] Y. Deng, D. Shichang, J. Shiyao, Z. Chen, and X. Zhiyuan, "Prognostic study of ball screws by ensemble data-driven particle filters," *J. Manuf. Syst.*, vol. 56, pp. 359–372, Jul. 2020.

[12] D. Li, J. Chen, R. Huang, Z. Chen, and W. Li, "Sensor-aware CapsNet: Towards trustworthy multisensory fusion for remaining useful life prediction," *J. Manuf. Syst.*, vol. 72, pp. 26–37, Feb. 2024.

[13] S. Jia, Y. Deng, J. Lv, S. Du, and Z. Xie, "Joint distribution adaptation with diverse feature aggregation: A new transfer learning framework for bearing diagnosis across different machines," *Measurement*, vol. 187, Jan. 2022, Art. no. 110332.

[14] Y. Deng, S. Du, D. Wang, Y. Shao, and D. Huang, "A calibration-based hybrid transfer learning framework for RUL prediction of rolling bearing across different machines," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.

[15] X. Zhang, L. Jiang, R. Huang, and T. Zhang, "A multiscale cross-channel attention network for remaining useful life prediction with variable sensors," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–10, 2025.

[16] S. Piao, R. Huang, and F. Tsung, "CRULP: Reliable RUL estimation inspired by conformal prediction," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–11, 2024.

[17] Y. Deng, D. Huang, S. Du, G. Li, C. Zhao, and J. Lv, "A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis," *Comput. Ind.*, vol. 127, May 2021, Art. no. 103399.

[18] Y. Xu et al., "Domain constrained cascadic multireceptive learning networks for machine health monitoring in complex manufacturing systems," *J. Manuf. Syst.*, vol. 80, pp. 563–577, Jun. 2025.

[19] W. Fu, R. Zhou, Y. Gao, Z. Guo, and Q. Yu, "A diffusion model-based deep learning approach for denoising acoustic emission signals in concrete," *Measurement*, vol. 251, Jun. 2025, Art. no. 117143.

[20] M. S. Nashed, J. Renno, M. S. Mohamed, and R. L. Reuben, "Gas turbine failure classification using acoustic emissions with wavelet analysis and deep learning," *Expert Syst. Appl.*, vol. 232, Dec. 2023, Art. no. 120684.

[21] M. Moradi, G. Galanopoulos, T. Kuiters, and D. Zarouchas, "A novel intelligent health indicator using acoustic waves: CEEMDAN-driven semi-supervised ensemble deep learning," *Mech. Syst. Signal Process.*, vol. 224, Feb. 2024, Art. no. 112156.

[22] C. Han, T. Liu, Y. Jin, and G. Yang, "Acoustic emission intelligent identification for initial damage of the engine based on single sensor," *Mech. Syst. Signal Process.*, vol. 169, Apr. 2022, Art. no. 108789.

[23] D. Li, Q. Chen, H. Wang, P. Shen, Z. Li, and W. He, "Deep learning-based acoustic emission data clustering for crack evaluation of welded joints in field bridges," *Autom. Constr.*, vol. 165, Sep. 2024, Art. no. 105540.

[24] A. Ebrahimkhanlou, B. Dubuc, and S. Salamone, "A generalizable deep learning framework for localizing and characterizing acoustic emission sources in riveted metallic panels," *Mech. Syst. Signal Process.*, vol. 130, pp. 248–272, Sep. 2019.

[25] Z. Han, J. Zhao, H. Leung, K. F. Ma, and W. Wang, "A review of deep learning models for time series prediction," *IEEE Sensors J.*, vol. 21, no. 6, pp. 7833–7848, Mar. 2021.

[26] L. Cheng, R. Khalitov, T. Yu, J. Zhang, and Z. Yang, "Classification of long sequential data using circular dilated convolutional neural networks," *Neurocomputing*, vol. 518, pp. 50–59, Jan. 2023.

[27] J. Li, X. Cao, R. Chen, C. Zhao, Y. Li, and X. Huang, "Prediction of remaining fatigue life of metal specimens using data-driven method based on acoustic emission signal," *Appl. Acoust.*, vol. 211, Aug. 2023, Art. no. 109571.

[28] J. Ma, Y. Fu, T. Cheng, D. He, H. Cao, and B. Yu, "SCADA data-driven spatio-temporal graph convolutional neural network for wind turbine fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–10, 2025.

[29] W. Yan, J. Liu, J. Luo, W. Liu, Y. Ma, and Q. Zhang, "A temporal–spatial embedding and dynamic aggregation network with adaptive weighting spectrum for EEG motion intention recognition," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–11, 2025.

[30] M. Zhao, C. Taal, S. Baggerohr, and O. Fink, "Graph neural networks for virtual sensing in complex systems: Addressing heterogeneous temporal dynamics," *Mech. Syst. Signal Process.*, vol. 230, May 2025, Art. no. 112544.

[31] Y. Wang, S. Peng, H. Wang, M. Zhang, H. Cao, and M. Liwei, "Remaining useful life prediction based on graph feature attention networks with missing multi-sensor features," *Reliab. Eng. Syst. Saf.*, vol. 258, Jun. 2025, Art. no. 110902.

[32] Y. He et al., "A systematic method of remaining useful life estimation based on physics-informed graph neural networks with multisensor data," *Reliab. Eng. Syst. Saf.*, vol. 237, Sep. 2023, Art. no. 109333.

[33] F. Zonzini, D. Bogomolov, T. Dhamija, N. Testoni, L. De Marchi, and A. Marzani, "Deep learning approaches for robust time of arrival estimation in acoustic emission monitoring," *Sensors*, vol. 22, no. 3, p. 1091, Jan. 2022.

[34] Y. Wang, M. Wu, D. Li, L. Xie, and Z. Chen, "A survey on graph neural networks for remaining useful life prediction: Methodologies, evaluation and future trends," *Mech. Syst. Signal Process.*, vol. 229, Apr. 2025, Art. no. 112449.

[35] C. Zhou and Y. Zhang, "Particle filter based noise removal method for acoustic emission signals," *Mech. Syst. Signal Process.*, vol. 28, pp. 63–77, Apr. 2012.

[36] J. K. Lancaster, "Accelerated wear testing of PTFE composite bearing materials," *Tribology Int.*, vol. 12, no. 2, pp. 65–75, Apr. 1979.

[37] T. Jing, P. Zheng, L. Xia, and T. Liu, "Transformer-based hierarchical latent space VAE for interpretable remaining useful life prediction," *Adv. Eng. Informat.*, vol. 54, Oct. 2022, Art. no. 101781.

[38] L. Lili, Y. Ma, and Y. Wu, "A robust detection scheme against selective forwarding attacks in wireless sensor networks under variable harsh environments," *Eng. Appl. Artif. Intell.*, vol. 138, Dec. 2024, Art. no. 109392.

[39] H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan, "A comprehensive survey on contrastive learning," *Neurocomputing*, vol. 610, Dec. 2024, Art. no. 128645.

[40] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[41] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[42] Y. Deng, J. Lv, D. Huang, and S. Du, "Combining the theoretical bound and deep adversarial network for machinery open-set diagnosis transfer," *Neurocomputing*, vol. 548, Sep. 2023, Art. no. 126391.

[43] S. Wang, L. Yan, S. Du, S. Li, and X. Chen, "Bearing prognostics and health management based on hybrid physical mechanism and data models: A systematic review," *Meas. Sci. Technol.*, vol. 36, no. 5, May 2025, Art. no. 052002.

**Danyue Shen** received the B.S. degree in logistics engineering from Wuhan University of Technology, Wuhan, China, in 2021. He is currently pursuing the M.S. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China.

His research interests include hybrid data-driven models for prognostics and health management, condition monitoring, and intelligent fault diagnosis.

**Shichang Du** (Member, IEEE) received the B.S. and M.S.E. degrees in mechanical engineering from Hefei University of Technology, Hefei, China, in 2000 and 2003, respectively, and the Ph.D. degree in industrial engineering and management from Shanghai Jiao Tong University, Shanghai, China, in 2008.

He was with the University of Michigan, Ann Arbor, MI, USA, from 2006 to 2007, as a Visiting Scholar. He is currently a Professor with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University. He has authored/co-authored 100 articles in journals, including IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, ASME Transaction, and *IISE Transactions*. His current research interests include quality engineering, including quality control with analysis of variation flow, monitoring, and intelligent diagnosis and prognosis of manufacturing systems.

Dr. Du's research has been supported by the National Natural Science Foundation of China, the Ministry of Science and Technology of China, and Shanghai Association for Science and Technology. He receives various awards for his research and teaching, including the IEEE T-ASE Googol Best New Application Paper Award, the QR2MSE Best Paper Award, and the Most Popular Professor Award of ME School, SJTU. He serves for several journals as an Associate Editor, including *Computers and Industrial Engineering Flexible Services and Manufacturing Journal Precision Engineering Measurement Journal of Intelligent Manufacturing*, and *Computers in Industry*.

**Shuo Wang** received the B.S. degree in industrial engineering from Xinjiang University, Ürümqi, China, in 2024. He is currently pursuing the Ph.D. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China.

His research interests include hybrid physical models and data-driven models for prognostics and health management, condition monitoring, and intelligent fault diagnosis.

**Liang Yan** is currently pursuing the B.S. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China. She plans to continue her studies at Shanghai Jiao Tong University by pursuing the Ph.D. degree in mechanical engineering after graduating in 2025.

Her research interests include quality and reliability methods integrating engineering models, industrial big data, artificial intelligence, and hybrid physical and data-driven models for fault diagnosis and remaining useful life prediction.

**Shanshan Li** received the B.S. degree in engineering mechanics from Shandong University, Jinan, China, in 2008, and the M.S. degree in engineering mechanics from China Institute of Aeronautics and Astronautics, Beijing, China, in 2011.

She is currently with the National Key Laboratory of Strength and Structural Integrity and the Aircraft Strength Research Institute of China (ASRI), Xi'an, China. Her research interests include fatigue reliability analysis and verification techniques for aircraft structures.

**Xianmin Chen** received the Ph.D. degree in engineering from Northwestern Polytechnical University, Xi'an, China, in 2018.

He is currently with the National Key Laboratory of Strength and Structural Integrity and the Aircraft Strength Research Institute of China (ASRI), Xi'an, as a Professor. His research interests include aircraft design, fatigue reliability analysis, and verification techniques for aircraft structures.