BearingFM: Towards A Foundation Model for Bearing Fault Diagnosis by Domain Knowledge and Contrastive Learning

Zou Lai[a], Chen Yang[b,*], Shulin Lan[c,*], Lihui Wang[d], Weiming Shen[e] and Liehuang Zhu[b]

[a]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[b]School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing, China
[c]School of Economics and Management, University of Chinese Academy of Science, Beijing, China
[d]Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
[e]State Key Lab of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Keywords:
Fault diagnosis
Foundation model
Semi-supervised learning
Cloud-edge-end collaboration
Contrastive learning

ABSTRACT

Monitoring bearing failures in production equipment can effectively prevent finished product quality issues and unplanned factory downtime, thereby reducing supply chain uncertainty and risk. This is important for improving supply chain sustainability. Due to the limitations in generalization of neural network models, specific models need to be trained for specific tasks. However, in real industrial scenarios, there is a severe lack of labeled samples, making it difficult to deploy fault diagnosis models across massive equipment in workshops. In order to solve the above issue, this paper proposes a cloud-edge-end collaborative semi-supervised learning framework, which provides multi-level computing power and data support for building the foundation model. A data augmentation method based on bearing fault mechanism is proposed, which effectively preserves the inherent essential characteristics in vibration signals by normalizing frequency and adding noise in specific frequency bands. A novel contrastive learning model has been designed, which pulls closer the distances between positive samples and pushes farther away the distances between negative samples in the high-dimensional space through cross comparison in the time dimension and knowledge dimension, thereby extracting the most essential characteristics from unlabeled signals. Multiple sets of experiments conducted on four datasets demonstrate that the proposed approach achieves approximately 98% fault classification accuracy with only 1.2% labeled samples.

1. Introduction

The sustainability of the semiconductor supply chain faces considerable risks due to the industry's reliance on precision manufacturing equipment. On the one hand, semiconductor fabrication facilities consist of many precisely controlled machines performing different operations, such as machining and inspection. Undetected mechanical failures in any of these machines can lead to reduced product quality and reliability (Ye et al., 2020). On the other hand, undetected mechanical failures can escalate and lead to unexpected downtime for entire production lines or even whole factories (Leoni et al., 2023). Therefore, effective fault diagnosis and preventive maintenance strategies can significantly reduce unforeseen interruptions during production, thereby positively influencing the efficiency and sustainability of the entire supply chain (Saihi et al., 2023; Wang et al., 2024).

Rolling bearings, as a crucial component of key equipment in semiconductor manufacturing, are extensively present in photolithography systems (Zhang et al., 2023a), industrial robots (Yin et al., 2023), and various industrial apparatus throughout the manufacturing process (Ni et al., 2023). The reliability of their performance directly impacts the stable operation of the production line. Therefore, it is necessary to continuously monitor the bearings of high-risk or high-maintenance-cost equipment for faults, so as to accurately detect and issue warnings, preventing further losses and risks from the spread of faults (Lu et al., 2023; Yang et al., 2023). Traditional fault classification algorithms, including support vector machines, extreme learning machines, k-nearest neighbors, and their enhanced versions, often require manual feature extraction tailored to specific task scenarios. They heavily rely on manual feature selection and prior knowledge, making it challenging to accurately model real-world environments, thus limiting the generality of these methods (Yan et al., 2023). Due to the dispensability of precise mechanistic models and prior knowledge, data-driven approaches, especially those based on deep learning, are becoming increasingly popular (Florian et al., 2021; Pournader et al., 2021).

However, their practical applicability in the industrial context is limited. The reasons are as follows:

The model's limited generality mandates the creation of specialized fault diagnosis models for nearly every specific tasks, contributing to high deployment costs. In an industrial setting, there are hundreds to thousands of bearings that require health monitoring and fault diagnosis. Existing models for bearing fault diagnosis perform well in the designed target tasks, but it is challenging to effectively generalize them to multiple scenarios (Li et al., 2022; Zhao et al., 2021). Therefore, designing a fault diagnosis model that can be widely used in multiple scenarios is a challenging task. To mitigate the challenge of high deployment costs, algorithmic model compression techniques, such as knowledge distillation (Long et al., 2023) and pruning (Yeom et al., 2021), have been efficacious in reducing the computational complexity of models. At the hardware level, deploying the compressed model on Internet of Things (IoT) devices has significantly reduced hardware costs (Kong et al., 2023; Shirin Abkenar et al., 2022). Nonetheless, the extant literature on model compression techniques and the IoT has neglected a crucial issue: the absence of a foundational model tailored for bearing fault diagnosis. Training lightweight models separately for hundreds or thousands of core components in industrial equipment is itself an impractical problem (Li et al., 2024; Lu et al., 2023). Therefore, building a foundation model for bearing fault diagnosis and achieving effective fine-tuning to adapt to downstream tasks is crucial to promoting the practical application of fault diagnosis models in real-world industrial scenarios.

Large amounts of unlabeled bearing vibration signals can be acquired during the operational processes of industrial equipment, but the cost of labeling them is prohibitively high. Existing deep learning models often rely on a significant number of well-labeled samples representing both healthy and various fault conditions (Zhang et al., 2023b). When only a small amount of labeled data is available, deep models will face the dilemma of overfitting and poor generalization capabilities (Jiao et al., 2020). Unlike data such as images and sound signals that align with human intuition, the labeling of mechanical fault relies heavily on domain expertise, which is time-consuming and costly (Zhao and Shen, 2023). An intuitive approach is data augmentation, increasing the quantity of labeled invariant data through various transformations (Liu et al., 2022). However, traditional data augmentation methods (such as rotation, cropping, or their combinations) were initially designed for images and are not suitable for bearing vibration signals (Hu et al., 2023; Zhuo and Ge, 2021).

Large-scale foundation models such as GPT-4 (OpenAI, 2023) have demonstrated highly intelligent cross-domain capabilities and zero-shot generalization abilities. Effective feature extraction models (Han et al., 2023) and self-supervised representation learning algorithms (Huang et al., 2023) are key to achieving the aforementioned performance leaps in foundation models (Li et al., 2024). Training a universal fault diagnosis model based on extensive unlabeled data, capable of efficient fine-tuning for specific target tasks, is crucial for addressing the aforementioned issues. Within this paradigm, adapting to specific downstream tasks is achievable solely through domain fine-tuning of the foundation model. However, as far as we are aware, there is currently no pertinent research available.

Inspired by GPT-like foundation models, this paper proposes BearingFM: a developing foundation model for bearing fault diagnosis by domain knowledge and contrastive learning. The main contributions of this paper are as follows:

- A cloud-edge collaborative bearing fault diagnosis foundation model is proposed. Constructing a foundation model on the cloud layer leverages large unlabeled data to train a highly generalizable model, reducing the need for specialized models. Efficient fine-tuning with a small labeled dataset on edge servers rapidly adapts this foundation model to specific industrial scenarios.

- A sample augmentation method rooted in the physical mechanism of bearing vibration data is proposed. By normalizing the mechanical rotation frequency and adding noise in specific frequency bands related to the fault mechanism, the differences in vibration signals between different devices are effectively reduced, providing data-level support for the construction of a basic model.

- Sample augmentation has achieved a certain degree of sample expansion, but it cannot completely solve the problem of lacking labeled samples. Therefore, a novel contrastive learning model has been designed. By predicting and cross-comparing across time and knowledge dimensions, the model focuses on extracting inherent commonalities across diverse bearing vibration signals. The cross-dimensional comparison directs the model to core shared characteristics, overcoming the limitation of insufficient labeled exemplars.

The rest of this paper is organized as follows. Section 2 introduced the related work. Section 3 provided a detailed explanation of the proposed cloud-edge-end collaborative framework. Section 4 introduced a bearing vibration signal enhancement method empowered by domain knowledge, along with a contrastive learning model that considers the dimensions of temporal level and knowledge level. Section 5 validated the effectiveness of the proposed methods through multiple experiments. Finally, Section 6 concluded this paper.

2. Related Work

2.1. Semi-supervised Learning

Semi-supervised learning proves adept at leveraging unlabeled data to enhance model generalization. By maximizing the similarity among positive samples and minimizing the distance between negative samples, the algorithm effectively captures the intrinsic structure of the data, acquiring invariant representations across diverse datasets. Yu et al. (2021) proposed a three-stage semi-supervised learning method that achieves accurate fault diagnosis with limited labeled data. Sarkar and Etemad (2022) introduced SSL-ECG, which enhances the model's ability to learn latent abstract representations from unlabeled time-series data by applying six transformations to the dataset. Oord et al. (2019) proposed an unsupervised learning method based on contrastive predictive coding. By using a powerful autoregressive model to predict future representations in latent space, it efficiently extracts useful features from high-dimensional data. Eldele et al. (2023) proposed a time-series representation learning framework based on time and context contrast (TS-TCC). It extracts useful representations from unlabeled data by contrasting different data augmentation views. Yue et al. (2022) introduced two contrastive losses to perform contrastive learning hierarchically. By contrasting the same sample at different time segments and different samples, it learns a universal representation of time series at arbitrary semantic levels. Hu et al. (2023) proposed a cross-instance and cross-time self-supervised learning framework for fault diagnosis. By integrating self-supervised learning from a large amount of unlabeled data and supervised learning from a small amount of labeled data, it enriches the learnable data capacity. Peng et al. (2023) proposed an open-set fault diagnosis method based on supervised contrastive learning. It simulates unknown faults by generating negative out-of-distribution data, achieving fault diagnosis under sample imbalance in open-set scenarios.

Semi-supervised learning, particularly contrastive learning, has achieved significant success in various fields. The key to the success of contrastive learning lies in data augmentation (Chen et al., 2020). However, existing research attempts to enhance data through noise addition, scaling, and other transformations at the data level, overlooking the mechanistic knowledge and expert experience accumulated over decades in specialized areas such as bearing fault diagnosis.

2.2. *Industrial Foundation Model*

In the past two years, large-scale foundation models (LSF-Models) such as GPT4 (OpenAI, 2023) and GLM-130B (Zeng et al., 2023) have demonstrated highly intelligent cross-domain capabilities and zero-shot generalization abilities, and can generalize to unseen data without additional training. Effective feature extraction models (Han et al., 2023) and self-supervised representation learning algorithms (Huang et al., 2023) are the key to the above-mentioned performance leaps of large-scale foundation models (Li et al., 2024). The emergence of LSF-Models in NLP and CV fields has realized the grand unification of models. By pre-training a LSF-model and fine-tuning in vertical domains, efficient downstream adaptation of models can be achieved. Jin et al. (2023) proposed the Time-LLM framework, which successfully applies large language models (LLMs) to time series prediction through re-programming and prompt-as-prefix methods. On the basis of keeping the basic model unchanged, this framework effectively improves the inference ability of LLMs on time series data through enriching the text prototypes and input contexts. Wu et al. (2020) proposed a new time series prediction method that utilizes self-attention mechanisms to learn complex patterns and dynamics of time series data using Transformer-based machine learning models. This general framework can be applied to univariate and multivariate time series data, as well as time series embeddings. Chang et al. (2023) proposed LLM4TS, which achieves time series prediction by pre-training LLMs in two stages. This framework improves time series patches and time encodings, enhancing the performance of LLMs in processing time series data.

The cross-domain capabilities and zero-shot generalization abilities of foundation models are the key to solving the problem of poor model generalizability. In recent months, foundation models in the industrial field have made some progress. However, the overall situation is still in the initial stage, and much of the relevant work has not yet been formally published. As far as we know, there is not yet research on foundation models applied to rotating equipment fault diagnosis.

3. Collaborative framework

3.1. *Overall framework*

In response to the challenges faced in deploying bearing fault diagnosis in large-scale industrial settings, such as the scarcity of labeled samples and high deployment costs, this paper proposes a cloud-edge-end collaborative framework for general bearing fault diagnosis. As illustrated in Fig. 1, the cloud server can cost-effectively gather unlabeled vibration signals from a wide range of similar rotating equipment. By integrating these with a limited number of labeled samples, the server is able to develop a foundational model. The edge server then performs domain fine-tuning on this foundational model to achieve efficient adaptation for specific tasks. Unlike training specialized models from scratch, the approach of employing a foundational model with domain fine-tuning effectively leverages the universal fault mechanism features inherent in the extensive unlabeled pre-training data. This strategy facilitates faster model convergence. The edge server, taking into account the different operating conditions of various end devices, further customizes the fine-tuned model through distillation and deploys it to the corresponding end devices. End devices, which are low-cost and low-power IoT devices deployed on workshop equipment for fault diagnosis, benefit from the advantages of low cost and low communication latency (Yang et al., 2022). Since the edge server has customized the model through distillation based on the operating conditions of different devices, the framework significantly reduces the storage and computational demands on IoT devices.

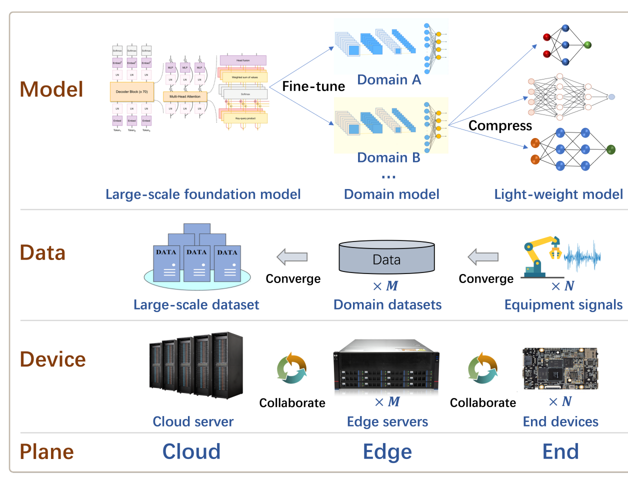

Fig. 1. General Bearing Fault Diagnosis Framework with Cloud-edge-end Collaboration.

Cloud layer: The powerful computational capabilities of the cloud server provide the computational support for constructing the foundation model. At the device level, the cloud server consists of a massive number of high-performance GPU/TPU units. On the data front, the cloud server aggregates publicly available bearing fault diagnosis and life prediction datasets from the internet, as well as bearing vibration signals collected from industrial equipment by subordinate edge nodes. The extensive bearing vibration signals collectively form the large-scale dataset within the cloud server. Leveraging the robust computational power of the cloud server and the large-scale dataset, a foundation model suitable for various bearing fault diagnosis tasks is trained. While it may not achieve optimal performance in all downstream tasks, the universal foundation model, having learned invariant features from a massive and diverse set of bearing data sources, can be fine-tuned slightly for specific downstream applications, yielding improved performance.

Edge layer: The edge layer serves as the link between the large cloud-based model and the specific downstream tasks at the end layer. At the device level, the edge server is equipped with GPU/TPU units, providing necessary computational support for the efficient domain fine-tuning of the foundation model. On the data level, the edge server aggregates bearing vibration signals collected by subordinate end devices and encapsulates data from similar devices into domain-specific datasets. On the model level, the edge server employs a two-stage domain adaptation paradigm. In the first stage, based on the specific requirements of fault diagnosis for the end devices, the foundation model is adapted to the corresponding domain through parameter-efficient fine-tuning. In the second stage, considering the operating conditions of the end devices, the foundation model undergoes personalized distillation/model compression. The compressed model, tailored to fulfill the fault diagnosis requirements of end devices, must be lightweight to ensure reliability on IoT devices at the end layer, particularly considering the constraints imposed by limited hardware resources.

End layer: The end layer forms the foundation of the cloud-edge-end collaborative framework. At the device level, the end layer comprises a vast array of workshop mechanical equipment along with corresponding IoT devices. Its primary role is to connect and manage sensors and manufacturing devices in the industrial field, ultimately supporting the construction of datasets and the deployment of models. In data layer, through IoT devices at the end layer, real-time collection and preprocessing of massive bearing vibration signals in the workshop can be achieved, resulting in data packets that record the historical vibration signals of the equipment. The collection and preprocessing of signals typically have lower computational demands. Compared to traditional data collection methods based on edge/cloud servers, the introduction of IoT devices greatly reduces the cost of data collection. In model layer, compressed AI models for fault diagnosis can be deployed on end IoT devices. Fault diagnosis often requires quick responses and timely actions. In contrast to traditional cloud computing paradigms, running fault diagnosis models on IoT devices not only lowers deployment costs but also reduces communication latency. This facilitates faster detection and identification of potential faults.

3.2. *Cloud-edge-end collaboration factors supporting foundation model construction*

Construction of Domain-Specific Datasets Based on IoT Devices: Data collection and preprocessing are the foundation of constructing domain-specific datasets. The introduction of IoT devices has significantly reduced the cost of data collection processes. These devices collect real-time vibration signals from a vast array of mechanical devices within the factory workshop. Industrial sites may feature sensors with varying noise profiles and sampling rates, necessitating that IoT devices perform appropriate filtering based on the hardware characteristics of the specific vibration signal sampling circuit to mitigate noise interference. Since IoT devices are installed in specific mechanical equipment, it is possible to design dedicated filter parameters and resampling algorithms tailored to the unique sampling circuit. By integrating the vibration data sampling and preprocessing stages into IoT devices, not only is real-time processing enhanced, but cloud communication costs are also reduced. Furthermore, this approach effectively leverages the computational capabilities of IoT devices, making data processing more intelligent and adaptable. Edge servers categorize the collected vibration signals from subordinate end nodes, creating domain-specific datasets.Cloud servers then aggregate these domain datasets from edge servers, integrating them into a more universally applicable and extensive dataset.

Efficient Fine-tuning: Each edge server's subordinate devices exhibit uniqueness in fault characteristics, data distribution, and business logic. In practical applications, it is often necessary to leverage knowledge closely related to a specific domain, and if the information in foundation models is not filtered and optimized, it may not only be ineffective but also lead to resource waste and reduced efficiency. Therefore, in the fault diagnosis process, efficient fine-tuning of the foundation model for a specific application domain becomes particularly important. Further domain fine-tuning of foundation models helps eliminate unnecessary information in a specific domain while enhancing the learning of key features. The fine-tuned model can more accurately identify and handle faults within the domain, thereby improving diagnostic accuracy and efficiency. Additionally, in practical applications, fault patterns may change over time or with technological development. This requires the model to quickly adapt to new fault features. By performing targeted fine-tuning of the foundation model, the model can integrate new data and knowledge rapidly while preserving its original knowledge structure, thus maintaining the model's flexibility and adaptability.

Personalized Model Compression: Deploying fault diagnosis models on IoT devices can significantly reduce costs and enhance the real-time capability of fault diagnosis. However, IoT devices typically have limited computational power, storage space, and power supply, making it impractical to run large fault diagnosis models directly on these devices. Therefore, model compression capability is a necessary feature in cloud-edge-end fault diagnosis frameworks.

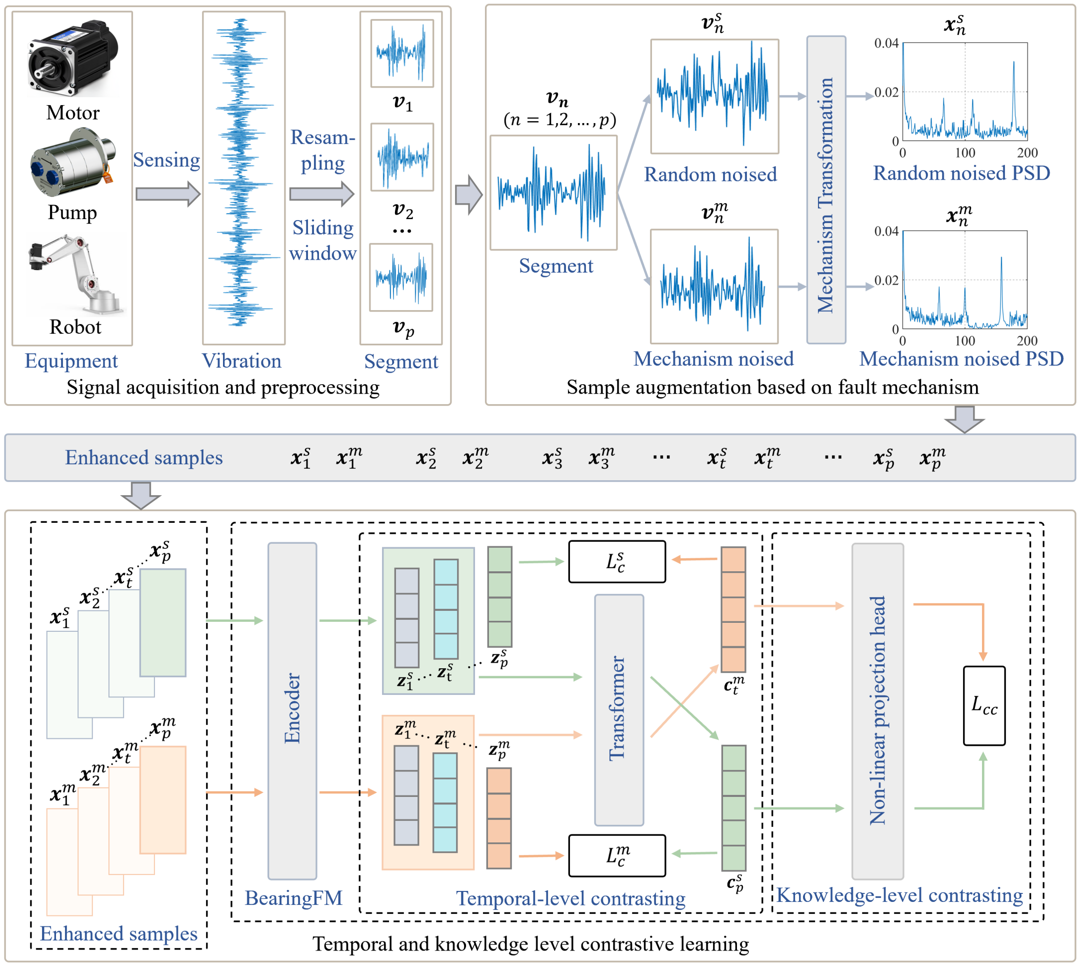

Fig. 2. Overall Framework of BearingFM.

Edge servers should perform personalized compression of the model based on the specific fault diagnosis scenarios of each IoT endpoint device. For example, a device may have 10 different operating conditions, but in actual production processes, it operates only in three of them. During the compression process, the edge server only needs to ensure the model achieves high accuracy in these three operating conditions.

4. Methodology

The core of contrastive learning lies in maximizing the similarity between different transformations of samples from the same class, while minimizing their similarity with other samples. Therefore, augmenting the vibration signal data is the key to the success of contrastive learning methods (Chen et al., 2020). Inspired by self-supervised learning in speech (Oord et al., 2019), we propose Bearing foundation Model (BearingFM). The overall framework of the proposed BearingFM is shown in Fig. 2. It includes three parts: Signal acquisition and preprocessing, Sample augmentation based on fault mechanisms, and Temporal and Knowledge Level Contrastive Learning.

4.1. Signal acquisition and preprocessing

The characteristic components in the signal of bearing faults are closely related to the rotation angle of the bearing, and the characteristic frequency of the fault signal is directly proportional to the rotational speed. By resampling the rotational frequency ω_0 to a fixed value, the influence of different speeds on the spectrum can be eliminated.

Rotating equipment that requires fault diagnosis is typically expensive or critical, and almost all of them are equipped with rotational speed sensors. In cases where a few do not have rotational speed sensors, their motor controllers are internally equipped with speed observers. For bearing fault diagnosis, obtaining the rotational speed is easy and does not incur additional costs.

Assume the sampling rate of the vibration signal sensor is f_s, and the original vibration signal $\hat{V}$ of duration t_s is collected. Resample it with time interval Δt_r to obtain the sample with normalized rotation speed and sampling rate:

$$\boldsymbol{V} = resample(\hat{V}, f_s, \Delta t_r) \tag{1}$$

where resample(·) represents the resampling operation, Δt_r is the resampled time interval and can be calculated as:

$$\Delta t_r = \frac{t_s \omega_m}{f_n \omega_n} \tag{2}$$

where ω_m is the angular velocity of the bearing, ω_n is the normalized standard angular velocity, and f_n is the normalized standard sampling rate.

Split $\boldsymbol{V}$ into p sub-sequences of length n using the sliding window method. For each sub-sequence, resample it at regular angular intervals as follows. $\boldsymbol{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_p\}$.

The above process normalizes the bearing vibration signals of different rotation speeds and sampling rates to a standard angular velocity and sampling rate, which helps the deep learning model focus on extracting the order characteristic features (Wang et al., 2019) specific to bearing faults.

4.2. Sample augmentation based on fault mechanisms

Recent research has demonstrated that combining various sample augmentation techniques can yield superior performance in image data analysis (Chen et al., 2020). In the realm of time series contrastive learning, a combination of transformations such as scaling, jittering, and time shifting is frequently employed to achieve time sample augmentation, thereby addressing the issue of insufficient labeled samples (Luo et al., 2023).

When bearing fault occurs, it generates characteristic signals at specific frequencies. Accurate extraction and identification of these characteristic signals are crucial for effective fault diagnosis of bearings. Drawing on the typical methods in existing contrastive learning research, this paper employs data augmentation techniques at the signal level, including noise addition, scaling, and shifting, to generate augmented views of the data. However, signal-level augmentation indiscriminately applies noise addition, scaling, and shifting operations to the vibration signals, which may potentially disrupt the specific frequency characteristics essential for fault diagnosis. Therefore, we have designed a sample augmentation method based on fault mechanisms. This method leverages prior knowledge of fault mechanisms to ensure that the unique signals indicative of faults are preserved while achieving sample augmentation. By generating both signal-augmented views and mechanism-augmented views of the same sample for comparison, the model can deeply learn fault characteristics that remain invariant under different interference conditions. This approach effectively enhances the model's generalization capability and robustness.

For the input sample $\boldsymbol{v}_p$, its signal-level enhanced view is represented as $\boldsymbol{v}_p^s$. Signal-level enhancement is achieved by applying finite changes to the shape of the original signal, including operations such as scaling, time shifting, and adding noise. Signal-level enhancement introduces significant disturbances to the signal shape while preserving temporal information.

The representation of the mechanistic-level enhanced view is denoted as $\boldsymbol{v}_p^m$. Mechanistic-level enhancement is achieved through analysis of bearing characteristic frequencies. After a bearing develops a fault, local impact forces are generated when mechanical contact occurs at the damaged location, resulting in pulse excitation and eliciting vibration responses from bearing components and bearing seats. The frequencies at which impacts occur at different fault locations (fault characteristic frequencies) vary. Characteristic frequencies can be obtained through motion analysis of the bearing, considering the bearing's rotational speed, the shape and dimensions of its components, and the relationships governing its movement.

Inner race fault:

$$f_i = \frac{n_b}{2}\left(1 + \frac{d_b}{D_m}\cos\alpha\right) f_r \tag{3}$$

Outer race fault:

$$f_o = \frac{n_b}{2}\left(1 - \frac{d_b}{D_m}\cos\alpha\right) f_r \tag{4}$$

Cage fault:

$$f_c = \frac{1}{2}\left(1 - \frac{d_b}{D_m}\cos\alpha\right) f_r \tag{5}$$

Ball fault:

$$f_b = \frac{D_m}{2d_b}\left(1 - \left(\frac{d}{D_m}\cos\alpha\right)^2\right) f_r \tag{6}$$

where n_b represents the number of balls, d_b is the diameter of the balls (mm), D_m is the diameter of the circle passing through the centers of the rolling elements (mm), f_r is the rotational frequency of the shaft (Hz), and α is the contact angle. Furthermore, a bandpass filter can be designed.

$$f_r = \frac{\min(f_i, f_o, f_c, f_b) + \max(f_i, f_o, f_c, f_b)}{2} \tag{7}$$

$$B = \max(f_i, f_o, f_c, f_b) + \min(f_i, f_o, f_c, f_b) \tag{8}$$

where f_r is the center frequency of the bandpass filter, and B is the bandwidth of the passband.

The vibrational signals generated due to malfunctions can be categorized into two primary components: those related to the fault mechanism itself and others comprising carrier waves, harmonics, and noise. Through the application of bandpass filters, it is feasible to isolate and eliminate the frequency bands in artificially generated noise signals that are pertinent to the fault mechanism. This approach facilitates the augmentation of noise in the signal while endeavoring to preserve the inherent characteristics of the bearing vibration signal at a mechanistic level, thereby not compromising its fundamental attributes.

To further analyze the vibrational signals associated with bearing faults, the original vibration signal sample, denoted as $\boldsymbol{v}_p$, is enhanced by adding two types of noise: Gaussian noise, which is unrelated to the fault mechanism, and band noise tailored to the fault mechanism, resulting in the creation of two modified samples, $\boldsymbol{v}_p^s$ and $\boldsymbol{v}_p^m$. Subsequently, these samples undergo a process known as Mechanism Transformation (MT). MT principally consists of three stages: transformation of the signal's Power Spectral Density (PSD), amplitude transformation, and scale transformation. The application of these steps is exemplified using the $\boldsymbol{v}_p^s$ sample.

To alleviate the impact of noise caused by mechanical resonance, the vibration signal is analyzed using Hilbert transform to extract its envelope spectrum:

$$H_p^s(\tau) = \int_{-\infty}^{\infty} \frac{v_p^s(\xi)}{\pi(\tau - \xi)} d\xi \tag{9}$$

where, $v_p^s(\tau)$ is the τ-th point of the sample $\boldsymbol{v}_p^s$.

Transform $H(\tau)$ to the frequency domain:

$$H_p^s(f) = \int_{-\infty}^{\infty} H_p^s(\xi) e^{-i2\pi f \xi} d\xi \tag{10}$$

where, f represents frequency, i is the imaginary unit.

The power spectrum $\hat{x}_p^s(f)$ is the square of the amplitude of the signal after Fourier transform, expressed as:

$$\hat{x}_p^s(f) = |H_p^s(f)|^2 \tag{11}$$

Horizontal, vertical stretching and shifting operations are performed on x_p^s:

$$\boldsymbol{x}_p^s = \text{enhance}(\hat{x}_p^s(f)) \tag{12}$$

where enhance$(\cdot)$ represents horizontal, vertical stretching and shifting operations performed on the samples to enhance the generalization of the deep learning models trained subsequently(Eldele et al., 2023).

The CMT transform process for x_p^m is similar to that for x_p^s, so the details are omitted here for brevity. The power spectrum provides the distribution of signal power in the frequency domain. It can reveal the natural frequencies, harmonic components, and potential fault features of the system. This provides a data-level basis for the subsequent construction of contrastive learning models.

4.3. Temporal and knowledge level contrastive learning

The proposed contrastive learning consists of two parts: time-contrastive module, and knowledge module. The time-contrastive module, using Transformer, predicts the future of the signal-enhanced view from the past of the mechanistic-enhanced view and vice versa. This design further enhances the model's ability to extract essential features related to the mechanism in the time-domain vibration signals of bearings. The Knowledge Contrasting module extracts high-dimensional features from the signal-enhanced view and the mechanistic-enhanced view, compares them, and calculates the loss.

4.3.1. Temporal-level contrast

Theoretical analysis and experimental validation both indicate that bearing damage is a gradual process, and the time-domain vibration signals generated during operation do not undergo sudden changes in a short time (on the order of seconds). In supervised learning, a common approach is to use a sliding window method to divide a long-time sequence of bearing vibration signals into multiple subsequences, where the subsequences share the same fault label. Therefore, subsequences from different time points of the same vibration signal sample serve as positive samples.

The core idea of temporal-level contrast is to optimize the similarity between representations at different time steps, thereby learning temporal dependencies. The specific implementation steps are as follows:

According to section 4.1, by slicing the temporal samples, we can obtain enhancement $\boldsymbol{x}^s = \{x_1^s, x_2^s, ..., x_t^s\}$ at the signal level and enhancement $\boldsymbol{x}^m = \{x_1^m, x_2^m, ..., x_t^m\}$ at the mechanism level, where k represents the sample sequence number. We employed the encoder proposed in reference (Wang et al., 2017) to extract high-dimensional features from the two enhanced views mentioned above. After encoding, $\boldsymbol{x}^s$ and $\boldsymbol{x}^m$ are respectively encoded as $z^s = \{z_1^s, z_2^s, ..., z_p^s\}$ and $z^m = z_1^m, z_2^m, ..., z_p^m\}$. Let p be the predicted depth $(1 < p < t)$, $\hat{\boldsymbol{x}}_p^s = \{x_p^s, ..., x_t^s\}$, $\hat{\boldsymbol{x}}_p^m = \{x_p^m, ..., x_t^m\}$. Further, the encoder encodes $\hat{\boldsymbol{x}}_p^s$ and $\hat{\boldsymbol{x}}_p^s$ into high-dimensional features, denoted as $\hat{\boldsymbol{z}}_p^s \in \mathbb{R}^k$ and $\hat{\boldsymbol{z}}_p^m \in \mathbb{R}^k$. The temporal regression module encodes $\boldsymbol{x}_p^s = \{x_1^s, x_2^s, ..., x_p^s\}$ and $\boldsymbol{x}_p^m = \{x_1^m, x_2^m, ..., x_p^m\}$ into signal-view feature $\boldsymbol{c}_p^s \in \mathbb{R}^w$ and mechanism-view feature $\boldsymbol{c}_p^m \in \mathbb{R}^w$, respectively. Training a linear neural network $\psi_{cz} : \mathbb{R}^{w \to k}$, mapping $\boldsymbol{c}_p^s$ and $\boldsymbol{c}_p^m$ to the same dimension as $\hat{\boldsymbol{x}}_p^s$ and $\hat{\boldsymbol{x}}_p^m$.

The temporal-level comparative loss is as follows

$$L_{TC}^s = -\frac{1}{K} \sum_{k=1}^{K} \log \frac{\exp((\psi_{cz}(\boldsymbol{c}_p^s))^T z_t^s)}{\sum_{i \in \mathrm{M}_{p,t}} \exp((\psi_{cz}(\boldsymbol{c}_p^s))^T z_i^s)} \tag{13}$$

$$L_{TC}^m = -\frac{1}{K} \sum_{k=1}^{K} \log \frac{\exp((\psi_{cz}(\boldsymbol{c}_p^m))^T z_t^m)}{\sum_{i \in \mathrm{M}_{p,t}} \exp((\psi_{cz}(\boldsymbol{c}_p^m))^T z_i^m)} \tag{14}$$

Furthermore, a Transformer was deployed as an autoregressive model to achieve autoregression from high-dimensional features z^s and z^m at the hierarchical level before time step p to high-dimensional features $\boldsymbol{c}_p^s$ and $\boldsymbol{c}_p^m$ from time step p to t. Considering that we did not make modifications to the Transformer model, we will not elaborate on it here. The specific structure of the Transformer model will be provided in the experiments in Section 5.

4.3.2. Knowledge-level contrast

In the temporal contrastive learning, a self-regressive Transformer module has been constructed to summarize the features of samples before time p into high-dimensional vectors $\boldsymbol{c}_p^s$ and $\boldsymbol{c}_p^m$. If a mini-batch contains M samples, after signal augmentation and mechanism enhancement, $2M$ enhanced views are obtained. Let $\boldsymbol{c}_p^k$ represent the k-th sample of the views enhanced through signal augmentation and mechanism enhancement for a batch. In each mini-batch, there are a total of 2 positive pairs and 2(M-1) negative pairs. Define $(\boldsymbol{c}_p^k, \boldsymbol{c}_p^{k+})$ as a positive pair, then the knowledge-level contrastive loss is given by

$$L(c_{\boldsymbol{p}}^k, \boldsymbol{c}_p^{k+}) = -\log \frac{\exp(sim(\boldsymbol{c}_p^k, \boldsymbol{c}_p^{k+})/\eta}{\sum_{j=1}^{2M} \delta_{i,j} \exp(sim(\boldsymbol{c}_p^k, \boldsymbol{c}_p^j)/\eta} \tag{15}$$

$$L_{cc} = \frac{1}{2N} \sum\nolimits_{v}^{2M} [L(2v-1, 2v) + L(2v, 2v-1)] \tag{16}$$

The overall self-supervised loss function is

$$L_{con} = \lambda_1 \cdot (L_{TC}^s + L_{TC}^m) + \lambda_2 \cdot L_{cc} \tag{17}$$

The weights of the temporal contrastive loss and the knowledge contrastive loss can be adjusted through the hyperparameters λ_1 and λ_2.

5. Experiment

5.1. Dataset description

To validate the feasibility and performance of the proposed BearingFM, relevant experiments were conducted. Bearings used in various types of mechanical equipment across different industries exhibit significant similarities. Due to the lack of publicly available datasets for bearing fault diagnosis in semiconductor manufacturing equipment, four publicly available general bearing fault diagnosis datasets were utilized for the experiments.

CWRU dataset: This publicly available dataset was sourced from Case Western Reserve University (CWRU) (Smith and Randall, 2015). The machinery operated under the conditions of 0-3 horsepower motor loads and motor speeds ranging from 1730 to 1797 revolutions per minute (rpm). The CWRU dataset encompasses three distinct bearing failure states in addition to the normal condition (NC), namely outer race fault (OR), inner race fault (IR), and ball fault (BF). Notably, the vibration signal was recorded at a sampling frequency of 12 kHz. To ensure alignment with fault types in other datasets, samples from the ball fault category were excluded. For the zero-shot testing set, samples from one operational condition of the remaining three fault types were selected. The data from the other operational conditions were then divided into training, validation, and test sets in a 6:2:2 ratio.

PU dataset: The experimental platform of the Paderborn University (PU) (Lessmeier et al., 2016) dataset comprises an electric motor, a torque measuring shaft, a rolling bearing test module, a flywheel, and a load motor. The dataset's experimental data is generated by installing ball bearings with various damage types in the bearing test module. Faulty bearings are categorized into artificial damage and real damage. Artificial damage results primarily from EDM (cracking), drilling (spalling), and electric engraving machines (pitting). Real damage bearings are obtained through accelerated life test benches. Both artificial and real damage encompass three main types of failure: normal condition (NC), outer race fault (OR), and inner race fault (IR). In the experiment, three types of bearing fault samples generated by an electric engraver under a specific operating condition were randomly selected as the zero-shot test set. The remaining bearing fault data generated by the electric engraver under different operating conditions were divided into training, validation, and test sets according to a 6:2:2 ratio.

MFPT dataset: This publicly available dataset provided by the Society for Machinery Failure Prevention Technology (SMFPT) comprises the vibration signals collected when a motor operates at a constant speed of 1500 rpm. These signals are categorized into three conditions: normal condition (NC), inner race fault (IR), and outer race fault (OR). Specifically, the dataset includes three sets of normal condition signals, three sets of outer race fault signals under identical conditions, seven sets of inner race fault signals under varying conditions, and seven sets of outer race fault signals under varying conditions. For the experiment, one set of inner race fault signals and one set of outer race fault signals under specific conditions were randomly selected as the zero-sample test set. The remaining samples were divided into training, validation, and test sets in a 6:2:2 ratio.

JNU dataset: This dataset provided by Jiangnan University in China, includes 12 categories under operational conditions of 600 rpm, 800 rpm, and 1000 rpm (Li et al., 2013). These categories comprise normal health condition (NC), inner race fault (IR), outer race fault (OR), and ball fault (BA). Vibration signals were collected at a sampling frequency of 50 kHz, with each data sample having a duration of 20 seconds. To ensure alignment with fault types in other datasets, samples from the ball fault category were excluded. For the zero-shot testing set, samples from one operational condition of the remaining three fault types were selected. The data from the other operational conditions were then divided into training, validation, and test sets in a 6:2:2 ratio.

5.2. Implementation details

The server are equipped with Intel Xeon Platinum 6133 20-core CPU, 128GB RAM, and four NVIDIA RTX A6000 GPUs. The experimental operating system is Ubuntu 22.04, and the PyTorch framework version is 2.0.1. For the Transformer used in the experiment, we set the number of layers to 35, the number of attention heads to 4, and the hidden layer size of the model to 100. The dropout is set to 0.35. Setting $\lambda_1 = 1.0$ and $\lambda_2 = 0.7$ to adjust the weights for temporal-level contrast and knowledge-level contrast.

We evaluate performance using two metrics: accuracy and the macro-averaged F1-score, providing a more comprehensive assessment of performance.

$$\text{Accuracy} = \frac{\sum_{i=1}^{M} TP_i}{N} \tag{18}$$

$$\text{F1}-\text{score} = \frac{2}{M}\sum_{i=1}^{M}\frac{Precision_i \times Recall_i}{Precision_i + \times Recall_i} \tag{19}$$

where $Precision_i = \frac{TP_i}{TP_i+FP_i}$, $Recall_i = \frac{TP_i}{TP_i+FN_i}$. TP_i, FP_i and FN_i denote the False Positives, and False Negatives for the i-th class respectively. N is the total number of samples, and M is the number of classes in the dataset.

5.3. Semi-supervised experiment

To verify the fault classification accuracy of the proposed BearingFM model, we conducted experiments on four publicly available bearing fault diagnosis datasets: CWRU, PU, MFPT, and JNU, as well as the "Merged" dataset obtained by combining the aforementioned four datasets. Each dataset was evaluated under three experimental conditions with 0.4%, 1.2%, and 2% labeled samples, respectively. The construction process of the BearingFM model in the experiments consisted of two phases. In the first phase, the unlabeled "Merged" dataset was used for unsupervised learning. This phase aimed to enable the model to learn the invariant features related to the fault mechanism from a large number of unlabeled samples of vibration signals from bearings of different mechanical equipment. In the second phase, a small amount of labeled samples was randomly selected from the 'Merged' dataset for preliminary fine-tuning of the model. In the third phase, the model is efficiently fine-tuned using a limited number of labeled samples from the target task. The experimental results for the first two phases are shown in the "Merged" row of Table 1. It is noted that, due to the inclusion of bearing vibration signals from various mechanical equipment and operating conditions in the "Merged" dataset, the 0.4% of labeled samples was insufficient for the model to achieve high classification accuracy. However, when the number of labeled samples reached 1.2%, the classification accuracy of the model significantly improved. When the number of labeled samples reached 2%, the model achieved a classification accuracy of nearly 90%. Although further increasing the number of labeled samples could still improve the model's performance, considering the high cost of data labeling in real industrial scenarios, we did not use more than 2% of labeled data in all subsequent experiments.

To further validate the efficient fine-tuning performance of the proposed BearingFM, we utilized the model trained with 0.4% labeled data from the "Merged" dataset as the foundation model and performed fine-tuning on four distinct datasets with a small number of labeled samples (the third stage of the model construction process). The experimental results are shown in Table 1. Utilizing only 0.4% labeled samples from the target tasks, the model achieved high accuracy. When the number of labeled samples reached 1.2%, the model attained 100% classification accuracy on the CWRU and MFPT datasets. With 2% labeled samples, the model reached 100% classification accuracy in three out of four datasets, excluding the PU dataset. The PU dataset presents more complex fault modes, and although the model's accuracy on the PU dataset was lower compared to the other three datasets, it still achieved over 98% classification accuracy with only 1.2% labeled samples for fine-tuning. In summary, across all four target tasks, the model achieved over 93% fault classification accuracy with just 0.4% labeled samples for fine-tuning. With 1.2% labeled samples from the target tasks, the model attained over 98% fault classification accuracy. The experimental results demonstrate that the BearingFM constructed with 0.4% labeled samples exhibits significant potential for effective fine-tuning.

5.4. Zero-shot experiment

In developing the BearingFM, we integrated a diverse dataset from multiple scenarios and employed a sample augmentation method based on bearing fault mechanisms. This approach endowed the model with robust generalization capabilities. To evaluate the zero-shot generalization performance of BearingFM, we adhered to the commonly used zero-sample test set partitioning method in existing bearing fault diagnosis research. Specifically, we randomly selected one operating condition from each dataset, completely excluded it from the training set, and used it as the zero-shot test set, thereby simulating a zero-shot generalization scenario in bearing fault diagnosis tasks.

The zero-shot test results are presented in Table 2. When the labeled data ratio was 0.40%, the model's performance remained robust across all datasets, achieving accuracies of 91.21%, 93.97%, 100.00%, and 99.47% for the CWRU, PU, MFPT, and JNU datasets, respectively. The "Merged" dataset, which includes data from various equipment and conditions, exhibited a relatively lower zero-shot test accuracy at a 0.40% labeling ratio. However, as the labeling ratio increased to 1.20%, the accuracy significantly improved to 87.12%. It is important to note that the test data used in the zero-shot experiments comprised entirely new operating conditions that were absent from the training and validation sets. Overall, the comparison between Tables 2 and 1 indicates that the accuracy of BearingFM on the zero-shot test set was only slightly lower than on the regular test set. This finding suggests that the model, having been trained on a large amount of unlabeled data, effectively captured the intrinsic invariant features of bearing vibration signals under different operating conditions, demonstrating strong zero-shot generalization performance.

Table 1
Performance comparison on different datasets with varying amounts of labeled data.

Dataset	0.4% Labeled data		1.2% Labeled data		2% Labeled data	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Merged	46.57%	42.32%	86.12%	85.87%	89.02%	89.13%
CWRU	95.56%	95.54%	100.00%	100.00%	100.00%	100.00%
PU	93.24%	93.27%	98.19%	98.19%	98.64%	98.64%
MFPT	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
JNU	98.59%	98.59%	99.67%	99.67%	100.00%	100.00%

Table 2
Performance comparison on different datasets with varying amounts of labeled data for zero-shot testing.

Dataset	0.4% Labeled data		1.2% Labeled data		2% Labeled data	
	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
Merged	35.75%	29.15%	87.12%	87.05%	88.06%	87.99%
CWRU	91.21%	90.96%	97.41%	97.40%	99.83%	99.83%
PU	92.27%	92.15%	97.44%	97.44%	97.70%	97.70%
MFPT	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
JNU	99.47%	99.46%	99.72%	99.72%	99.10%	99.09%

5.5. *Convergence rate of the fundamental model*

In the cloud-edge collaborative framework proposed in this paper, the model training process unfolds in three stages. In the first stage, the cloud server utilizes a large amount of mixed data from multiple unlabeled datasets for unsupervised training. In the second stage, the cloud server performs fine-tuning using a small amount of labeled samples from the mixed datasets, thereby producing a foundation model. The aim of the first two stages is to leverage the cloud server's abundant computational and storage resources to build a foundational model with strong generalization capabilities, which can be quickly fine-tuned with a small number of target task samples. In the third stage, the edge server fine-tunes the foundational model using a limited number of labeled samples specific to the target task. Compared to traditional methods of training a model from scratch, this training paradigm effectively accelerates the model's convergence speed and enhances its performance. To validate these conclusions, an experiment was conducted to evaluate the convergence speed of BearingFM using the PU dataset with 1.2% labeled data.

To substantiate these conclusions, an experiment was conducted to evaluate the convergence speed of BearingFM using the PU dataset, with only 1.2% of the data labeled. The experimental results are depicted in Fig. 3. In this figure, "Pretrain" denotes the model training process initiated with the pre-trained model's weights as the starting point, whereas "Normal" signifies training the model from scratch. It is evident from the figure that initializing the model with the pre-trained weights substantially accelerates convergence.

Moreover, the pre-trained model, having been exposed to a large corpus of unlabeled data, effectively assimilates the underlying fault knowledge present in the data.

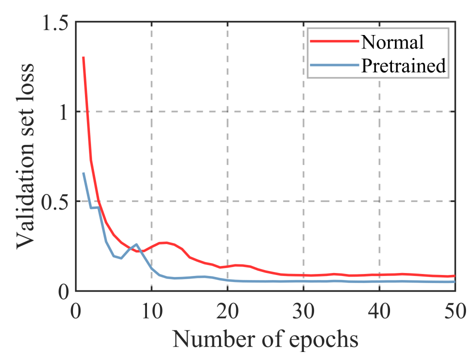

Fig. 3. Comparative of Convergence Speed between Efficient Fine-Tuning and Direct Model Training.

Consequently, during the latter stages of training, when the loss value stabilizes, the process that commences with the pre-trained model demonstrates a markedly lower loss value—approximately 60% of that observed when training from scratch. This clearly indicates that the pre-trained model confers significant advantages in terms of both model performance and training efficiency.

5.6. *Ablation experiment*

The careful selection of augmentations is paramount in contrastive learning, given their sensitivity to the choice of augmentation methods (Chen et al., 2020). To validate the effectiveness of the proposed sample augmentation method based on bearing fault mechanisms, ablation experiments were conducted on the CWRU dataset.

Fig. 4 reveals that, compared to traditional signal-level data augmentation methods, the proposed mechanism-level data augmentation method effectively enhances both accuracy and F1-score.

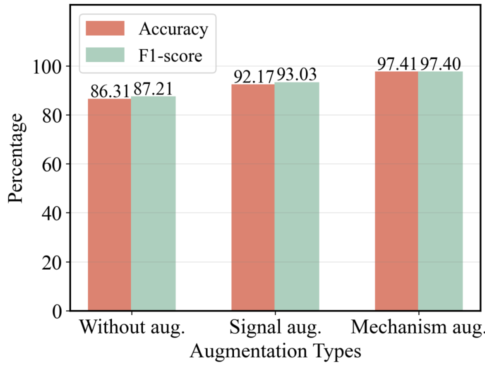

Fig. 4. Accuracy and F1-score under different data augmentation methods.

6. Conclusion

This paper proposes a cloud-edge collaborative semi-supervised bearing fault diagnosis method empowered by domain knowledge. The proposed application paradigm integrates cloud-based pre-training of a foundational model, followed by precise domain-specific fine-tuning at the edge layer. Leveraging a cloud-edge collaborative framework, it provides computational power and data support for constructing the foundation model for bearing fault diagnosis. The proposed domain knowledge empowered contrastive learning method facilitates the construction of a foundation model by extracting invariant features from vibration signals of bearings. Compared to traditional fault diagnosis models, BearingFM exhibits potent feature learning capabilities, enabling it to abstract universal knowledge from data. BearingFM possesses the capacity to concurrently perform fault diagnosis tasks across a diverse array of mechanical equipment bearings. When confronted with novel bearing fault diagnosis tasks that deviate substantially from the pre-training data, BearingFM can capitalize on the extensive knowledge acquired during the pre-training phase as an efficacious foundation. Leveraging a minimal quantity of target training samples, BearingFM can rapidly adapt to the new task.

The scale of the publicly available bearing dataset is limited, so the model we built is just a developing foundation model. Compared to foundation models like ChatGPT, there is still a gap. However, this developing foundation model serves as a stepping stone in building a truly comprehensive and specialized foundation model for bearings and rotating devices in the future. Although the current amount of data is insufficient, the method and framework we proposed have the potential to support a comprehensive foundational model. In our future work, we aim to enhance BearingFM on three fronts. Firstly, the scope of the bearing fault diagnosis datasets will be further expanded to include data from mechanical equipment bearings in specific industries, such as semiconductor manufacturing. This expansion aims to enhance the generalization performance of BearingFM across more diverse fields. Secondly, we will investigate methodologies for multi-domain alignment, involving the isolated extraction of the classification module from the comprehensive model, implemented as adapters, to facilitate enhanced adaptability across diverse scenarios. Lastly, the detection of unknown faults is equally crucial in industrial fault diagnosis. Hence, we contemplate extending the applicability of BearingFM to open-set fault diagnosis scenarios, which is an important direction for future exploration.

References

Chang, C., Peng, W.C., Chen, T.F., 2023. LLM4TS: Two-Stage Fine-Tuning for Time-Series Forecasting with Pre-Trained LLMs. arXiv:2308.08469.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A Simple Framework for Contrastive Learning of Visual Representations, in: Proceedings of the 37th International Conference on Machine Learning, PMLR. pp. 1597–1607. doi:10.1190/geo2021-0590.1.

Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.K., Li, X., Guan, C., 2023. Self-Supervised Contrastive Representation Learning for Semi-Supervised Time-Series Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1–15doi:10.1109/TPAMI.2023.3308189.

Florian, E., Sgarbossa, F., Zennaro, I., 2021. Machine learning-based predictive maintenance: A cost-oriented model for implementation. International Journal of Production Economics 236, 108114. doi:10.1016/j.ijpe.2021.108114.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., Tao, D., 2023. A Survey on Vision Transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 87–110. doi:10.1109/TPAMI.2022.3152247.

Hu, C., Wu, J., Sun, C., Yan, R., Chen, X., 2023. Interinstance and Intratemporal Self-Supervised Learning With Few Labeled Data for Fault Diagnosis. IEEE Transactions on Industrial Informatics 19, 6502–6512. doi:10.1109/TII.2022.3183601.

Huang, G., Laradji, I., Vázquez, D., Lacoste-Julien, S., Rodríguez, P., 2023. A Survey of Self-Supervised and Few-Shot Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 4071–4089. doi:10.1109/TPAMI.2022.3199617.

Jiao, J., Zhao, M., Lin, J., Ding, C., 2020. Classifier Inconsistency-Based Domain Adaptation Network for Partial Transfer Intelligent Diagnosis. IEEE Transactions on Industrial Informatics 16, 5965–5974. doi:10.1109/TII.2019.2956294.

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.Y., Liang, Y., Li, Y.F., Pan, S., Wen, Q., 2023. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. arXiv:2310.01728.

Kong, L., Tan, J., Huang, J., Chen, G., Wang, S., Jin, X., Zeng, P., Khan, M., Das, S.K., 2023. Edge-computing-driven Internet of Things: A Survey. ACM Computing Surveys 55, 1–41. doi:10.1145/3555308.

Leoni, L., De Carlo, F., Tucci, M., 2023. Developing a framework for generating production-dependent failure rate through discrete-event simulation. International Journal of Production Economics 266, 109034. doi:10.1016/j.ijpe.2023.109034.

Lessmeier, C., Kimotho, J.K., Zimmer, D., Sextro, W., 2016. Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification. PHM Society European Conference 3. doi:10.36001/phme.2016.v3i1.1577.

Li, K., Ping, X., Wang, H., Chen, P., Cao, Y., 2013. Sequential Fuzzy Diagnosis Method for Motor Roller Bearing in Variable Operating Conditions Based on Vibration Analysis. Sensors 13, 8013–8041. doi:10.3390/s130608013.

Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., Yan, R., Gryllias, K., 2022. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. Mechanical Systems and Signal Processing 167, 108487. doi:10.1016/j.ymssp.2021.108487.

Li, Y.F., Wang, H., Sun, M., 2024. ChatGPT-like large-scale foundation models for prognostics and health management: A survey and roadmaps. Reliability Engineering & System Safety 243, 109850. doi:10.1016/j.ress.2023.109850.

Liu, S., Jiang, H., Wu, Z., Li, X., 2022. Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. Mechanical Systems and Signal Processing 163, 108139. doi:10.1016/j.ymssp.2021.108139.

Long, Z., Ma, F., Sun, B., Tan, M., Li, S., 2023. Diversified branch fusion for self-knowledge distillation. Information Fusion 90, 12–22. doi:10.1016/j.inffus.2022.09.007.

Lu, S., Lu, J., An, K., Wang, X., He, Q., 2023. Edge Computing on IoT for Machine Signal Processing and Fault Diagnosis: A Review. IEEE Internet of Things Journal 10, 11093–11116. doi:10.1109/JIOT.2023.3239944.

Luo, D., Cheng, W., Wang, Y., Xu, D., Ni, J., Yu, W., Zhang, X., Liu, Y., Chen, Y., Chen, H., Zhang, X., 2023. Time Series Contrastive Learning with Information-Aware Augmentations. Proceedings of the AAAI Conference on Artificial Intelligence 37, 4534–4542. doi:10.1609/aaai.v37i4.25575.

Ni, Q., Ji, J.C., Feng, K., 2023. Data-Driven Prognostic Scheme for Bearings Based on a Novel Health Indicator and Gated Recurrent Unit Network. IEEE Transactions on Industrial Informatics 19, 1301–1311. doi:10.1109/TII.2022.3169465.

Oord, A., Li, Y., Vinyals, O., 2019. Representation Learning with Contrastive Predictive Coding. arXiv:1807.03748.

OpenAI, 2023. GPT-4 Technical Report. arXiv:2303.08774.

Peng, P., Lu, J., Xie, T., Tao, S., Wang, H., Zhang, H., 2023. Open-Set Fault Diagnosis via Supervised Contrastive Learning With Negative Out-of-Distribution Data Augmentation. IEEE Transactions on Industrial Informatics 19, 2463–2473. doi:10.1109/TII.2022.3149935.

Pournader, M., Ghaderi, H., Hassanzadegan, A., Fahimnia, B., 2021. Artificial intelligence applications in supply chain management. International Journal of Production Economics 241, 108250. doi:10.1016/j.ijpe.2021.108250.

Saihi, A., Ben-Daya, M., As'ad, R., 2023. Underpinning success factors of maintenance digital transformation: A hybrid reactive Delphi approach. International Journal of Production Economics 255, 108701. doi:10.1016/j.ijpe.2022.108701.

Sarkar, P., Etemad, A., 2022. Self-Supervised ECG Representation Learning for Emotion Recognition. IEEE Transactions on Affective Computing 13, 1541–1554. doi:10.1109/TAFFC.2020.3014842.

Shirin Abkenar, F., Ramezani, P., Iranmanesh, S., Murali, S., Chulerttiyawong, D., Wan, X., Jamalipour, A., Raad, R., 2022. A Survey on Mobility of Edge Computing Networks in IoT: State-of-the-Art, Architectures, and Challenges. IEEE Communications Surveys & Tutorials 24, 2329–2365. doi:10.1109/COMST.2022.3211462.

Smith, W.A., Randall, R.B., 2015. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. Mechanical Systems and Signal Processing 64–65, 100–131. doi:10.1016/j.ymssp.2015.04.021.

Wang, Y., Tse, P.W., Tang, B., Qin, Y., Deng, L., Huang, T., Xu, G., 2019. Order spectrogram visualization for rolling bearing fault detection under speed variation conditions. Mechanical Systems and Signal Processing 122, 580–596. doi:10.1016/j.ymssp.2018.12.037.

Wang, Y., Xia, T., Xu, Y., Ding, Y., Zheng, M., Pan, E., Xi, L., 2024. Joint optimization of flexible job shop scheduling and preventive maintenance under high-frequency production switching. International Journal of Production Economics , 109163doi:10.1016/j.ijpe.2024.109163.

Wang, Z., Yan, W., Oates, T., 2017. Time series classification from scratch with deep neural networks: A strong baseline, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, Anchorage, AK, USA. pp. 1578–1585. doi:10.1109/IJCNN.2017.7966039.

Wu, N., Green, B., Ben, X., O'Banion, S., 2020. Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. arXiv:2001.08317.

Yan, S., Shao, H., Xiao, Y., Liu, B., Wan, J., 2023. Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises. Robotics and Computer-Integrated Manufacturing 79, 102441. doi:10.1016/j.rcim.2022.102441.

Yang, C., Lai, Z., Wang, Y., Lan, S., Wang, L., Zhu, L., 2023. A Novel Bearing Fault Diagnosis Method based on Stacked Autoencoder and End-edge Collaboration, in: 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, Rio de Janeiro, Brazil. pp. 393–398. doi:10.1109/CSCWD57460.2023.10152598.

Yang, C., Wang, Y., Lan, S., Wang, L., Shen, W., Huang, G.Q., 2022. Cloud-edge-device collaboration mechanisms of deep learning models for smart robots in mass personalization. Robotics and Computer-Integrated Manufacturing 77, 102351. doi:10.1016/j.rcim.2022.102351.

Ye, Z., Cai, Z., Si, S., Zhang, S., Yang, H., 2020. Competing Failure Modeling for Performance Analysis of Automated Manufacturing Systems With Serial Structures and Imperfect Quality Inspection. IEEE Transactions on Industrial Informatics 16, 6476–6486. doi:10.1109/TII.2020.2967030.

Yeom, S.K., Seegerer, P., Lapuschkin, S., Binder, A., Wiedemann, S., Müller, K.R., Samek, W., 2021. Pruning by explaining: A novel criterion for deep neural network pruning. Pattern Recognition 115, 107899. doi:10.1016/j.patcog.2021.107899.

Yin, T., Lu, N., Guo, G., Lei, Y., Wang, S., Guan, X., 2023. Knowledge and data dual-driven transfer network for industrial robot fault diagnosis. Mechanical Systems and Signal Processing 182, 109597. doi:10.1016/j.ymssp.2022.109597.

Yu, K., Lin, T.R., Ma, H., Li, X., Li, X., 2021. A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. Mechanical Systems and Signal Processing 146, 107043. doi:10.1016/j.ymssp.2020.107043.

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., Xu, B., 2022. TS2Vec: Towards Universal Representation of Time Series. arXiv:2106.10466.

Zeng, A., Liu, X., Du, Z., Wang, Z., Lai, H., Ding, M., Yang, Z., Xu, Y., Zheng, W., Xia, X., Tam, W.L., Ma, Z., Xue, Y., Zhai, J., Chen, W., Zhang, P., Dong, Y., Tang, J., 2023. GLM-130B: An Open Bilingual Pre-trained Model. arXiv:2210.02414.

Zhang, G., Huang, M., Chen, G., Li, J., Liu, Y., He, J., Zheng, Y., Tang, S., Cui, H., 2023a. Design and optimization of fluid lubricated bearings operated with extreme working performances-A comprehensive review. International Journal of Extreme Manufacturing doi:10.1088/2631-7990/ad1825.

Zhang, J., Zhang, K., An, Y., Luo, H., Yin, S., 2023b. An Integrated Multitasking Intelligent Bearing Fault Diagnosis Scheme Based on Representation Learning Under Imbalanced Sample Condition. IEEE Transactions on Neural Networks and Learning Systems , 1–12doi:10.1109/TNNLS.2022.3232147.

Zhao, C., Shen, W., 2023. Mutual-assistance semisupervised domain generalization network for intelligent fault diagnosis under unseen working conditions. Mechanical Systems and Signal Processing 189, 110074. doi:10.1016/j.ymssp.2022.110074.

Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., Chen, X., 2021. Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study. IEEE Transactions on Instrumentation and Measurement 70, 1–28. doi:10.1109/TIM.2021.3116309.

Zhuo, Y., Ge, Z., 2021. Auxiliary Information-Guided Industrial Data Augmentation for Any-Shot Fault Learning and Diagnosis. IEEE Transactions on Industrial Informatics 17, 7535–7545. doi:10.1109/TII.2021.3053106.