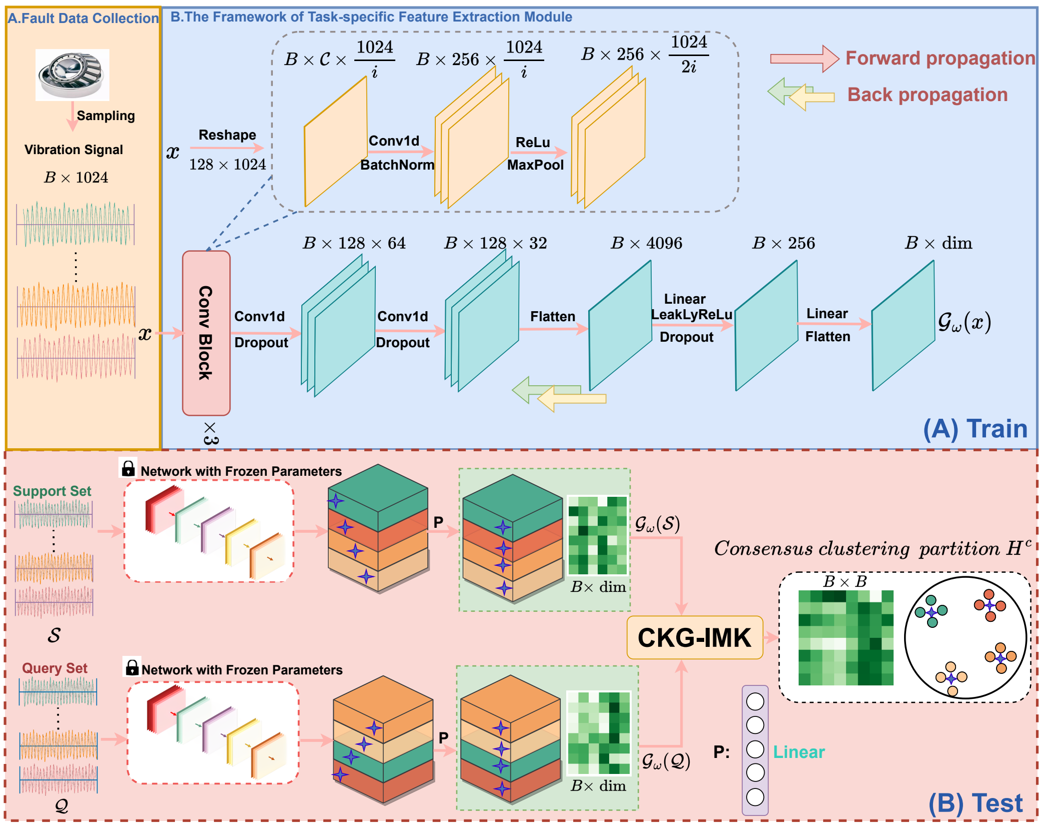
Graphical Abstract

Category Knowledge-Guided Few-Shot Bearing Fault Diagnosis

Feng Zhan, Lingkai Hu, Wenkai Huang, Yikai Dong



Highlights

Category Knowledge-Guided Few-Shot Bearing Fault Diagnosis

Feng Zhan, Lingkai Hu, Wenkai Huang, Yikai Dong

- Analyzing and Accurately Locating Pertinent Features for Novel Tasks.
- The Category Knowledge-Guided (CKG) Framework is Introduced in This Study.
- Tackling Cold-Start Cross-Domain Rotating Machinery Fault Diagnosis and Early Fault Detection.
- Attaining Exceptional Clustering Performance with High Sensitivity for Unknown Samples.

Category Knowledge-Guided Few-Shot Bearing Fault Diagnosis

Feng Zhan[a,*], Lingkai Hu[a,*], Wenkai Huang[a,**], Yikai Dong[a]

[a]*School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, 510006, China*

Abstract

Real-time bearing fault diagnosis is crucial for ensuring the safety and reliability of complex industrial systems. Few-shot learning (FSL) has shown promise in automatically extracting features from severe fault signals and accurately identifying them. However, real-world challenges like data scarcity and environmental noise hinder existing methods from detecting early-stage faults effectively. These challenges arise because current FSL approaches do not adequately consider inter-class correlations in noisy environments, limiting their ability to generalize known features to new classes. Thus, there is an urgent need for an FSL approach that can leverage inter-class correlations to address data scarcity and environmental complexities, enabling early-stage fault detection in few-shot scenarios. This paper presents a novel category-knowledge-guided model tailored for few-shot multitask scenarios. By utilizing attribute information from base category samples and similarities between new category samples, the model quickly establishes mapping relationships in unseen tasks, enhancing its generalization for early stage fault diagnosis and multitask settings. This model enables rapid and accurate FSL fault diagnosis under unknown operating conditions. Experimental results on the Case Western Reserve University bearing dataset and the Early Mild Fault Traction Motor bearing dataset demonstrate superior performance compared to state-of-the-art FSL and transfer learning approaches.

Keywords: Few-Shot Learning, Fault Diagnosis, Early-stage Fault Detection, Knowledge-Guide.

*Co-first author.

**Corresponding author.

Email address: `smallkat@gzhu.edu.cn` (Wenkai Huang)

1. Introduction

Mechanical equipment failure incidents typically result from the successive occurrence of structural damage, malfunctions, and eventual failure in functional capabilities, often leading to severe human casualties and economic losses[1, 2, 3]. In order to prevent such incidents, a range of intelligent fault diagnosis techniques has been developed, including deep autoencoders (DAE) [4], convolutional neural networks (CNN) [5], recurrent neural networks (RNN), long short-term memory networks (LSTM) [6], generative adversarial networks (GAN)[7], and graph neural networks (GNN) [8]. However, these methods are still constrained by the laborious process of manual parameter tuning and their costly computational resource requirements. With the advancement of modern industry, the demand for fault diagnosis has expanded from basic components (e.g., bearings, gears, and rotating shafts) to large-scale equipment (e.g., wind turbines, aviation engines, high-speed trains) [9, 10, 11, 12]. This has given rise to complex fault diagnosis scenarios characterized by the following: 1). Data scarcity: It is challenging to collect data from devices in faulty or failure states [13, 14] due to safety concerns in safety-critical systems; 2). Varied operating conditions: Real-world data can originate from various operating conditions [15, 16, 17, 18], including variations in speed and workload; 3). Poor quality of training data: Complex environmental conditions during operation make it difficult to substitute simulated laboratory data for real data [19] in generating well-performing models; 4). Model efficiency: Most deep learning models have grown deeper and more complex for feature extraction [20, 21], making it challenging to deploy them in practical production environments; and 5). Early fault diagnosis: Subtle differences between normal and early fault features pose challenges in extracting knowledge that reflects early-stage bearing faults, rendering early fault diagnosis highly challenging [22]. Hence, there is a pressing need for an approach that automates hyperparameter optimization, operates with limited training data, exhibits high sensitivity, and possesses strong generalization capabilities to meet the demand for rapid and accurate fault diagnosis in practical settings. Fortunately, FSL has emerged as a practical solution [23].

FSL [24, 25] is a widely recognized fault diagnosis approach, allowing rapid adaptation to unforeseen tasks with only a limited set of available samples. In essence, researchers do need not to be overly concerned with data volume, and users are not compelled to excessively rely on expert experience [26]. Hence, the application of FSL to bearing fault diagnosis demonstrates rationality and holds promising research prospects. However, FSL also has some evident limitations. First, the scarcity of data can lead to model overfitting [27], thus affecting diagnostic accuracy [28]. Additionally, FSL typically utilizes only a limited dataset [23], resulting in weaker generalization capabilities for many models, especially in the context of recognizing new categories [29]. When faced with new and complex operating conditions, such as variable loads and speeds [23, 30], the distribution of collected vibration signals may also undergo changes. This can result in differences between the source and the target domains [31], thereby impacting the generalization of pretrained models. Particularly in real-world industrial applications, work conditions frequently vary to accommodate intricate production plans and are often accompanied by various environmental noises. These variations can cause the same fault characteristics to manifest differently in different working conditions. Consequently, in situations where samples from the source and target domains are limited, it is challenging to apply pretrained fault diagnosis models directly to cross-domain fault diagnosis tasks [32]. The essence of FSL lies in leveraging acquired knowledge to learn new categories under data scarcity. Existing FSL methods primarily fall into three main branches: optimization-based, metric-based, and deep transfer learning-based methods.

The purpose of optimization-based methods is to learn a well-initialized set of parameters through extensive prior experience with similar tasks, allowing rapid adaptation to new tasks within a few optimization steps. Among these methods, model-agnostic meta-learning (MAML), proposed by Finn et al. [33], stands out as one of the most representative approaches. In the field of fault diagnosis, research has shown that methods combining MAML with two-dimensional CNNs (2D-CNN) [34] and methods integrating MAML with multilabel CNNs (MLCNN) [35] both demonstrate the practicality of MAML in addressing few-shot fault diagnosis challenges. However, while optimization-based meta-learning methods establish both inner and outer learning processes, they require the specification of numerous hyperparameters, leading to increased time consumption and computational intensity.

The core objective of metric-based methods is to learn how to measure the distance between the support set and the query set in an embedding space. Classical metric-based methods include Siamese networks, prototypical networks, matching networks, and relation networks. In [36], Wang and colleagues proposed a fault contrast deep neural network based on Siamese networks, enhancing model metric performance by introducing label-smoothing techniques. To address domain shift issues, a novel approach combining supervised domain adaptation with prototypical networks for fault diagnosis was introduced [37]. Researchers like Feng [38] designed semi-supervised machine learning networks, incorporating squeeze and excitation attention mechanisms to correct prototypes using unlabeled samples, thus improving fault recognition performance. Zhang and colleagues proposed an iterative matching network[39], achieving more comprehensive utilization of unlabeled samples by employing a selective sample reuse strategy. Further, in the realm of relation networks for fault diagnosis, Wang and colleagues introduced reinforce relation networks [40] for bearing fault diagnosis, enhancing network classification performance through label smoothing and the Adabound algorithm. However, training methods solely based on metrics typically teach the model to focus on the relative similarity information between sample groups while overlooking the attribute information of each specific category [41]. This implies that the provided labeled source data are not fully utilized.

Methods based on deep transfer learning are designed to apply knowledge or patterns learned from a source domain or task to different but related target domains or problems [42]. Currently, research on fault diagnosis using transfer learning can be categorized into three classes: instance-based (including instance selection or reweighting strategies), model-based (sharing neural network structures and parameters, including fine-tuning), and feature-based methods (sharing or learning common feature representations, including difference-based domain adaptation, adversarial-based domain adaptation, and reconstruction-based domain adaptation) [43]. The performance of instance-based methods depends on the quantity of target instances and their similarity to source instances [31]; however, such methods still face the challenge of cold-start problems [44]. Model-based methods heavily rely on the quantity and quality of training data, and the trained diagnostic models require further adjustments to accommodate target domain data to accomplish the target fault diagnosis task [45, 46]. Meanwhile, feature-based methods are susceptible to negative transfer, require an assessment of transferability, and often struggle to achieve robust generalization in target fault diagnosis tasks [47].

According to the literature review [22, 44, 48], existing deep transfer learning-based fault diagnosis methods primarily focus on permanent or severe faults. When the fault labels in the target domain differ from those in the source domain, this label inconsistency issue increases the complexity of knowledge transfer [49]. Additionally, optimization-based methods fail to adequately discover and capture the attribute information of each specific category. These methods lack the ability to precisely identify features relevant to the new task in the source task, making it challenging to effectively detect early faults. However, many severe faults in rolling bearings actually evolve from mild early-stage faults [50, 51]. Therefore, early fault diagnosis is crucial for effective component maintenance or replacement. Nonetheless, due to the small magnitude of early faults, they are susceptible to interference from strong background noise. Traditional FSL methods have failed to explicitly convey the attribute information of relevant classes, adding to the complexity of early fault diagnosis [52]. Moreover, the subtle differences between normal features and early fault features pose challenges in extracting features that reflect early-stage faults, making early fault diagnosis in rolling bearings extremely challenging.

To analyze and precisely locate salient features relevant to new tasks [53, 54, 55] as well as to explicitly leverage inter-category relationships [56], encompassing commonalities for facilitating generalization across related categories [57] and uniqueness to reduce misclassification among similar categories, this paper introduces a novel category knowledge-guided (CKG) framework for cold-start cross-domain rotating machinery fault diagnosis and early fault detection. Specifically, this research adopts an incomplete multikernel clustering matrix to quantify the inter-category correlations, thus generating a consensus clustering matrix to guide the joint learning for more robust tasks. This capability allows CKG to simultaneously handle multiple support categories, unlike the majority of existing methods, which require repeated meta-learning for each individual category. Further, the proposed framework is extended to more challenging tasks in the realm of FSL, including cold-start scenarios and early fault diagnosis, and its outstanding performance in data-scarce and complex environments is demonstrated. Finally, a theoretical analysis of the category knowledge-guided incomplete multikernel clustering(CKG-IMK) algorithm is provided, and comprehensive experiments and analyses are conducted to validate the stability and exceptional clustering performance of the proposed method. The primary contributions of this study are as follows:

1. A novel multiclass-based feature learning model is proposed for the

analysis and localization of task-relevant features.

2. The introduced incomplete multikernel clustering method can effectively leverage the relative similarities among different categories in multitask scenarios, achieving high sensitivity clustering performance for unknown samples. Additionally, this research theoretically investigates the effectiveness of the proposed CKG framework in terms of clustering generalization error.
3. Dependency on data volume and label quality is reduced, thus reducing the cost of data collection and processing while significantly enhancing the sensitivity and reliability of early diagnosis. Further, the model's robustness and interpretability are improved, enabling it to effectively address various challenges, including noise, anomalies, and uncertainties.

2. Theoretical foundation

Guided by category knowledge, this paper primarily explores the theories of FSL and incomplete multikernel clustering [58]. In this section, a brief introduction to FSL [59] learning methods and the theoretical research on multiple kernel K-means (MKKM) [60] is provided.

2.1. Few-shot Learning

FSL is a machine learning approach that leverages prior knowledge from multiple related tasks [61] (the support set, denoted as $\mathcal{S}$) to enhance performance on new target tasks (the query set, denoted as $\mathcal{Q}$). This approach allows the training of robust models that can classify a small sample of annotated data, recognize new classes, and improve the model's generalization and portability. Achieving this objective necessitates the effective utilization of prior knowledge and limited data to mitigate model bias and variance while avoiding overfitting and underfitting. Simultaneously, considerations must be made for factors like data quality, distribution, noise, heterogeneity, as well as the complexity, diversity, and dynamics of tasks in meta-test scenarios.

In fault classification tasks, faults occurring in the same equipment typically exhibit a certain degree of similarity, and faults in the same category of equipment tend to share similar features [62]. Metric-based FSL methods capitalize on this characteristic by learning methods to represent the similarity between support and query samples in an embedding space to identify unknown samples. Specifically, this approach effectively measures the similarity between support and query samples in the embedding space. By learning similarity metrics in this embedding space, the model can accurately classify unknown samples into support categories that are similar to them [63]. This demonstrates the method's ability to handle various operating conditions or fault categories in practical applications, showcasing its feasibility in fault classification tasks.

Metric-based approaches utilize certain distance or similarity measures to compare samples in the test set with those in the support set (a limited amount of labeled data). However, metric-based FSL methods suffer from a drawback in that features unrelated to the classification task might mislead the model. Additionally, due to the limited number of samples in the support set, they often fail to identify the target features relevant to the task [64]. A model that could extract fault feature information more comprehensively from the data and explicitly learn significant features related to the task would be more reliable model, reducing attention to task-irrelevant information and thereby enhancing the performance of fault classification tasks [65].

2.2. Multiple Kernel K-Means (MKKM)

As a non-linear technique, multikernel methods [66] can handle linearly inseparable data and achieve satisfactory clustering results in high-dimensional spaces [67]. However, in the context of multitask settings, a single kernel function may not effectively handle heterogeneous data. To address this, the concept of multikernel subspace clustering (MKSC) [68] has been introduced. The principle behind MKSC involves extracting more information from the data using various kernel functions, thereby enhancing clustering performance. Given a set of observed data $\{x_i\}_{i=1}^{n}$ and the kernel mapping $\mathcal{F}(\cdot)$, the objective of MKKM is to partition the samples into k clusters by minimizing the sum of squares loss. This objective can be expressed as follows:

$$\min_{\mathcal{W},\widehat{\mathcal{C}}} \sum_{i=1}^{n} \sum_{j=1}^{m} \|\mathcal{F}(x_i) - \widehat{\mathcal{C}}_j\|_F^2$$

$$s.t. \sum_{j=1}^{m} \mathcal{W}_{ij} = 1 \tag{1}$$

where $\mathcal{W} \in \{0,1\}^{n\times k}$ represents the clustering assignments for each sample, and $\widehat{\mathcal{C}}_j$ denotes the centroid of the j-th cluster.

In most cases, $\mathcal{F}(x_i) \in \mathcal{R}^d$, where $d \gg n$ or even infinite. Therefore, Eq.(1) cannot be directly optimized. Consequently, it is equivalently rewritten in matrix-vector form a follows:

$$\min_{\mathcal{W}} \mathrm{Tr}(\mathcal{K}_\xi) - \mathrm{Tr}(\xi^{1/2}\mathcal{W}^T\mathcal{K}_\xi\mathcal{W}\xi^{1/2}) \tag{2}$$

where, $\mathcal{K}_\xi^{ij}(x_i,x_j) = \mathcal{F}(x_i)^T\mathcal{F}(x_j)$, $\xi = diag([n_1^{-1}, n_2^{-1}, \cdots, n_k^{-1}])$,$n_j = \sum_{i=1}^n \mathcal{W}_{ij}$, $\mathrm{Tr}(\cdot)$ denotes the trace norm, and ξ_j represents the fundamental kernel for the j-th weight. Discrete $\mathcal{W}$ makes Eq.(2) challenging to solve, and a common technique is to relax it, allowing for arbitrary values. MKKM can simultaneously learn ξ and the clustering assignment matrix $\mathrm{H}^{\mathcal{C}}$ by defining $\mathrm{H}^{\mathcal{C}} = \mathcal{W}\xi^{-1}$. The aforementioned problem can thus be transformed as follows:

$$\begin{aligned} &\min_{\xi,\mathrm{H}^{\mathcal{C}}} \mathrm{Tr}(\mathcal{K}_\xi(\mathrm{I}_n - \mathrm{H}^{\mathcal{C}}(\mathrm{H}^{\mathcal{C}})^T)) \\ &s.t.\ \mathrm{H}^{\mathcal{C}} \in \mathbb{R}^{n\times k}, (\mathrm{H}^{\mathcal{C}})^T\mathrm{H}^{\mathcal{C}} = \mathrm{I}_k, \|\xi\| \geq 0. \end{aligned} \tag{3}$$

Existing algorithms typically solve Eq.(3) through alternating optimization of $\mathrm{H}^{\mathcal{C}}$ and ξ: (i) Fixing ξ to optimize $\mathrm{H}^{\mathcal{C}}$. For a specific kernel coefficient ξ, optimizing $\mathrm{H}^{\mathcal{C}}$ in Eq.(3) is equivalent to the following Eq.(4):

$$\begin{aligned} &\min_{\xi,\mathrm{H}^{\mathcal{C}}} \mathrm{Tr}(\mathcal{K}_\xi(\mathrm{I}_n - \mathrm{H}^{\mathcal{C}}(\mathrm{H}^{\mathcal{C}})^T)) \\ &s.t.\ \mathrm{H}^{\mathcal{C}} \in \mathbb{R}^{n\times k}, (\mathrm{H}^{\mathcal{C}})^T\mathrm{H}^{\mathcal{C}} = \mathrm{I}_k, \|\xi\| \geq 0 \end{aligned} \tag{4}$$

Eq.(4) is a classic kernel k-means equation that can be easily optimized. An optimized kernel matrix $\mathcal{K}_\xi$ is parameterized in the following form: $\mathcal{K}_\xi = \sum_{j=1}^k \xi_j^2\mathcal{K}_\xi^j$, where $\{\mathcal{K}_\xi^j\}_{j=1}^k$ represents a set of precomputed kernel matrices. (ii) Fix $\mathrm{H}^{\mathcal{C}}$ to optimize ξ. For a specific $\mathrm{H}^{\mathcal{C}}$, the optimization of ξ in Eq.(4) simplifies to the following:

$$\begin{aligned} &\min_{\xi} \sum_{j=1}^m \xi_j^2 \mathrm{Tr}(\mathcal{K}_\xi(\mathrm{I}_n - \mathrm{H}^{\mathcal{C}}(\mathrm{H}^{\mathcal{C}})^T)) \\ &s.t.\ \mathrm{H}^{\mathcal{C}} \in \mathbb{R}^{n\times k}, (\mathrm{H}^{\mathcal{C}})^T\mathrm{H}^{\mathcal{C}} = \mathrm{I}_k, \|\xi\| \geq 0. \end{aligned} \tag{5}$$

Algorithm 1 presents a detailed optimization procedure for MKKM, where $\mathrm{H}^{\mathcal{C}}$ and ξ are alternately optimized until convergence.

Algorithm 1 Multiple Kernel K-Means

Require: $\{\mathcal{K}_{\xi}^{j}\}_{j=1}^{m}, k, t = 1$
  Initialization $\xi = 1/\sqrt{m}$
  repeat
    Compute $(\mathrm{H}^{\mathcal{C}})^t$ in Eq. (2) with $\mathcal{K}_{\xi^t} = \sum_{j=1}^{m}(\xi_j^t)^2\mathcal{K}_{\xi}^{j}$
    Update ξ^t and $(\mathrm{H}^{\mathcal{C}})^t$ in Eq.(3)
    $t \leftarrow t + 1$
  until $|\xi^{(t+1)} - \xi^t| \leq e^{-4}$

3. Methodology

In this section, a category knowledge-guided incomplete multikernel clustering few-shot learning method is introduced, which is applicable to the problem of few-shot bearing fault diagnosis under various limited data conditions. This approach has demonstrated high sensitivity and accuracy in fault diagnosis performance on the Case Western Reserve University bearing dataset (CWRU) [69] for permanent and severe faults as well as on the Early Mild Fault Traction Motor bearing dataset (EMF-TM).

3.1. Problem definition

In contrast to traditional machine learning approaches, in FSL, training samples are treated as tasks or events rather than mere data instances. In the context of few-shot classification problems, this is typically formalized as an N-way K-shot classification problem. In this problem, the model is required to acquire K labeled fault samples from N different categories and accurately classify unlabeled faults [59, 70].

Specifically, given a dataset $D = \{(x_i, y_i), y_i \in \mathcal{L}\}_{i=1}^{I}$, D is divided into the meta-training set $D^{train} = \{(x_i, y_i), y_i \in \mathcal{L}^{train}\}_{i=1}^{I^{train}}$ and the meta-test set $D^{test} = \{(\widetilde{x}_i, \widetilde{y}_i), \widetilde{y}_i \in L^{test}\}_{i=1}^{I^{test}}$, where (x_i, y_i) represents the original features and label information of the i-th bearing sample in the meta-training set, and $D^{train} \cup D^{test} = D$, $L^{train} \cup L^{test} = L$. There is no intersection between the meta-training set and the meta-test set ($D^{train} \cap D^{test} = \emptyset$).

The FSL algorithm requires learning general meta-knowledge from multiple training sets to acquire new tasks. In accordance with the prior literature, we consider a meta-training set, denoted as T, comprising tasks $\mathfrak{T} = \{\mathfrak{T}^1, \mathfrak{T}^2, \cdots, \mathfrak{T}^T\}$. To construct each task $\mathfrak{T}^j$, we randomly select N categories from D^{train}, with each category containing M samples. Within each selected category, the M samples are further divided into two sets, each containing K and $M-K$ bearing fault samples, respectively. These sets are referred to as the support set $\mathcal{S}^j = \{(x_i^j, y_i^j), y_i^j \in L^train\}_{i=1}^{N\times K}$ and the query set $\mathcal{Q}^j = \{(\widehat{x_i^j}, \widehat{y_i^j}), \widehat{y_i^j} \in L^{train}\}_{i=1}^{N\times(M-k)}$. Similarly, D^{test} is divided into the labeled support set $\mathcal{S}^\zeta = \{(x_i^\zeta, y_i^\zeta), y_i^\zeta \in L^{test}\}_{i=1}^{N\times K}$ and the unlabeled query set $\mathcal{Q}^\zeta = (\widehat{x_j^\zeta})_{j=1}^{N\times(M-k)}$, with no intersection between these datasets($\mathcal{S}^\zeta \cap \mathcal{Q}^\zeta = \emptyset$).

During the training phase of FSL, the model is initially trained using the meta-training set, and the performance of the meta-trained model is evaluated using the meta-test set. The advantage of this task training strategy is that it enables the model to demonstrate good generalization performance on entirely new class samples. Consequently, our model can better adapt to various tasks and data, thereby enhancing its applicability and performance in practical applications.

3.2. Data preprocessing

To strike a balance between training efficiency and accuracy while showcasing the excellent performance of the FSL network and its adaptability to real industrial production environments, this study applied a random segmentation operation to the original signal sequences, dividing them into multiple data segments, each containing 1024 sampling points. There were only 500 samples per class. Data augmentation techniques were not employed in this research to increase the dataset size, as doing so might introduce signal redundancy between different data segments and consequently affect the reliability of experimental results [5].

It is worth noting that, before feeding the data into the model, data standardization was conducted. Specifically, a global standardization method, known as the z-score, was used to eliminate scale differences in the data, thereby aligning the bearing signal sequences with a standard Gaussian distribution.

3.3. Model architecture of Category-Knowledge-Guided

The proposed model consists of three modules, including a task-related feature extraction module, a category knowledge-guided incomplete multikernel clustering(CKG-IMK) algorithm, and the algorithm's extension.

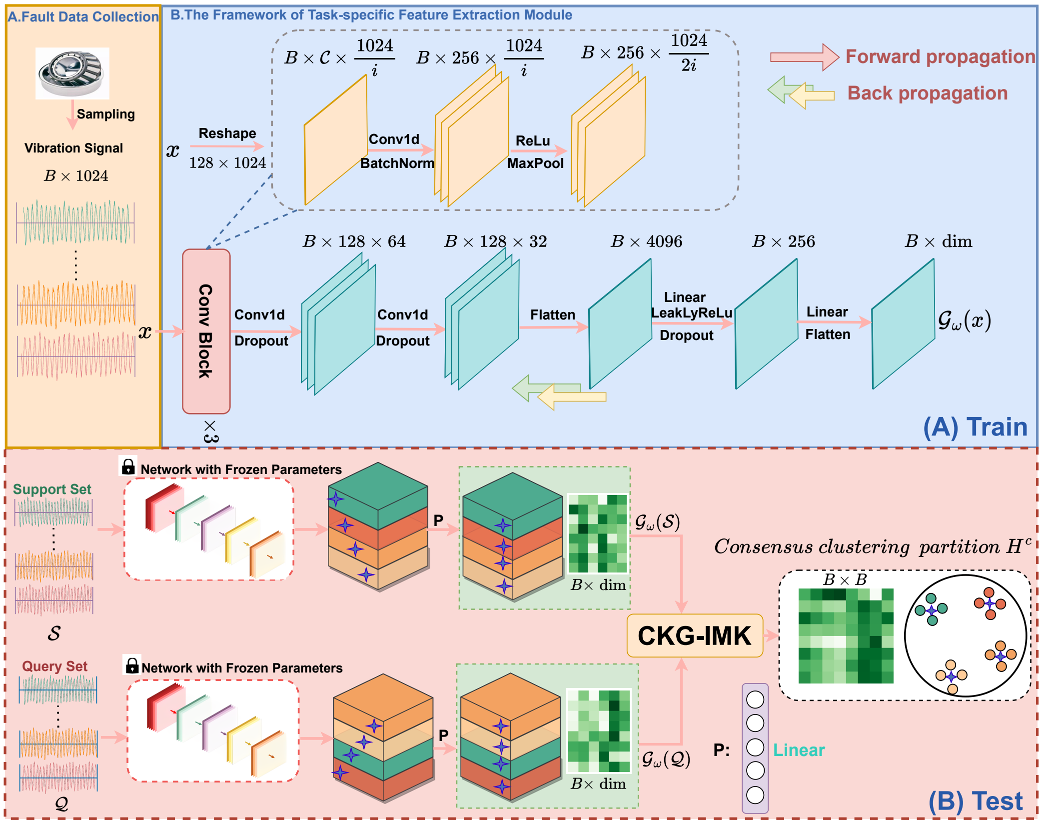


Fig. 1. Structure diagram of the CKG framework.

3.3.1. *Task-specific feature extraction module*

A novel task-related feature extraction module has been proposed in this study (Fig.1), aimed at extracting complex and salient fault-related features while addressing the issue of overfitting. This objective is accomplished through strategic enhancements to the model architecture and optimizations within the network layers, which increase the model's depth while simultaneously reducing the overall number of parameters. Fig.1 illustrates the workflow of the proposed CKG framework, while Table 1 provides a more comprehensive breakdown of the task-related feature extraction module.

The module, which is shown in Fig.1, is employed for the processing of preprocessed fault signals, denoted as $x_i \in \mathbb{R}^{C\times1024}$, where C represents the channels of input features. The feature calibration process begins with the application of a basic 1×1 convolutional layer (BC) feature extractor. The design of this layer is aimed at efficiently capturing complex patterns within the preprocessed fault signals while preserving crucial features associated with local faults. This facilitates the learning of local features in subsequent multiclassification tasks.

Table 1
Details of the task-specific feature extraction module.

Layer	Filter	Kernel/Stride	Padding	Input size	Output size
Conv1d/BN/ReLU	256	(3,3)/1	1	(B,1,1024)	(B,256,1024)
MaxPool1d	/	(2,2)/2	/	(B,256,1024)	(B,256,512)
Conv1d/BN/ReLU	256	(3,3)/1	1	(B,256,512)	(B,256,512)
MaxPool1d	/	(2,2)/2	/	(B,256,512)	(B,256,256)
Conv1d/BN/ReLU	256	(3,3)/1	1	(B,256,256)	(B,256,256)
MaxPool1d	/	(2,2)/2	/	(B,256,256)	(B,256,128)
Conv1d/BN/ReLU	128	(3,3)/1	1	(B,256,128)	(B,128,128)
MaxPool1d	/	(2,2)/2	/	(B,128,128)	(B,128,64)
Dropout	/	/	/	(B,128,64)	(B,128,64)
Conv1d/BN/ReLU	128	(3,3)/1	1	(B,128,64)	(B,128,64)
MaxPool1d	/	(2,2)/2	/	(B,128,64)	(B,128,32)
Dropout	/	/	/	(B,128,32)	(B,128,32)
Flatten	/	/	/	(B,128,32)	(B,4096)
Linear/LeaklyReLU	256	/	/	(B,4096)	(B,256)
Dropout	/	/	/	(B,256)	(B,256)
Linear	/	/	/	(B,256)	(B,dim)
Flatten	/	/	/	(B,dim)	(B,dim)

During the process of task-related feature extraction, the proposed model employs a series of cascading CNN layers. Each layer consists of batch normalization, rectified linear unit (ReLU) activation, and max-pooling operations. Through these layers, the spatial dimensions of the feature maps gradually decrease, while the depth increases, thereby incrementally enhancing the representational capacity of the features.

To further enhance model performance, batch normalization (BN) modules are introduced in subsequent layers. These modules not only preserve fundamental local features but also alleviate the issues of gradient vanishing or explosion within the constraints of limited model parameters. Additionally, BN modules play a crucial role in enhancing the model's robustness, aiding in reducing overfitting and improving generalization capabilities. For efficient computation, ReLU is employed as the activation function. This function ensures that the model introduces non-linearity in feature extraction, which is crucial for capturing complex fault-related patterns.

It is worth noting that the max-pooling operation, while reducing spatial dimensions, retains critical information. To prevent overfitting, strategic dropout layers are strategically inserted after certain convolutional layers, with a dropout rate of 0.5 employed in all dropout layers. This regularization technique helps prevent the model from relying excessively on specific neurons, thereby enhancing generalization to unseen data. Following the convolutional layers, the feature maps are flattened in preparation for input to fully connected layers. These fully connected layers further abstract and refine the features extracted from the preceding layers. A leaky ReLU activation function is introduced after the linear layers. This is done to avoid the issue of neuron death, ensuring a small gradient value even when the input is negative, enabling neurons to continue updating during training. This helps mitigate the problem of gradient vanishing and can enhance the model's convergence speed.

After the linear layers, another dropout layer is applied to prevent overfitting at this stage. The final linear layer maps the learned features to the output classes, thereby generating accurate fault diagnosis predictions. Our proposed model is designed to effectively capture both local and higher-level features, thus enhancing early fault diagnosis performance.

In summary, the proposed feature extraction module strikes a balance between module depth, parameter efficiency, and feature representation. The combined application of convolutional layers, batch normalization, ReLU activation, and dropout regularization enables the module to capture complex fault-related patterns while preventing overfitting. This innovative module exhibits significant potential for enhancing the field of fault diagnosis.

3.3.2. Category Knowledge-Guided Incomplete Multikernel Clustering Algorithm

To jointly optimize the optimal kernel, maximum-margin hyperplane, and optimal clustering labels, a CKG-IMK algorithm is proposed to construct a consensus partition. The detailed specifics of this algorithm are presented below.

Assuming that the meta-training support set$\{(x_i, y_i), y_i \in L^{train}\}_{i=1}^{n}$ comprises a collection of n samples, and $\{\mathcal{G}(\cdot)\}_{\varrho=1}^{m} : x_i \in X_{\mathcal{S}}^{test} \Rrightarrow \{\mathcal{H}\}_{\varrho=1}^{m}$ represents an encoder that maps different inputs x_i to Hilbert spaces $\{\mathcal{H}\}_{\varrho=1}^{m}$, yielding m multikernel observations $[(z_i)_1, (z_i)_2, \cdots, (z_i)_m]_{i=1}^{n} \in \mathbb{R}^n$, where $(z_i)_\varrho$ denotes the ϱ-th base kernel for the i-th sample. These m base kernels are obtained through the encoder $\mathcal{G}(\cdot)_{\varrho=1}^{m}$. These concepts can be precisely formulated mathematically as follows:

$$\{\mathcal{G}(x_i)\}_{\varrho=1}^{m} = [\xi_1(z_i)_1, \xi_2(z_i)_2, \cdots, \xi_m(z_i)_m] \tag{6}$$

where $\xi = [\xi_1, \xi_2, \cdots, \xi_m]$ represents a matrix containing normalization coefficients. These coefficients are adaptively optimized during the training process. ξ is capable of normalizing the multikernel observations $(z_i)_\varrho$, enabling local kernel alignment.

Hence, it can be assumed that the m basic kernels $(z_i)_\varrho$ share a latent proxy $h_i \in \mathbb{R}^k$ to represent each kernel sample z_i in the latent embedding space. Specifically, the m kernel samples $\{(z_i)_\varrho\}_{\varrho=1}^{m}$ can be represented through the latent proxies h_j and the corresponding mapping matrices $\{\ddot{\mathcal{P}}_\varrho\}_{\varrho=1}^{m} \in \mathbb{R}^{n\times k}$. These concepts can be precisely formulated mathematically as follows:

$$\begin{aligned} &\min_{\ddot{\mathcal{P}}_\varrho, h_i, \xi} \sum_{i=1}^{n} \sum_{\varrho=1}^{m} \|\xi_\varrho (z_i)_\varrho \ddot{\mathcal{P}}_\varrho - h_i\|_F^2 \\ &s.t.\ \|\xi\|_\varrho \geq 0, \ddot{\mathcal{P}}_\varrho^T \ddot{\mathcal{P}}_\varrho = \mathrm{I}_k. \end{aligned} \tag{7}$$

Based on the definition of the base kernel $(z_i)_\varrho$ and the fact that the samples x_i can be transformed into $\{\mathcal{G}(x_i)\}_{\varrho=1}^{m}$ through the encoder $\{\mathcal{G}(\cdot)\}_{\varrho=1}^{m}$, the kernel function can be expressed as follows:

$$\kappa_{\mathcal{G}}(x_i, x_i) = \{\mathcal{G}(x_i)\}_{\varrho=1}^{m} (\{\mathcal{G}(x_i)\}_{\varrho=1}^{m})^{\mathrm{T}} = \sum_{\varrho=1}^{m} \xi_\varrho^2 \kappa_\varrho((z_i)_\varrho, ((z_i)_\varrho). \tag{8}$$

Subsequently, the kernel matrix $\mathcal{K}_{\mathcal{G}}^i$ is computed by employing the defined kernel function $\kappa_{\mathcal{G}}(x_i, x_i)$. $\mathcal{K}_{\mathcal{G}}^i$ not only ensures the existence of potential partitions within a low-rank space but also enables the integration of complementary information among multiple base kernels, resulting in a consensus clustering partition, denoted as $\mathrm{H}^{\mathcal{C}}$. The aforementioned concept can be realized as follows:

$$\begin{aligned} &\min_{\ddot{\mathcal{P}}_\varrho, \mathrm{H}^{\mathcal{C}}, \xi} \sum_{\varrho=1}^{m} \|\xi_\varrho (\mathcal{K}_{\mathcal{G}}^i) \ddot{\mathcal{P}}_\varrho - \mathrm{H}^{\mathcal{C}}\|_F^2 \\ &s.t.\ \|\xi\|_\varrho \geq 0, \ddot{\mathcal{P}}_\varrho^T \ddot{\mathcal{P}}_\varrho = \mathrm{I}_k. \end{aligned} \tag{9}$$

By solving Eq.(9), one can infer a latent consensus partition, denoted as $\mathrm{H}^{\mathcal{C}}$, which fundamentally characterizes the data and uncovers the underlying structures shared by different kernels. Initially, a consensus partition matrix

$\mathrm{H}^{\mathcal{C}}$ is derived from the feature vectors $\{\mathrm{H}_{\varrho}\}_{\varrho=1}^{m}$, and subsequently, the partially overlapping partitions are computed using the learned consensus matrix $\mathrm{H}^{\mathcal{C}}$. In this manner, these two learning processes can seamlessly intertwine, allowing them to mutually negotiate and achieve enhanced clustering. The aforementioned concept can be implemented as follows:

$$\begin{aligned}
&\max_{\mathrm{H}^{\mathcal{C}},\{\mathrm{H}_{\varrho},\ddot{\wp}_{\varrho}\}_{\varrho}^{m}} \mathrm{Tr}[(\mathrm{H}^{\mathcal{C}})^{\mathrm{T}}(\sum_{\varrho=1}^{m}\mathrm{H}_{\varrho}\ddot{\wp}_{\varrho})] \\
&s.t.\ \mathrm{H}^{\mathcal{C}} \in \mathbb{R}^{n\times k}, (\mathrm{H}^{\mathcal{C}})\mathrm{H}^{\mathcal{C}} = \mathrm{I}_k, \\
&\ddot{\wp}_{\varrho} \in \mathbb{R}^{k\times k}, \ddot{\wp}_{\varrho}^{\mathrm{T}}\ddot{\wp}_{\varrho} = \mathrm{I}_k, \\
&\mathrm{H}_{\varrho} \in \mathbb{R}^{k\times k}, \mathrm{H}_{\varrho}^{\mathrm{T}}\mathrm{H}_{\varrho} = \mathrm{I}_k
\end{aligned} \tag{10}$$

where $\mathrm{H}^{\mathcal{C}}$ and H_{ϱ} represent the consensus clustering matrix and the ϱ-th base clustering matrix, respectively, with k denoting the number of clusters and $ddot\wp_{\varrho}$ denoting the transposition matrix for the ϱ-th base, which aids in better alignment between $\mathrm{H}^{\mathcal{C}}$ and H_{ϱ}. This necessitates the imputation of all incomplete elements and the deliberate decomposition of the entire inferred similarity to facilitate clustering. This enhances the model's robustness throughout the optimization process. Ultimately, m partially incomplete base kernels $\{(z_i)_{\varrho}\}_{\varrho=1}^{m}$ are obtained, along with the clustering indicator matrix $\mathrm{H}^{\mathcal{C}}$. Let $\widehat{\mathcal{C}} = [\widehat{\mathcal{C}}_1, \widehat{\mathcal{C}}_2, \cdots, \widehat{\mathcal{C}}_k]$, where $\widehat{\mathcal{C}}_k$ represents the centroids of each cluster, to reduce redundancy and enhance the diversity of the selected base kernels. Finally, K-means is employed to minimize the reconstruction loss:

$$\mathbb{E}[\min_{y\in\{e_1,e_2,\cdots,e_k\}} \|\{\mathcal{G}(x_i)\}_{\varrho=1}^{m} - \widehat{\mathcal{C}}_y\|_F^2] \tag{11}$$

where $\{e_1, e_2, \cdots, e_k\}$ form the orthogonal bases of $\mathbb{R}^k$.

During the query phase of the meta-testing, a query set is constructed using samples $\{(x_i^{\mathcal{S}}, y_i^{\mathcal{S}}), x_i^{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}^{test}, y_i^{\mathcal{S}} \in \mathcal{L}^{test}\}_{i=1}^{n}$ from k classes, where $x_i^{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}^{test}$ represents the labeled data. Additionally, unlabeled data $\{(x_i^{\mathcal{Q}}), x_i^{\mathcal{Q}} \in \mathcal{X}_{\mathcal{Q}}^{test}\}_{i=1}^{N\times K}$ are employed as samples within the query set. Consistent with the previous approach, the kernel function is computed using unobserved elements $\{x_j^{\mathcal{Q}}\}_{j=1}^{N\times(M-K)}$:

$$\kappa_{\mathcal{G}}(x_i^{\mathcal{S}}, x_j^{\mathcal{Q}}) = \{\mathcal{G}(x_i^{\mathcal{S}})\}_{\varrho=1}^{m}(\{\mathcal{G}(x_j^{\mathcal{Q}})\}_{\varrho=1}^{m})^{\mathrm{T}} = \sum_{\varrho=1}^{m}\xi_{\varrho}^2\kappa_{\varrho}(x_i^{\mathcal{S}}, x_j^{\mathcal{Q}}). \tag{12}$$

The encoder $\{\mathcal{G}(\cdot)\}_{\varrho=1}^{m}$ is employed to encode $x_i^{\mathcal{S}}$ and $x_j^{\mathcal{Q}}$, resulting in encoding matrices $\widehat{\Theta}_i \in \mathbb{R}^{n\times c}$ and $\widehat{\delta}_j \in \mathbb{R}^{n\times c}$, respectively. $\mathcal{K}_{\mathcal{G}}^{i}$ represents the consensus clustering matrix used to measure the correlation between $\mathcal{X}_{\mathcal{S}}^{test}$ and $\mathcal{X}_{\mathcal{Q}}^{test}$, where alignment is only required for the similar samples of each data point with its nearest neighbors. The relevance aggregation algorithm is a critical step in cross-class computation for the meta-aggregation model, where the alignment of similarity between query features and support set categories is aggregated, and this can be achieved using the following equation:

$$\min_{\widehat{\Theta}_i,\ddot{\mathcal{P}}_\varrho,\mathrm{H}^{\mathcal{C}},\widehat{\delta},\mathcal{B},\xi} \sum_{\varrho=1}^{m} \|\xi_\varrho \mathcal{K}_{\mathcal{G}}^{i}\ddot{\mathcal{P}}_\varrho - \widehat{\Theta}_i\mathcal{B}\|_F^2 + \|\mathrm{H}^{\mathcal{C}} - \widehat{\delta}_j\mathcal{B}\|_F^2$$

$$s.t.\ \widehat{\delta}_j\widehat{\delta}_j^T = \mathrm{I}_k, \ddot{\mathcal{P}}_\varrho^T\ddot{\mathcal{P}}_\varrho = \mathrm{I}_k, \widehat{\Theta}_i\widehat{\Theta}_i^{\mathrm{T}} = \mathrm{I}_n, (\mathrm{H}^{\mathcal{C}})\mathrm{H}^{\mathcal{C}} = \mathrm{I}_k \tag{13}$$

Eq.(7) and (13) are combined to yield:

$$\min_{\widehat{\Theta}_i,\ddot{\mathcal{P}}_\varrho,\mathrm{H}^{\mathcal{C}},\widehat{\delta},\mathcal{B},\xi} \sum_{\varrho=1}^{m} \|\xi_\varrho \mathcal{K}_{\mathcal{G}}^{i}\ddot{\mathcal{P}}_\varrho - \mathrm{H}^{\mathcal{C}}\|_F^2 + \varpi\left(\sum_{\varrho=1}^{m} \|\xi_\varrho \mathcal{K}_{\mathcal{G}}^{i}\ddot{\mathcal{P}}_\varrho - \widehat{\Theta}_i\mathcal{B}\|_F^2 + \|\mathrm{H}^{\mathcal{C}} - \widehat{\delta}_j\mathcal{B})\|_F^2\right)$$

$$s.t.\ \widehat{\delta}_j\widehat{\delta}_j^T = \mathrm{I}_k, \ddot{\mathcal{P}}_\varrho^T\ddot{\mathcal{P}}_\varrho = \mathrm{I}_k, \widehat{\Theta}_i\widehat{\Theta}_i^{\mathrm{T}} = \mathrm{I}_n, (\mathrm{H}^{\mathcal{C}})\mathrm{H}^{\mathcal{C}} = \mathrm{I}_k \tag{14}$$

where ϖ governs the consistency of cluster centers, dimension k controls the partitioning of latent dimensions, and $\mathcal{B}$ represents the consensus clustering center matrix. Notably, Eq.(14) utilizes the consensus clustering center $\mathcal{B}$ to connect the incomplete consensus partition matrix $\mathrm{H}^{\mathcal{C}}$ with embedded cluster representations, characterizing this model as CKG. Furthermore, to ensure the aggregation of similar data within the clustering center matrix $\mathcal{B}$ and the separation of dissimilar data, guidance is drawn from global distribution information. The kernel function $\kappa_{\mathcal{F}}$ is employed to capture cross-category correlations, allowing the proposed algorithm to effectively utilize intra-cluster variations among samples and leverage inter-feature relationships for aggregated representations, thereby reducing misclassification and enhancing the model's generalization capabilities.

Moreover, for enhanced scalability of the method, balancing the redundancy information among kernels and kernel details, the final dimensions of the input central matrix are set to be the same, allowing for matrix multiplication after transposition, enabling the method to handle inputs of different sizes. This approach enhances the method's value in industrial applications. This concept can be implemented as follows:

$$\mathcal{L}'(\omega;\mathcal{K}^i_{\mathcal{G}},\mathcal{A}) = \left(\frac{\mathcal{K}^i_{\mathcal{G}} + |\mathcal{K}^i_{\mathcal{G}}|(\mathcal{J}_n - \mathcal{A}) + \mathcal{A}^2 - 2\mathcal{A}}{\mathcal{J}_n - \mathcal{A}}\right)^2 + \psi\Omega(\omega)$$
$$s.t.\ \mathcal{L}^{test} \in \mathbb{R}^{n\times k}, \mathcal{L}^{test}(\mathcal{L}^{test})^T = \mathcal{A} \in \mathbb{R}^{n\times n} \tag{15}$$

where $\mathcal{L}^{test}$ represents a one-hot encoded matrix of size $n \times k$, $\mathcal{J}_n$ is an all-ones matrix,ω represents the training parameters of the model, and k denotes the number of categories. The regularization coefficient ψ is set to 0.005, and $\Omega(\omega)$ is a penalty term used to penalize model complexity. This is because model complexity is positively correlated with the number of coefficients, and the more coefficients there are, the more complex the model becomes. To control model complexity, it is possible to reduce the number of coefficients, which means limiting the number of non-zero elements in the vector. This can be achieved by introducing constraints into the optimization problem:

$$s.t.\ \|\omega\|_2 \leq \mathbb{C} \tag{16}$$

where $\mathcal{A}_{ij} \in \{0, 1\}$, and thus the output expression of Eq.(15) is as follows:

$$\mathcal{L}'(\omega;\mathcal{K}^i_{\mathcal{G}},\mathcal{A}) = \begin{cases} \mathcal{K}^i_{\mathcal{G}} + \psi\Omega(\omega) & if\ \mathrm{I}_k = 1 \\ \mathcal{A} + \psi\Omega(\omega) & if\ \mathrm{I}_k = 0. \end{cases} \tag{17}$$

The loss in training the network is computed by utilizing the mean squared loss function, which measures the discrepancy between the model's clustering output and the actual labels. The network is trained by minimizing the loss, represented as ℓ:

$$\ell = \|\mathcal{L}'(\omega;\mathcal{K}^i_{\mathcal{G}} - \mathcal{A}) - \mathcal{A}\|_F. \tag{18}$$

During the model training process, the training loss ℓ is propagated through the network using the backpropagation algorithm. Gradients are computed to update the network's parameters, ω, aiming to minimize the loss ℓ within the network. After each training epoch, the network is tested using data from unforeseen classes to evaluate its generalization ability. Visual results from t-SNE [71] visualization experiments demonstrate a significant enhancement in the distinctiveness and reliability of class separations, reducing misclassification, and strengthening the model's generalization capabilities.

The final consensus clustering matrix $\mathrm{H}^{\mathcal{C}}$ obtained from model training can effectively classify new samples.

Algorithm 2 Category Knowledge-guided Incomplete Multikernel Algorithm

---

Require: $\{\mathrm{H}_\varrho\}_{\varrho=1}^m$, k, $\mathcal{G}_\theta(\mathcal{S})$, $\mathcal{G}_\theta(\mathcal{Q})$, $\ddot{\mathcal{P}}_\varrho = \mathrm{I}_k$
Ensure: Consensus clustering matrix $\mathrm{H}^\mathcal{C}$
  Initialize $\{\mathrm{H}_\varrho\}_{\varrho=1}^m = \mathrm{I}_{m\times k}$, $\{\ddot{\mathcal{P}}_\varrho\}_{\varrho=1}^m = \mathrm{I}_k$, $m = k$, $\xi = 1/\sqrt{m}$
  repeat
    Update $\mathcal{G}$ using the joint Eq.(8) and Eq.(9);
    Update $\mathrm{H}^\mathcal{C}$ using Eq.(9);
    Update $\ddot{\mathcal{P}}_\varrho$ using Eq.(10);
    Update $\mathcal{K}_\mathcal{G}^i$ using Eq.(11);
    Update $\mathcal{B}$ using Eq.(13);
    Update $\hat{\Theta}$ using Eq.(14);
    Update ξ using Eq.(14);
    Update ω using Eq.(15);
  until Eq.(18) is reached for convergence

3.3.3. Extension of the Algorithm

This study introduces a CKG-IMK algorithm that estimates incomplete base clustering matrices $\{\mathrm{H}_\varrho\}_{\varrho=1}^m$ from multiple few-shot tasks by enhancing task-relevant features during the clustering process [72]. In the matching phase, the algorithm accurately distinguishes samples. Specifically, the algorithm defines the obtained $\mathrm{H}^\mathcal{C} \in \mathbb{R}^{n\times k}$ as the incomplete multikernel clustering matrix, where m represents the number of clusters for each clustering task, n is the number of samples in each input, and k is the embedded dimension. To strike a balance between computational complexity and information retention in experiments, m is set equal to k.

In the CKG-IMK algorithm, the process begins with the computation of an incomplete similarity matrix $\mathcal{K}_\mathcal{G}$ to perform clustering and generate a set of incomplete base clustering matrices $\{\mathrm{H}_\varrho\}_{\varrho=1}^m$ for each base kernel. These obtained base clustering matrices are then used to learn the consensus clustering matrix $\mathrm{H}^\mathcal{C}$, which is subsequently employed to estimate each incomplete base clustering matrix. These two steps are iteratively performed until convergence is achieved. The underlying concept is to maximize alignment between the consensus clustering matrix and adaptively weighted base clustering matrices with the optimal arrangement. This prior knowledge is capable of integrating features from different categories [73], facilitating the learning of the consensus clustering matrix and thereby enhancing the performance and efficiency of clustering.

4. Experimental results and analysis

4.1. Experimental setup

In the experiment, the model's training iterations were set to 100, with a batch size of 128, and the Adam optimizer with a learning rate of 1e-4 was employed. Additionally, the learning rate underwent logarithmic decay based on the number of iterations. During the meta-training phase, the training dataset was utilized for supervised learning. In the meta-testing phase, the model with the highest accuracy was loaded and used for clustering the data in the query set. This study adhered to the standard principles of small-sample classification, incorporating three query modes: 1-shot, 3-shot, and 5-shot. To ensure fairness and consistency in our experiments, all competing models were tested under the same running environment, which included an Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz, NVIDIA GeForce GTX3090 GPU, CUDA 11.0, and PyTorch 1.12.

Regarding the evaluation of the results, various common validation methods were employed to comprehensively assess the fault diagnosis outcomes. Given the presence of different types of fault data in practical applications, including permanent faults and early-stage minor faults, the CWRU dataset [69] and the EMF-TM bearing dataset were chosen as the basis for two case studies. In the first case, the CWRU dataset containing vibration signals from 10 different bearing conditions was utilized. In the second case, fault diagnosis across different interdisciplinary scenarios was explored, encompassing rolling bearings under various loads and different fault severities. Multiple experiments were conducted to thoroughly validate the robust generalization capabilities and outstanding diagnostic performance of our proposed CKG method, highlighting its applicability across different fault contexts.

4.1.1. Case1: CWRU Dataset

1)Description of the CWRU Dataset: The CWRU dataset, which is provided by CWRU Bearing, has become a widely used benchmark for diagnosing faults in bearings. The data are collected using accelerometers from fan- and drive-end deep groove ball bearings, which are sampled at the frequency of 48 kHz. The dataset consists of vibration signals from three predesigned bearing faults obtained through electrical discharge machining (EDM): inner race fault (IF), outer race fault (OF), and ball fault (BF). Each bearing fault includes three fault sizes: 0.007 inches, 0.014 inches, and 0.021 inches. The signals were collected at a sampling frequency of 48 kHz under different loads (1HP, 2HP, and 3HP). Thus, for each load condition, there are 10 bearing states (one normal state and nine fault states).

Table 2
Details of the CWRU dataset.

Label	Fault Type	Defect Size (inch)	Accelerometer	Load (hp)
(a)	BF	0.007	Drive end	1
(b)	BF	0.014	Drive end	1
(c)	BF	0.021	Drive end	1
(d)	IF	0.007	Drive end	1
(e)	IF	0.014	Drive end	1
(f)	IF	0.021	Drive end	1
(g)	OF	0.007	Drive end	1
(h)	OF	0.014	Drive end	1
(i)	OF	0.021	Drive end	1

In practical industrial settings, the partitioning of data is crucial, especially when dealing with different fault sizes and types. To better simulate the variations in rolling bearing conditions encountered in real-world usage, we selected three different fault types and healthy conditions from among the 10 available categories as the meta-testing set, while the remaining six fault types were used as the meta-training set. This data partitioning approach aimed to reflect the model's robustness and generalization ability in classifying unknown non-stationary faults during actual healthy operational conditions. Given the significant disparities in data distribution, the model needs to adapt to various fault types and sizes, thereby better accommodating the diversity and complexity of real-world conditions.

2)Performance Comparison with Existing Methods: Nine recently published fault diagnosis methods were employed in this experiment to validate the performance of the proposed CKG method on the CWRU dataset. These methods include K-nearest neighbor algorithms (KNN) [74], DPDAN [37], CNN-MMD [75], MANN [76], UCPMnet [43], MAML [77], ProtoNet [78], MD-SN [79], and RelationNet [80]. To comprehensively demonstrate the effectiveness of the CKG method, clustering methods like KNN and two domain adaptation methods were also included, as transfer learning methods like DPDAN and CNN-MMD are widely applied in real-world scenarios. In the experiment, three different fault types were merged with the normal state to construct three meta-test sets, each consisting of four categories. Multiple repeated experiments were conducted, including 1-shot, 3-shot, and

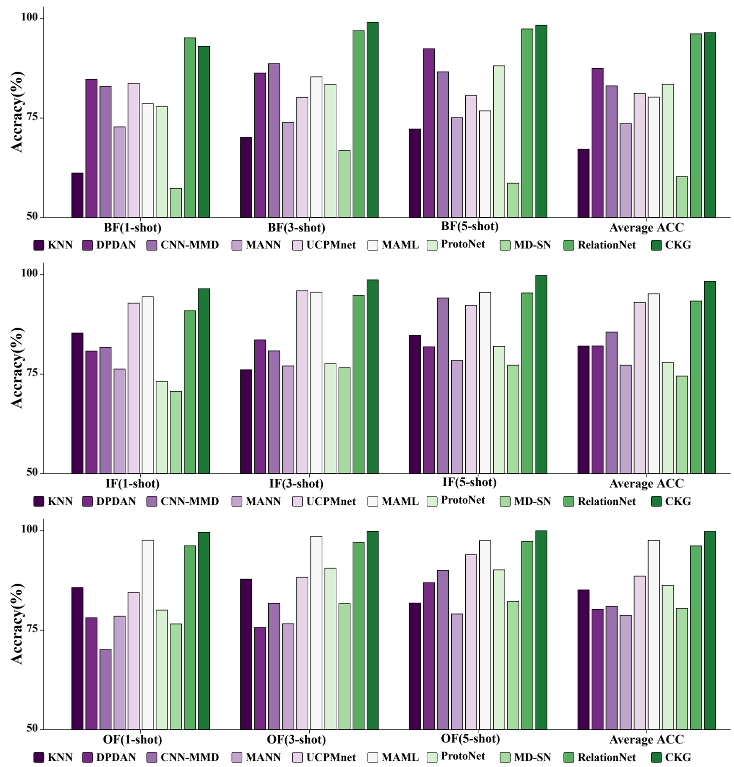


Fig. 2. Performance of various methods on three types of bearing operation data, with healthy data included as the test set: (a) ball Fault (BF);(b) inner race Fault (IF);(c) outer race Fault (OF).

5-shot tasks. To ensure fairness, all methods were trained using the same backbone architecture and hyperparameters, and each method underwent 10 experiments to determine the average accuracy, minimizing the impact of uncertainties arising from random network initialization and neural network training. The experimental results are presented in Fig.2 and Table 3. For ease of reading, the optimal and suboptimal accuracy results are indicated with bold and underlined formatting in Table 3.

Based on the experimental data presented in Table 3, it can be observed that the accuracy of KNN fluctuates significantly under different operating conditions, indicating its subpar clustering performance in the presence of shifting data distributions. Meanwhile, domain adaptation methods, such as DPDAN and CNN-MMD, show a declining trend in performance when confronted with knowledge transfer across different labels. This decline can be attributed to their high reliance on data distribution similarity between the source and target domains as well as the similarity of learning tasks between the two domains. Additionally, the representative FSL method, MD-SN, has some limitations. Its complex structure is prone to overfitting, and it struggles to capture inter-task correlations, leading to poorer performance on the meta-test set. Compared to CKG, MD-SN exhibits a significant performance gap of up to 35.68 percentage points. It is worth noting that classical methods like MANN require an ample amount of labeled data for training the diagnostic model. Limited meta-training samples may cause model instability and hinder its generalizability to target fault diagnosis tasks.

Table 3
Performance of CKG and nine existing methods.

Method	BF (1-shot)	BF (3-shot)	BF (5-shot)	Average ACC
KNN	61.18%	70.13%	72.24%	67.85%
DPDAN	84.76%	86.31%	92.42%	87.50%
CNN-MMD	82.97%	88.69%	86.61%	83.09%
MANN	72.78%	73.91%	75.11%	73.60%
UCPMnet	83.74%	80.20%	80.66%	81.20%
MAML	78.62%	85.36%	76.81%	80.26%
ProtoNet	77.88%	83.49%	88.13%	83.50%
MD-SN	57.34%	66.88%	58.63%	60.28%
RelationNet	**95.16**%	96.95%	97.40%	96.17%
CKG	93.02%	**99.09**%	**98.34**%	**96.48**%
Method	IF (1-shot)	IF (3-shot)	IF (5-shot)	Average ACC
KNN	85.33%	76.10%	84.76%	82.06%
DPDAN	80.79%	83.60%	81.86%	82.08%
CNN-MMD	81.74%	80.84%	94.13%	85.57%
MANN	76.28%	77.05%	78.43%	77.25%
UCPMnet	92.84%	95.95%	92.30%	93.03%
MAML	94.45%	95.59%	95.54%	95.19%
ProtoNet	73.15%	77.62%	81.97%	77.91%
MD-SN	70.67%	76.61%	77.25%	74.51%
RelationNet	90.93%	94.78%	95.40%	93.37%
CKG	**96.46**%	**98.69**%	**99.79**%	**98.31**%
Method	OF (1-shot)	OF (3-shot)	OF (5-shot)	Average ACC
KNN	85.72%	87.83%	81.81%	85.12%
DPDAN	78.13%	75.68%	86.95%	80.25%
CNN-MMD	70.13%	81.76%	90.05%	80.98%
MANN	78.52%	76.61%	79.08%	78.74%
UCPMnet	84.47%	88.31%	93.99%	88.59%
MAML	97.61%	98.59%	97.51%	97.57%
ProtoNet	80.05%	90.58%	90.16%	86.26%
MD-SN	76.57%	81.70%	82.23%	80.50%
RelationNet	96.21%	97.05%	97.32%	96.19%
CKG	**99.59**%	**99.85**%	**99.99**%	**99.81**%

For the three types of bearing operation, the CKG method has significantly improved ball fault diagnostic accuracy compared to other methods.

Specifically, the CKG method achieved an improvement of 8.26% - 31.84% in 1-shot ball fault diagnosis, 2.14% - 32.21% in 3-shot ball fault diagnosis, and 1.94% - 39.71% in 5-shot ball fault diagnosis. Furthermore, the CKG method also demonstrated improved accuracy in diagnosing inner race faults, with an improvement of 2.74% - 22.59% in 1-shot diagnosis, 2.74% - 22.54% in 3-shot diagnosis, and 4.25% - 22.54% in 5-shot diagnosis. Similarly, for outer race faults, the CKG method achieved an improvement of 3.38% - 29.46% in 1-shot diagnosis, 1.26% - 24.17% in 3-shot diagnosis, and 2.67% - 20.91% in 5-shot diagnosis.

Overall, the proposed CKG method demonstrates greater robustness and superior performance in these nine experiments when compared to the recently published nine other fault diagnosis methods. These experimental results comprehensively showcase the robustness and superiority of the CKG method across diverse working conditions.

4.1.2. Case2: EMF-TM Dataset

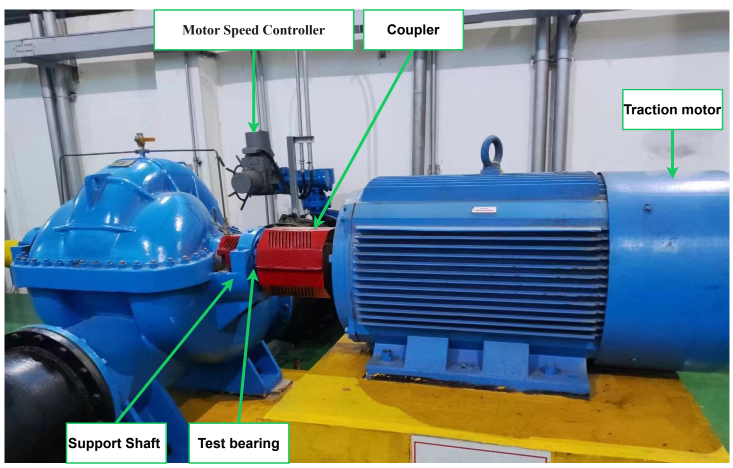


Fig. 3. Case 2 - EMF-TM Bearing Test Bench.

1)Description of the EMF-TM Dataset: The dataset was acquired from motors with broken rotor bars operating at different load currents (0.2A, 0.5A, and 0.8A, with the device capable of a maximum load of 1A) and varying levels of fault severity. The load currents of the motors fall into three categories: 0.2A (light load [LL]), 0.5A (moderate load [ML]), and 0.8A (heavy load [HL]). Additionally, three levels of fault severities were considered, namely, 0.005 (trivial fault), 0.200 (moderate fault), and 1.000 (severe fault), resulting in a total of 10 bearing conditions, which also include the healthy state (H). The data were sampled at a frequency of 2.4 KHz. The experimental arrangement for data collection is depicted in Fig.3, while a comprehensive dataset description can be found in Tabel 4.

Table 4
Details of the EMF-TM dataset.

Label	Load Current	Degree of Fault	Accelerometer	Temperature(°C)
(a)	-	-	Drive end	45
(b)	LL	0.005	Drive end	45
(c)	LL	0.200	Drive end	45
(d)	LL	1.000	Drive end	45
(e)	ML	0.005	Drive end	45
(f)	ML	0.200	Drive end	45
(g)	ML	1.000	Drive end	45
(h)	HL	0.005	Drive end	45
(i)	HL	0.200	Drive end	45
(j)	HL	1.000	Drive end	45

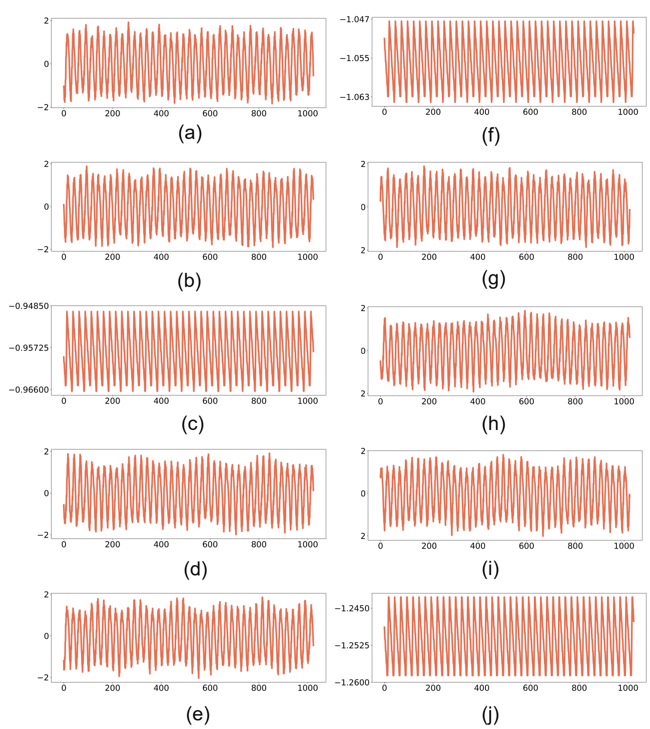


Fig. 4. Raw vibration signals for 10 bearing states of the EMF-TM dataset.

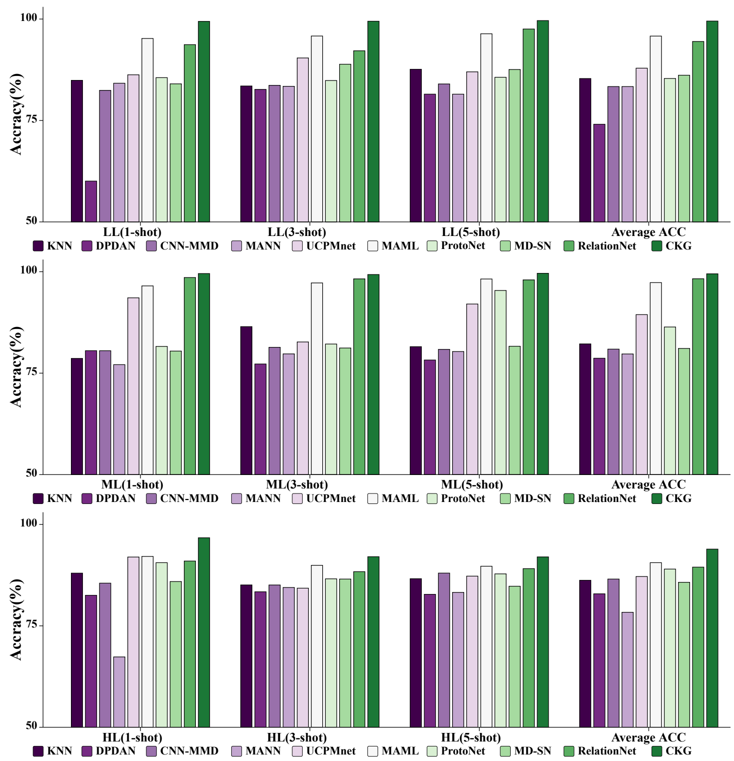


Fig. 5. Performance of various methods on three types of bearing operation data, with healthy data included as the test set:(a) Light load data;(b) Moderate load data;(c) Heavy load data.

2) Performance Comparison with Existing Methods: In this experiment, nine recently published fault diagnosis methods were employed to evaluate the performance of the proposed CKG method on the EMF-TM dataset. These methods include KNN, DPDAN, CNN-MMD, MANN, uncertainty-based contrastive prototype-matching network (UCPMnet), MAML, ProtoNet, MD-SN, and RelationNet. In the experiment, three different load types were merged with the normal state to construct three meta-test sets, each consisting of four categories, to assess the model's ability to sensitively classify minor faults. Multiple repeated experiments were conducted, including 1-shot, 3-shot, and 5-shot tasks. To prevent overfitting, an early stopping strategy was employed during training. In the meta-training phase, the model with the highest accuracy was saved to perform meta-test tasks.To ensure fairness, all methods were trained using the same backbone architecture and hyperparameters, and each method underwent 10 repeated experiments to determine the average accuracy, minimizing the impact of uncertainties arising from random network initialization and neural network training. The experimental results are presented in Fig.5 and Table 5. For ease of reading, the optimal and suboptimal accuracy results are indicated with bold and underlined formatting in Table 5.

Table 5
Performance of CKG and nine existing methods.

Method	LL (1-shot)	LL (3-shot)	LL (5-shot)	Average ACC
KNN	84.89%	83.50%	87.62%	85.34%
DPDAN	60.06%	82.66%	81.49%	74.07%
CNN-MMD	82.42%	83.66%	84.00%	83.36%
MANN	84.18%	83.42%	81.48%	83.36%
UCPMnet	86.26%	90.41%	86.99%	87.89%
MAML	<u>95.21%</u>	<u>95.82%</u>	<u>96.37%</u>	<u>95.80%</u>
ProtoNet	85.56%	84.85%	85.65%	85.35%
MD-SN	84.03%	88.86%	87.56%	86.15%
RelationNet	93.69%	92.18%	97.54%	94.47%
CKG	**99.42%**	**99.46%**	**99.62%**	**99.50%**
Method	ML (1-shot)	ML (3-shot)	ML (5-shot)	Average ACC
KNN	78.63%	86.47%	81.53%	82.21%
DPDAN	80.54%	77.26%	78.24%	78.68%
CNN-MMD	80.53%	81.37%	80.86%	80.92%
MANN	77.10%	79.75%	80.33%	79.73%
UCPMnet	93.57%	82.69%	92.04%	89.43%
MAML	96.51%	97.24%	<u>98.21%</u>	97.32%
ProtoNet	81.59%	82.18%	95.38%	86.38%
MD-SN	80.44%	81.20%	81.63%	81.09%
RelationNet	<u>98.57%</u>	<u>98.24%</u>	97.99%	<u>98.27%</u>
CKG	**99.55%**	**99.31%**	**99.61%**	**99.49%**
Method	HL (1-shot)	HL (3-shot)	HL (5-shot)	Average ACC
KNN	87.99%	85.08%	86.62%	86.23%
DPDAN	82.53%	83.40%	82.77%	82.90%
CNN-MMD	85.52%	85.07%	88.00%	86.53%
MANN	67.34%	84.45%	83.24%	78.34%
UCPMnet	91.96%	84.28%	87.25%	87.16%
MAML	<u>92.11%</u>	<u>89.92%</u>	<u>89.69%</u>	<u>90.57%</u>
ProtoNet	90.57%	86.59%	87.81%	88.99%
MD-SN	85.92%	86.53%	84.76%	85.73%
RelationNet	90.97%	88.35%	89.10%	89.47%
CKG	**96.71%**	**92.04%**	**91.99%**	**93.91%**

Based on the experimentation, it was found that CNN-MMD has a faster training speed but performs poorly in terms of test accuracy, often struggling to converge. The DPDAN method performs inadequately under light loads, indicating that it learns specific features during training, lacking transferability. The MANN and MAML methods require meticulous feature engineering and higher computational resources, with fluctuating performance across different datasets. The ProtoNet method has limitations in capturing effective and information-rich fault features. Finally, the RelationNet method appears to be sensitive to the size and quality of the dataset, resulting in unstable accuracy levels.

The CKG method has shown significant improvement in light load diagnostic accuracy compared to other methods for the three types of load currents. Specifically, the CKG method achieved an improvement of 4.21% - 39.36% in 1-shot ball fault diagnosis, 3.64% - 16.80% in 3-shot ball fault diagnosis, and 3.84% - 25.55% in 5-shot ball fault diagnosis. Furthermore, the CKG method also demonstrated improved accuracy in diagnosing moderate load faults, with an improvement of 0.98% - 22.45% in 1-shot diagnosis, 1.07% - 22.05% in 3-shot diagnosis, and 1.40% - 21.40% in 5-shot diagnosis. Similarly, for heavy load faults, the CKG method achieved an improvement of 4.60% - 29.37% in 1-shot diagnosis, 2.12% - 8.64% in 3-shot diagnosis, and 2.30% - 9.22% in 5-shot diagnosis.

The proposed CKG method efficiently clusters based on inter-task correlated features, achieving the highest diagnostic accuracy while demonstrating stability and superiority in multiple aspects. It not only exhibits high-speed operation and low computational memory usage but also attains high accuracy during the training process. Importantly, CKG can effectively distinguish minor early-stage faults from healthy states under the interference of strong background noise, even when there are slight variations in load conditions. Thus, CKG represents a valuable method for addressing complex, dynamic, and high-demand real-world applications, especially in the realm of early and rapid fault diagnosis.

3) Advanced Feature Visualization: The feature extraction capabilities of the aforementioned methods were visualized using the t-SNE method, and the results are depicted in Fig.6. In comparison to the nine recently published methods, the proposed CKG method exhibits well-separated and compact data clusters. Nearly all samples are clustered within their respective regions, and there are reasonable inter-class differences among all four states. It is noteworthy that under the interference of strong background noise, the subtle distinctions between the healthy state and early fault features pose a significant challenge for early fault diagnosis in rolling bearings. Other models tend to confuse early minor faults with healthy operational conditions, which can be fatal in the context of real industrial vulnerabilities.

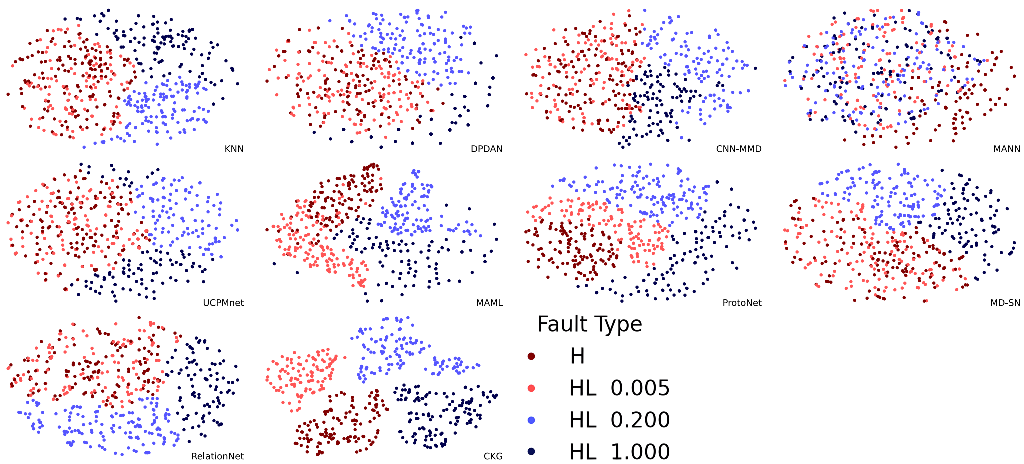


Fig. 6. t-SNE-based visualization of extracted features by different methods. The labels H, HL 0.005, HL 0.200, and HL 1.000 correspond to healthy state, trivial fault, moderate fault, and severe fault, respectively.

The traditionally used KNN clustering method yielded dispersed learned features and had a poor ability to learn inter-task correlated features. In addition, it exhibited low convergence accuracy. Transfer learning methods like DPDAN and CNN-MMD incurred high tuning costs in the target domain. ProtoNet, while capable of distinguishing between the four categories, lacked clarity and was prone to misclassification. Visual results indicate that the CKG method effectively extracts discriminative features from vibration signals in fault diagnosis. It maximizes classification boundaries and achieves optimal clustering results, surpassing the performance of the nine recently published methods.

5. Conclusion

In order to address the more challenging cross-domain cold-start tasks and early fault diagnosis tasks within FSL and to reduce the complexity of FSL, this study introduced an innovative CKG method. This method effectively utilizes an incomplete multikernel clustering algorithm to capture category information among data. Unlike traditional inductive approaches, this method is capable of classifying unlabeled query instances in a single pass, thus overcoming the challenges of small-sample category knowledge imbalances. Further, the proposed CKG method was compared with nine other cross-domain fault diagnosis methods. Multiple experiments were conducted on the CWRU bearing dataset and the traction motor dataset, assessing the performance of these methods across classic 4-way, 1-shot, 3-shot, and 5-shot tasks. The experimental results demonstrate that, even with limited samples in both support and query sets, the proposed category-guided incomplete clustering method outperforms the other nine fault diagnosis methods.

Furthermore, the visual results also demonstrate that the proposed CKG jointly optimizes the best kernel, maximum-margin hyperplane, and optimal clustering labels while effectively distinguishing minor faults from normal states. This highlights the high fault diagnosis accuracy of the method. Notably, the CKG method was tested and applied in an actual industrial production context, offering a more reliable and efficient solution for bearing fault diagnosis. In summary, this study provides a viable solution for the field of small-sample bearing fault diagnosis and offers valuable insights for future research endeavors.

Acknowledgments

The authors would like to express their gratitude to Guangzhou Sanki Automotive Gasket Co., Ltd. for providing test data samples and verifying the algorithm studied in this article during actual production work. This work was supported by Guangzhou Youth Science and Technology Education Project under Grant KP2023243 & KP2023245.

References

[1] J. Mi, Y.-F. Li, W. Peng, H.-Z. Huang, Reliability analysis of complex multi-state system with common cause failure based on evidential networks, Reliability Engineering & System Safety 174 (2018) 71–81.

[2] J. Zhu, N. Chen, W. Peng, Estimation of bearing remaining useful life based on multiscale convolutional neural network, IEEE Transactions on Industrial Electronics 66 (2018) 3208–3216.

[3] C.-G. Huang, H.-Z. Huang, Y.-F. Li, W. Peng, A novel deep convolutional neural network-bootstrap integrated method for rul prediction of rolling bearing, Journal of Manufacturing Systems 61 (2021) 757–772.

[4] Z. Yang, B. Xu, W. Luo, F. Chen, Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review, Measurement 189 (2022) 110460.

[5] Y. Xue, R. Yang, X. Chen, Z. Tian, Z. Wang, A novel local binary temporal convolutional neural network for bearing fault diagnosis, IEEE Transactions on Instrumentation and Measurement (2023).

[6] K. Liu, N. Lu, F. Wu, R. Zhang, F. Gao, Model fusion and multiscale feature learning for fault diagnosis of industrial processes, IEEE Transactions on Cybernetics (2022).

[7] Q. Sun, F. Peng, X. Yu, H. Li, Data augmentation strategy for power inverter fault diagnosis based on wasserstein distance and auxiliary classification generative adversarial network, Reliability Engineering & System Safety 237 (2023) 109360.

[8] M. Shi, C. Ding, R. Wang, C. Shen, W. Huang, Z. Zhu, Graph embedding deep broad learning system for data imbalance fault diagnosis of rotating machinery, Reliability Engineering & System Safety (2023) 109601.

[9] K. Zhong, J. Wang, S. Xu, C. Cheng, H. Chen, Overview of fault prognosis for traction systems in high-speed trains: A deep learning perspective, Engineering Applications of Artificial Intelligence 126 (2023) 106845.

[10] S.-P. Zhu, H.-Z. Huang, W. Peng, H.-K. Wang, S. Mahadevan, Probabilistic physics of failure-based framework for fatigue life prediction of aircraft gas turbine discs under uncertainty, Reliability Engineering & System Safety 146 (2016) 1–12.

[11] X.-Y. Li, X. Li, J. Feng, C. Li, X. Xiong, H.-Z. Huang, Reliability analysis and optimization of multi-phased spaceflight with backup missions and mixed redundancy strategy, Reliability Engineering & System Safety 237 (2023) 109373.

[12] Y. Li, Z. Liu, Z. He, L. Tu, H.-Z. Huang, Fatigue reliability analysis and assessment of offshore wind turbine blade adhesive bonding under the coupling effects of multiple environmental stresses, Reliability Engineering & System Safety (2023) 109426.

[13] K. Zhao, J. Hu, H. Shao, J. Hu, Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy, Reliability Engineering & System Safety 236 (2023) 109246.

[14] T. Zhang, J. Chen, J. Xie, T. Pan, Sasln: Signals augmented self-taught learning networks for mechanical fault diagnosis under small sample condition, IEEE Transactions on Instrumentation and Measurement 70 (2021).

[15] P. Lei, C. Shen, D. Wang, L. Chen, Z. Zhou, Z. Zhu, A new transferable bearing fault diagnosis method with adaptive manifold probability distribution under different working conditions, Measurement 173 (2021) 108565.

[16] X. Cong, Y. Song, Y. Li, L. Jia, Federated domain generalization with global robust model aggregation strategy for bearing fault diagnosis, Measurement Science and Technology 34 (2023) 115116.

[17] T. Zhang, J. Chen, S. Liu, Z. Liu, Domain discrepancy-guided contrastive feature learning for few-shot industrial fault diagnosis under variable working conditions, IEEE Transactions on Industrial Informatics (2023).

[18] Y.-F. Li, H.-Z. Huang, J. Mi, W. Peng, X. Han, Reliability analysis of multi-state systems with common cause failures based on bayesian network and fuzzy probability, Annals of Operations Research (2022) 1–15.

[19] Y. Ma, J. Yang, L. Li, Meta bi-classifier gradient discrepancy for noisy and universal domain adaptation in intelligent fault diagnosis, Knowledge-Based Systems (2023) 110735.

[20] D. Huang, W.-A. Zhang, F. Guo, W. Liu, X. Shi, Wavelet packet decomposition-based multiscale cnn for fault diagnosis of wind turbine gearbox, IEEE Transactions on Cybernetics (2021).

[21] J. Chen, C. Lin, B. Yao, L. Yang, H. Ge, Intelligent fault diagnosis of rolling bearings with low-quality data: A feature significance and diversity learning method, Reliability Engineering & System Safety 237 (2023) 109343.

[22] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, Z. Wang, Deep transfer learning for bearing fault diagnosis: A systematic review since 2016, IEEE Transactions on Instrumentation and Measurement (2023).

[23] J. Chen, W. Hu, D. Cao, Z. Zhang, Z. Chen, F. Blaabjerg, A meta-learning method for electric machine bearing fault diagnosis under varying working conditions with limited data, IEEE Transactions on Industrial Informatics 19 (2022) 2552–64.

[24] A. Parnami, M. Lee, Learning from few examples: A summary of approaches to few-shot learning, arXiv preprint arXiv:2203.04291 (2022).

[25] S. X. Hu, D. Li, J. Stühmer, M. Kim, T. M. Hospedales, Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9068–77.

[26] Y. Feng, J. Chen, J. Xie, T. Zhang, H. Lv, T. Pan, Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects, Knowledge-Based Systems 235 (2022) 107646.

[27] D. Zhang, K. Zheng, Y. Bai, D. Yao, D. Yang, S. Wang, Few-shot bearing fault diagnosis based on meta-learning with discriminant space optimization, Measurement Science and Technology 33 (2022) 115024.

[28] L. Qiao, Y. Zhang, Q. Wang, Fault detection in wind turbine generators using a meta-learning-based convolutional neural network, Mechanical Systems and Signal Processing 200 (2023) 110528.

[29] J. Liu, H. Cao, Y. Luo, An information-induced fault diagnosis framework generalizing from stationary to unknown nonstationary working conditions, Reliability Engineering & System Safety 237 (2023) 109380.

[30] Y. Shi, A. Deng, M. Deng, M. Xu, Y. Liu, X. Ding, W. Bian, Domain augmentation generalization network for real-time fault diagnosis under unseen working conditions, Reliability Engineering & System Safety 235 (2023) 109188.

[31] Y. Feng, J. Chen, Z. Yang, X. Song, Y. Chang, S. He, E. Xu, Z. Zhou, Similarity-based meta-learning network with adversarial domain adaptation for cross-domain fault identification, Knowledge-Based Systems 217 (2021) 106829.

[32] P. Ding, X. Zhao, H. Shao, M. Jia, Machinery cross domain degradation prognostics considering compound domain shifts, Reliability Engineering & System Safety 239 (2023) 109490.

[33] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: International conference on machine learning, PMLR, 2017, pp. 1126–35.

[34] J. Zhang, Y. Zhou, B. Wang, Z. Wu, Bearing fault diagnosis base on multi-scale 2d-cnn model, in: 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), IEEE, 2021, pp. 72–5.

[35] P. Li, H. Xiao, W. Jiang, D. Ning, Compound fault diagnosis of industrial robot based on improved multi-label one-dimensional convolutional neural network, in: Data Mining and Big Data: 6th International Conference, DMBD 2021, Guangzhou, China, October 20–22, 2021, Proceedings, Part II 6, Springer, 2021, pp. 205–16.

[36] C. Wang, Z. Xu, An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis, Neurocomputing 456 (2021) 550–62.

[37] H. Wang, X. Bai, J. Tan, J. Yang, Deep prototypical networks based domain adaptation for fault diagnosis, Journal of Intelligent Manufacturing (2022) 1–11.

[38] Y. Feng, J. Chen, T. Zhang, S. He, E. Xu, Z. Zhou, Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis, ISA transactions 120 (2022) 383–401.

[39] K. Zhang, J. Chen, T. Zhang, S. He, T. Pan, Z. Zhou, Intelligent fault diagnosis of mechanical equipment under varying working condition via iterative matching network augmented with selective signal reuse strategy, Journal of Manufacturing Systems 57 (2020) 400–15.

[40] S. Wang, D. Wang, D. Kong, J. Wang, W. Li, S. Zhou, Few-shot rolling bearing fault diagnosis with metric-based meta learning, Sensors 20 (2020) 6437.

[41] Y. An, K. Zhang, Y. Chai, Q. Liu, X. Huang, Bearing fault diagnosis under variable working conditions base on contrastive domain adaptation method, IEEE Transactions on Instrumentation and Measurement 71 (2022) 1–11.

[42] Q. Zhang, L. Tang, J. Xuan, T. Shi, R. Li, An uncertainty relevance metric-based domain adaptation fault diagnosis method to overcome class relevance caused confusion, Reliability Engineering & System Safety 231 (2023) 109040.

[43] T. Zhang, J. Jiao, J. Lin, H. Li, J. Hua, D. He, Uncertainty-based contrastive prototype-matching network towards cross-domain fault diagnosis with small data, Knowledge-Based Systems 254 (2022) 109651.

[44] Z. Ren, T. Lin, K. Feng, Y. Zhu, Z. Liu, K. Yan, A systematic review on imbalanced learning methods in intelligent fault diagnosis, IEEE Transactions on Instrumentation and Measurement (2023).

[45] Y. Chen, D. Zhang, R. Yan, Domain adaptation networks with parameter-free adaptively rectified linear units for fault diagnosis under variable operating conditions, IEEE Transactions on Neural Networks and Learning Systems (2023).

[46] Y. An, K. Zhang, Y. Chai, Z. Zhu, Q. Liu, Gaussian mixture variational based transformer domain adaptation fault diagnosis method and its application in bearing fault diagnosis, IEEE Transactions on Industrial Informatics (2023).

[47] G. Liu, W. Shen, L. Gao, A. Kusiak, Automated broad transfer learning for cross-domain fault diagnosis, Journal of manufacturing systems 66 (2023) 27–41.

[48] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily, A. Abdellatif, A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations, Ain Shams Engineering Journal 14 (2023) 101945.

[49] R. Ma, T. Han, W. Lei, Cross-domain meta learning fault diagnosis based on multi-scale dilated convolution and adaptive relation module, Knowledge-Based Systems 261 (2023) 110175.

[50] C. Zhang, Y. Liu, A two-step denoising strategy for early-stage fault diagnosis of rolling bearings, IEEE Transactions on Instrumentation and Measurement 69 (2020) 6250–61.

[51] Z. Gao, Y. Liu, Q. Wang, J. Wang, Y. Luo, Ensemble empirical mode decomposition energy moment entropy and enhanced long short-term memory for early fault prediction of bearing, Measurement 188 (2022) 110417.

[52] C. Yin, Y. Wang, G. Ma, Y. Wang, Y. Sun, Y. He, Weak fault feature extraction of rolling bearings based on improved ensemble noise-reconstructed emd and adaptive threshold denoising, Mechanical Systems and Signal Processing 171 (2022) 108834.

[53] Y. Ding, J. Zhuang, P. Ding, M. Jia, Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings, Reliability Engineering & System Safety 218 (2022) 108126.

[54] Z. Wu, L. Su, Q. Huang, Decomposition and completion network for salient object detection, IEEE transactions on image processing 30 (2021) 6226–39.

[55] L. Zhao, G. Liu, D. Guo, W. Li, X. Fang, Boosting few-shot visual recognition via saliency-guided complementary attention, Neurocomputing 507 (2022) 412–27.

[56] Z. Li, C. Lang, J. H. Liew, Y. Li, Q. Hou, J. Feng, Cross-layer feature pyramid network for salient object detection, IEEE Transactions on Image Processing 30 (2021) 4587–98.

[57] C. Chen, X. Yang, J. Zhang, B. Dong, C. Xu, Category knowledge-guided parameter calibration for few-shot object detection, IEEE Transactions on Image Processing 32 (2023) 1092–1107.

[58] J. Wen, Z. Zhang, L. Fei, B. Zhang, Y. Xu, Z. Zhang, J. Li, A survey on incomplete multiview clustering, IEEE Transactions on Systems, Man, and Cybernetics: Systems 53 (2022) 1136–49.

[59] S. Zhang, F. Ye, B. Wang, T. G. Habetler, Few-shot bearing fault diagnosis based on model-agnostic meta-learning, IEEE Transactions on Industry Applications 57 (2021) 4754–64.

[60] X. Liu, Simplemkkm: Simple multiple kernel k-means, IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (2022) 5174–86.

[61] X. Li, Z. Sun, J.-H. Xue, Z. Ma, A concise review of recent few-shot meta-learning methods, Neurocomputing 456 (2021) 463–8.

[62] Z. Wang, J. Xuan, T. Shi, Alternative multi-label imitation learning framework monitoring tool wear and bearing fault under different working conditions, Advanced Engineering Informatics 54 (2022) 101749.

[63] C. Chen, C. Liu, T. Wang, A. Zhang, W. Wu, L. Cheng, Compound fault diagnosis for industrial robots based on dual-transformer networks, Journal of Manufacturing Systems 66 (2023) 163–78.

[64] Y. Guan, Z. Meng, D. Sun, J. Liu, F. Fan, Rolling bearing fault diagnosis based on information fusion and parallel lightweight convolutional network, Journal of Manufacturing Systems 65 (2022) 811–21.

[65] Y. Xu, X. Yan, B. Sun, Z. Liu, Global contextual residual convolutional neural networks for motor fault diagnosis under variable-speed conditions, Reliability Engineering & System Safety 225 (2022) 108618.

[66] B. Yang, X. Zhang, Z. Lin, F. Nie, B. Chen, F. Wang, Efficient and robust multiview clustering with anchor graph regularization, IEEE Transactions on Circuits and Systems for Video Technology 32 (2022) 6200–13.

[67] Z. Li, C. Tang, X. Zheng, X. Liu, W. Zhang, E. Zhu, High-order correlation preserved incomplete multi-view subspace clustering, IEEE Transactions on Image Processing 31 (2022) 2067–80.

[68] X. Zhang, S. Zhao, J. Wang, L. Guo, X. Wang, H. Sun, Purity-preserving kernel tensor low-rank learning for robust subspace clustering, IEEE Transactions on Circuits and Systems for Video Technology (2023).

[69] K. Loparo, Case western reserve university bearing data center, Bearings Vibration Data Sets, Case Western Reserve University (2012) 22–8.

[70] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, Advances in neural information processing systems 30 (2017).

[71] D. Kobak, P. Berens, The art of using t-sne for single-cell transcriptomics, Nature communications 10 (2019) 5416.

[72] M. Li, Y. Zhang, C. Ma, S. Liu, Z. Liu, J. Yin, X. Liu, Q. Liao, Regularized simple multiple kernel k-means with kernel average alignment, IEEE Transactions on Neural Networks and Learning Systems (2023).

[73] X. Liu, Hyperparameter-free localized simple multiple kernel k-means with global optimum, IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).

[74] B. Song, S. Tan, H. Shi, B. Zhao, Fault detection and diagnosis via standardized k nearest neighbor for multimode process, Journal of the Taiwan Institute of Chemical Engineers 106 (2020) 1–8.

[75] J. Jiao, M. Zhao, J. Lin, K. Liang, Residual joint adaptation adversarial network for intelligent transfer fault diagnosis, Mechanical Systems and Signal Processing 145 (2020) 106962.

[76] T. Li, X. Su, W. Liu, W. Liang, M.-Y. Hsieh, Z. Chen, X. Liu, H. Zhang, Memory-augmented meta-learning on meta-path for fast adaptation cold-start recommendation, Connection Science 34 (2022) 301–18.

[77] J. Lin, H. Shao, X. Zhou, B. Cai, B. Liu, Generalized maml for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals, Expert Systems with Applications (2023) 120696.

[78] W. Zhou, N. Li, Semi-supervised prototype network with cbam and data selector for few-shot bearing fault diagnosis, in: 2022 4th International Conference on Data-driven Optimization of Complex Systems (DOCS), IEEE, 2022, pp. 1–6.

[79] X. Xing, W. Guo, X. Wan, An improved multidimensional distance siamese network for bearing fault diagnosis with few labelled data, in: 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), IEEE, 2021, pp. 1–6.

[80] H. Ruan, Y. Wang, X. Li, Y. Qin, B. Tang, An enhanced non-local weakly supervised fault diagnosis method for rotating machinery, Measurement 189 (2022) 110433.