Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Confidence-Guided Selective Refinement Network for Imbalanced Gearbox Fault Diagnosis

Wei Zhou
North China Electric Power University
Guiji Tang
tanggjlk@ncepu.edu.cn
North China Electric Power University
Xiaolong Wang
North China Electric Power University
Zhenli Xu
North China Electric Power University
Kai Sun
North China Electric Power University
Lihong Cheng
North China Electric Power University

Research Article

Keywords: Gearbox fault diagnosis, Class imbalance, Mamba, Dynamic routing, Selective refinement, Vibration signal analysis

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11579574/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Confidence-Guided Selective Refinement Network for Imbalanced Gearbox Fault Diagnosis

Wei Zhou, Guiji Tang[a], Xiaolong Wang, Zhenli Xu, Kai Sun and Lihong Cheng
Department of Mechanical Engineering, North China Electric Power University, Baoding, 071003, Hebei, China
[a] Correspondence: tanggjlk@ncepu.edu.cn

Abstract: In intelligent gearbox fault diagnosis, class imbalance remains a major challenge: majority-class samples dominate the training process, whereas minority fault samples are scarce and difficult to characterize. Existing methods usually ignore differences in diagnostic difficulty among samples, resulting in insufficient recognition of minority and ambiguous fault categories. To address this problem, this paper proposes a confidence-guided selective refinement network, termed CGH-RMNet, for imbalanced gearbox fault diagnosis. The proposed framework decomposes the original diagnosis task into two stages: majority-class coarse screening and difficult-sample fine-grained refinement. In the first stage, a lightweight ResNet branch performs coarse discrimination and confidence estimation, allowing high-confidence majority-class samples to be directly accepted. In the second stage, samples predicted as complex categories or assigned low confidence are routed to a Mamba-based expert branch for refined full-class identification. Furthermore, an imbalance-aware optimization strategy is designed by combining class-weighted learning and label smoothing to enhance the representation of minority and hard samples. Experiments are conducted on the XJTU gearbox dataset and the MCC-THU Gearbox dataset under constant-speed, variable-speed, and variable-load conditions with different imbalance ratios. The results show that CGH-RMNet consistently outperforms several representative comparison methods in terms of diagnostic accuracy, confusion matrix distribution, and t-SNE-based feature separability. Ablation experiments further verify the effectiveness of the Mamba expert branch, the low-confidence routing mechanism, and the imbalance-aware learning strategy. The proposed method provides an efficient and robust solution for gearbox fault diagnosis under complex operating conditions and severe class imbalance.

Keywords: Gearbox fault diagnosis, Class imbalance, Mamba, Dynamic routing, Selective refinement, Vibration signal analysis

1 Introduction

Gearboxes are critical transmission components in wind turbines, high-speed trains, aerospace systems, and advanced manufacturing equipment. Their health condition directly affects system reliability, operational safety, and maintenance costs. Under harsh service conditions, gearboxes are frequently subjected to variable speeds, heavy loads, impact excitations, and other adverse factors, which may induce faults such as pitting, wear, tooth breakage, and bearing defects. Previous studies have shown that gear surface pitting can alter time-varying mesh stiffness and generate characteristic sidebands in vibration responses(Wang and Zhu 2023), while journal bearing failures in wind turbine gearboxes may become a key bottleneck affecting reliability and frictional losses (Yin et al. 2025). Therefore, accurate and timely gearbox fault diagnosis is essential for condition-based maintenance and predictive maintenance. In addition, data-driven prognostic studies have demonstrated the practical value of

intelligent models in predicting gearbox-related failures and supporting maintenance scheduling(Zhang et al. 2025).

Vibration signals and Supervisory Control and Data Acquisition (SCADA) data are two major data sources for gearbox health monitoring. SCADA-based methods provide a cost-effective solution because such data are generally already available in wind turbine control systems. For example, temperature-prediction-based machine learning models have been used to detect gearbox anomalies before severe failures occur(Rashid et al. 2026), and fine-tuned Transformer encoders have been developed for semi-supervised early detection of wind turbine gearbox faults using SCADA data(España et al. 2026). Meanwhile, vibration-based monitoring remains highly sensitive to local mechanical damage. Recent studies have further combined vibration features with explainable machine learning to classify the fault severity of wind turbine gearboxes and improve diagnostic transparency(Arika et al. 2026).

With the rapid development of intelligent diagnosis, deep learning-based fault diagnosis has become an important approach for gearbox fault diagnosis. Deep learning can automatically mine latent fault-related features from vibration signals, thereby overcoming the strong dependence of traditional diagnostic methods on expert knowledge and adapting to the increasing sample complexity caused by the rapid growth of data. Reference(Ahmad et al. 2024)indicated that deep learning methods for planetary gearbox diagnosis have been widely applied to variable working conditions, strong noise environments, imbalanced data, limited labeled samples, and zero-fault-sample scenarios. In addition to end-to-end deep networks, advanced feature extraction methods remain effective for noisy gearbox signals. For example, EMHFDE combined with RFE and NGO-LSSVM has been proposed to extract discriminative fault features from wind turbine gearbox vibration signals(Wang et al. 2026). Furthermore, operator-learning-based frameworks have also been employed to decouple degradation-related dynamics from operating-condition variations under variable speed–load conditions(Lakshmaiya 2026). These studies indicate that intelligent gearbox diagnosis is gradually evolving from conventional feature classification toward robust, adaptive, and transferable representation learning.

However, practical gearbox diagnosis is still constrained by fault data scarcity and class imbalance. In industrial systems, gearboxes usually operate in healthy states for most of their service life, whereas fault samples are rare, costly to obtain, and sometimes unsafe to collect. To address this issue, physics-based data generation methods have been introduced to synthesize mechanism-consistent fault samples for wind turbine gearbox diagnosis under scarce fault data conditions (Han et al. 2026). For imbalanced gearbox fault diagnosis, dynamic normalized supervised contrastive learning has been proposed to enhance the representation capability of minority-class samples and hard samples(Dong et al. 2024), while wavelet-transform-based improved residual networks have also been used to diagnose single and simultaneous gearbox faults under imbalanced data conditions (Wang et al. 2024). More recently, generative modeling has emerged as a promising solution to fault-sample scarcity. Long et al. developed a gated attention-enhanced improved conditional diffusion model, termed GA-ICDM, to generate class-conditional one-dimensional vibration signals for planetary gearboxes (Long et al. 2026) .Although these methods improve diagnostic performance, they are still mainly limited to a single-stage classification paradigm.

Another challenge is that class imbalance often coexists with domain shift, open-set categories, and varying operating conditions. To address extreme imbalanced domain adaptation, DACA introduced class-aware adversarial clustering, category alignment, reweighting, and re-margining for cross-domain diagnosis(Li et al. 2024). CIJDAN further considered class-imbalanced cross-domain diagnosis by combining domain adversarial training, augmented balancing, and dual-weight subdomain alignment to

improve diagnostic performance (Xiang et al. 2026). For gearbox data scarcity, vibration-mechanism-driven wavelet attention transfer learning and time-frequency transfer learning methods have been proposed to improve transferable feature extraction (Benattia et al. 2026; Zhu et al. 2026). Wang et al. further developed a dynamic dual-scale normalized fusion network for cross-domain gearbox diagnosis under variable-speed and imbalanced conditions (Wang et al. 2025a). The framework employed parallel graph-convolution branches to extract multi-scale features, aligned fine-grained global and local subdomain distributions through dual-scale normalized fusion, and used a dynamic soft-threshold feedback strategy to assign pseudo-labels while reducing the bias toward majority classes. In addition, open-set gearbox diagnosis has been investigated through reinforcement-learning-guided domain-calibrated prototype learning, aiming to identify unknown faults in unseen domains(Liu et al. 2026b).

Recent studies have extended imbalanced fault diagnosis to more complex diagnostic scenarios. A multi-task dynamic balanced learning network reformulated compound gearbox fault diagnosis as a component-level multi-task learning problem under severe imbalance (Xie et al. 2026). Reinforcement learning with balanced reward and cost-equilibrium sampling has been proposed for imbalanced gearbox diagnosis to actively guide the model in exploring minority-class samples (Liu et al. 2026c). In long-term monitoring scenarios, progressive calibrated feature association networks and convolutional-attention fusion networks have been developed for imbalanced or few-shot class-incremental fault diagnosis(Guo et al. 2025; Li et al. 2026). In addition, TFDL-GAN introduced a time-frequency dual-loss generation mechanism to improve the fidelity of minority fault samples for aero-engine accessory gearbox diagnosis(Sun et al. 2025). Beyond improving the generative model itself, Liu et al. proposed a diffusion-based augmentation framework with dual-stage quality gating (Liu et al. 2026a). In this framework, samples generated by a denoising diffusion probabilistic model were screened using both physical constraints and statistical consistency criteria, thereby improving the physical plausibility and practical usability of synthetic minority-class samples. The method achieved substantial Macro-F1 improvements under imbalance ratios ranging from 1:5 to 1:20 and demonstrated robustness across gearbox and bearing datasets. Collectively, these studies show that class imbalance is no longer treated as an isolated data-distribution problem, but is increasingly coupled with fault-sample scarcity, synthetic-sample quality, domain shift, compound faults, incremental learning, and computational efficiency.

Despite the above advances, most existing methods still process all samples using the same diagnostic model. This design ignores the differences in diagnostic difficulty among samples. Majority-class samples with clear features can be reliably classified by a lightweight model, whereas minority-class, ambiguous, or low-confidence samples require stronger representation learning capability. Moreover, current imbalanced diagnosis methods usually emphasize improvements in diagnostic accuracy, while paying insufficient attention to confidence reliability, selective sample routing, and the trade-off between diagnostic performance and computational cost.

To overcome these limitations, this paper proposes a confidence-guided two-stage selective refinement framework for gearbox fault diagnosis under imbalanced data. A confidence threshold is introduced to balance diagnostic performance and the activation cost of the expert branch. In the first stage, a lightweight residual network is employed for coarse-grained diagnosis and confidence estimation. Samples predicted as representative majority classes with high confidence are directly accepted, whereas ambiguous or low-confidence samples are routed to a Mamba-based refinement network for fine-grained diagnosis. Temperature scaling is introduced to calibrate the confidence of the coarse classifier, and a validation-based threshold selection strategy is adopted to balance diagnostic accuracy and expert model activation cost. In this way, the proposed method allocates diagnostic resources according to sample

difficulty and improves minority-class recognition performance without unnecessarily increasing inference cost.

The remainder of this paper is organized as follows. Section 2 presents the related theoretical background and recent advances in imbalanced gearbox fault diagnosis and sequence modeling. Section 3 describes the proposed CGH-RMNet in detail, including the majority-class screening branch, the Mamba-based refinement branch, the confidence-guided routing mechanism, and the imbalance-aware optimization strategy. Section 4 verifies the superiority and effectiveness of the proposed method through comparative experiments, visualization analysis, and ablation studies. Finally, Section 5 concludes this paper and discusses future research directions.

2 Related work

State space sequence modeling has recently emerged as an effective paradigm for long-sequence representation learning. Unlike conventional CNNs and RNNs, structured state space models can describe temporal evolution through latent states and provide an efficient way to capture long-range dependencies in sequential data(Gu et al. 2022). This property is particularly valuable for mechanical fault diagnosis, where vibration signals usually contain periodic impulses, transient components, and long-term degradation-related patterns. However, early structured state space models are generally based on time-invariant transition mechanisms, which may limit their ability to adaptively focus on fault-sensitive components under complex operating conditions.

To overcome this limitation, Mamba introduces a selective state space mechanism in which model parameters are dynamically conditioned on the input sequence(Gu and Dao 2024). By selectively propagating or forgetting information along the temporal dimension, Mamba improves context-aware sequence modeling while maintaining linear computational complexity. This advantage makes it suitable for rotating machinery fault diagnosis, where diagnostic models must simultaneously capture local fault impulses, global temporal dependencies, and noise-robust representations from long vibration sequences.

Recently, Mamba and selective state space models have been gradually introduced into intelligent fault diagnosis. Wang et al. proposed SSTG, an interpretable spatio-temporal selective state-space model for multi-sensor data fusion, which integrates selective state space modeling with graph-based spatio-temporal representation to capture fault evolution over time(Wang et al. 2025b). Xia et al. developed BMTM-net by combining 2D–1D fusion with bidirectional multi-granularity Transformer-Mamba, aiming to enhance inter-sensor correlation modeling, temporal feature extraction, and noise robustness in rotating machinery diagnosis(Xia et al. 2025). In addition, Pang et al. proposed Mamba-FeNet, which incorporates physical prior knowledge into a Mamba-based framework for interpretable bearing fault feature extraction, showing the potential of combining mechanism information with selective sequence modeling(Pang et al. 2026).

Although these studies demonstrate the effectiveness of state space sequence modeling in fault diagnosis, most existing methods still focus on improving feature extraction within a unified diagnostic model. The relationship among sample difficulty, prediction confidence, selective inference, and computational cost has not been sufficiently explored. Therefore, introducing state space sequence modeling into a selective refinement framework may provide a more flexible solution for imbalanced gearbox fault diagnosis, especially for ambiguous or minority-class samples that require stronger temporal representation capability.

3 Proposed method

3.1 Overall architecture of the proposed two-stage cascade model

In gearbox fault diagnosis under class-imbalanced conditions, majority-class samples are sufficiently abundant, enabling the network to learn their representative feature distributions more comprehensively. In contrast, minority-class samples are limited in number, and their fault-related features are prone to being overwhelmed by those of the majority classes during training. Owing to the learning bias caused by differences in sample size, the model tends to favor the recognition of majority-class samples, while exhibiting insufficient discriminative capability for minority-class samples and complex samples located near class boundaries. This may consequently lead to misclassification and degrade the overall classification performance.

To address these issues, this study proposes a confidence-guided hierarchical ResNet--Mamba network, termed CGH-RMNet. The overall architecture of the proposed model is illustrated in Fig.1. Instead of directly performing full-class classification in a single step, CGH-RMNet decomposes the original imbalanced diagnosis task into two collaborative stages: majority-class coarse screening and difficult-sample fine-grained refinement. The main idea is to first identify samples that can be reliably determined by a lightweight model, and then allocate stronger temporal modeling capability only to samples that are complex or uncertain.

Specifically, the first stage adopts a one-dimensional ResNet as the majority-screening backbone. The original fine-grained label space is reconstructed into a coarse-grained label space, in which representative majority classes are retained as independent categories, while minority and ambiguous categories are merged into a complex category. This design reduces the learning difficulty of the first-stage classifier and prevents difficult minority samples from being prematurely forced into unreliable fine-grained decisions. For a given input vibration signal, the ResNet branch outputs coarse-grained logits and the corresponding prediction confidence. High-confidence samples predicted as majority classes are directly accepted and mapped back to the original label space, thereby enabling efficient diagnosis for easily distinguishable samples.

For samples that are predicted as the complex category or assigned low confidence by the ResNet branch, CGH-RMNet activates the second-stage expert refinement branch. This branch adopts Mamba as the temporal modeling backbone and performs refined classification in the original full-class label space. Unlike a simple minority-class classifier, the Mamba branch is designed as a full-class refinement expert. Therefore, even when a low-confidence majority-class sample is routed to this branch, it can still be corrected and reassigned to its original majority class. By selectively modeling these difficult samples, the Mamba branch strengthens the extraction of long-range temporal dependencies, weak fault-sensitive patterns, and dynamic evolution characteristics in vibration signals, which are especially important under variable-speed, variable-load, and severe class-imbalance conditions.

In addition, a confidence-guided dynamic routing mechanism is introduced to connect the two stages. The routing decision is jointly determined by the coarse prediction category and its confidence score. Samples belonging to the complex category or with confidence lower than a predefined threshold are forwarded to the Mamba-based expert branch, whereas reliable majority-class samples are directly output by the ResNet branch. In this way, the proposed model adaptively allocates diagnostic resources according to sample difficulty, rather than applying the same computational process to all samples.

To further alleviate the optimization bias caused by imbalanced class distributions, an imbalance-aware learning strategy is incorporated into the hierarchical framework. The coarse-grained branch is optimized using class-weighted cross-entropy loss, so that the merged complex category is not overwhelmed by the majority categories. Meanwhile, the fine-grained branch adopts a class-weighted label smoothing loss to enhance the learning of minority and hard samples while reducing overfitting to limited routed samples. The final training objective jointly optimizes the coarse screening branch and the expert refinement branch, enabling CGH-RMNet to achieve both efficient majority-class recognition and

robust minority-class discrimination. Overall, the proposed CGH-RMNet has three key characteristics. First, it reformulates imbalanced gearbox fault diagnosis as a hierarchical decision process, which reduces the difficulty of direct full-class classification under severe imbalance. Second, it introduces confidence-guided selective routing, allowing the model to distinguish easy samples from difficult samples and allocate modeling capacity accordingly. Third, it employs a Mamba-based expert branch to refine low-confidence and complex samples by capturing long-range temporal dependencies in vibration signals. Through the cooperation of these components, CGH-RMNet improves the recognition performance of minority and ambiguous fault samples while maintaining the efficiency and stability of the overall diagnostic process.

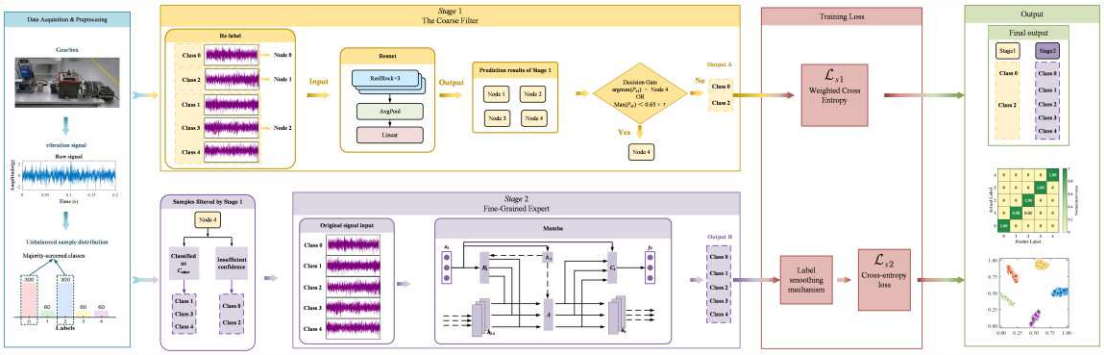


Fig.1 Overall architecture of the proposed CGH-RMNet for imbalanced gearbox fault diagnosis.

3.2 Stage 1: Majority-screening backbone for primary condition discrimination

In the proposed model, the coarse-grained discrimination branch does not directly undertake the complete classification task. Taking the constant-speed experimental dataset used in this study as an example, the diagnosis task consists of one healthy state and four fault states, forming a five-class classification problem. Accordingly, the original samples are assigned five labels, ranging from 0 to 4. Among them, labels 0 and 2 correspond to majority-class samples, whereas the remaining labels correspond to minority-class samples.

Therefore, in the coarse-grained discrimination branch, the original labels 0 and 2 are treated as two independent easy-to-discriminate categories and are reassigned pseudo-labels 0 and 1, respectively. Meanwhile, labels 1, 3, and 4 are uniformly mapped into a complex category and assigned the pseudo-label 2. For convenience in the following discussion, this category C_{other} is referred to as the complex class. The mapping relationship can be expressed as:

$$\tilde{y}=\begin{cases}0, & y=0,\\ 1, & y=2,\\ 2, & y\in\{1,3,4\}.\end{cases} \tag{1}$$

This label reconstruction serves two purposes. First, it reduces the difficulty of the initial classification task, enabling the network to prioritize the learning of stable categories with highly separable feature distributions. Second, it preserves the opportunity for samples belonging to the complex class to be processed by the subsequent fine-grained branch, thereby preventing the model from making premature and irreversible decisions on difficult-to-classify samples at the first stage.Therefore, the objective of the first branch is not to finely distinguish all gearbox health states. Instead, it aims to preferentially identify categories characterized by stable and discriminative feature distributions using a model with relatively low computational complexity, while retaining samples that cannot be reliably determined for further analysis in the subsequent branch.

It is worth noting that, in the coarse-grained branch, the samples classified as C_{other} consist not

only of the complex categories with original labels 1, 3, and 4, but also of the hard-to-distinguish samples from the majority classes. In contrast, the samples classified as pseudo-labels 0 and 1, corresponding to the majority classes, are directly output as part of the final diagnostic results. Only the samples assigned to C_{other} are forwarded to the second-stage fine-grained branch for further discrimination.

The primary objective of this branch is to perform efficient and stable preliminary discrimination under relatively low model complexity, while preserving uncertain samples for further analysis in the subsequent fine-grained branch. In accordance with the principle of parsimony implied by Occam's razor, the complexity of the model should be consistent with the requirements of the task. Accordingly, ResNet, a simple yet effective convolutional classification network, is employed as the backbone of the coarse-grained discrimination branch in this study.

The one-dimensional ResNet can effectively extract local impact features and periodic fluctuation information from vibration signals. Therefore, this study adopts the one-dimensional ResNet as the backbone of the coarse-grained discrimination branch, which is used for preliminary learning of the input vibration signals and discrimination among majority classes. This branch first encodes the input signal through a front-end module composed of one-dimensional convolution, batch normalization, ReLU activation, and max pooling. The purpose of this stage is to map the original two-channel signal into a higher-dimensional feature space and preliminarily compress redundant high-frequency fluctuations, thereby providing a more stable input representation for subsequent residual feature learning.

After the front-end module, the features are sequentially passed through three groups of one-dimensional residual layers. The first group maintains a relatively low number of channels and is used to extract basic local features. The second and third groups gradually expand the channel dimension and perform temporal downsampling through strided convolution, enabling the network to model the relationship between fault impacts and periodic responses within a larger receptive field. After multiple residual mappings, the network compresses the temporal features into a sample-level global representation through adaptive average pooling, and then outputs the three-class coarse-grained logits through a fully connected layer. This process can be expressed as:

$$z_i^{s_1}, h_i^{s_1} = f_{\text{res}}(X_i) \tag{2}$$

where $h_i^{s_1}$ denotes the sample-level feature extracted by the coarse-grained branch, and $z_i^{s_1}$ represents the three-class coarse-grained output.

In summary, under class-imbalanced training conditions, the coarse-grained branch also plays an important role in alleviating the dominance of majority classes during model optimization. By merging several categories into the C_{other} class, the model can reduce the competition among fine-grained minority classes at the initial training stage, thereby encouraging the optimization objective to focus more on distinguishing stable categories from complex categories. This strategy helps enhance the model's ability to preserve complex samples and provides more discriminative fine-grained targets for the fine-grained temporal modeling branch.

3.3 Stage 2: Expert refinement module for hard-sample identification

The fine-grained temporal modeling branch is designed to perform refined discrimination in the original label space for samples selected through dynamic filtering. This branch directly models the five-class fault states. However, its input does not include all training samples; instead, it consists of complex samples selected by the confidence-gating mechanism. Specifically, the samples entering this branch can be divided into two categories: first, samples predicted as C_{other} by the coarse-grained decision branch; second, samples that are predicted as majority-class samples but have confidence scores lower than a predefined threshold. Therefore, this branch essentially performs refined full-class classification.

The fine-grained temporal modeling branch mainly focuses on minority-class samples and low-confidence complex samples. This requires the network to possess a stronger capability for temporal dependency modeling, so that it can capture weak impacts, intermittent fluctuations, and dynamic variation patterns across different temporal segments from long-sequence vibration signals. To this end, Mamba is adopted as the backbone structure of the fine-grained temporal modeling branch in this study.

Gearbox vibration signals contain not only local impulsive components but also state-evolution information across temporal segments. Conventional convolutional structures are advantageous in capturing local patterns. However, when fault-related features are distributed over a relatively long temporal range, or when dependencies exist among different temporal segments, relying solely on local convolution may be insufficient to characterize global dynamic features. The Mamba structure can model long-range temporal dependencies while maintaining high computational efficiency, making it suitable for refined discrimination of complex samples.

Specifically, the fine-grained branch first employs two one-dimensional convolutional layers to perform downsampling and channel mapping on the input signal. The first convolutional layer maps the original dual-channel input into a lower-dimensional intermediate representation and performs preliminary temporal compression. The second convolutional layer further increases the feature dimension to d_{model}, thereby satisfying the input requirement of the Mamba module. Subsequently, the feature tensor is transformed from the convolutional representation format [B, C, L] to the sequential representation format [B, L, C], where B denotes the batch size, L denotes the downsampled sequence length, and $C = d_{\text{model}}$ denotes the feature dimension.This transformation allows each temporal position to be treated as an individual sequence unit, thereby facilitating temporal dependency modeling by the Mamba module.

After passing through two Mamba modules, the network obtains sequential feature representations containing long-range dependency information. Subsequently, layer normalization is employed to enhance the stability of the feature distribution, followed by average aggregation along the temporal dimension to obtain the sample-level representation $h_i^{s_2}$. Finally, the sample-level representation is fed into a classifier composed of a fully connected layer, ReLU activation, dropout, and a linear classification layer, which outputs the logits corresponding to the five original classes:

$$z_i^{s_2}, h_i^{s_2} = f_{\text{mamba}}(X_i) \tag{3}$$

where $z_i^{s_2} \in \mathbb{R}^5$ denotes the prediction result of the fine-grained branch in the original label space.

In summary, the fine-grained branch is not simply a minority-class classifier. Although its primary role is to enhance the discriminative capability for complex classes and boundary samples, it is still trained and evaluated within the complete five-class label space. This design prevents the model from introducing overly strong prior assumptions in the coarse-grained decision branch. For example, when a low-confidence sample whose true label is class 0 or class 2 is forwarded by the coarse-grained branch to the fine-grained branch, the Mamba branch can still reassign it to class 0 or class 2, rather than being restricted to classes 1, 3, and 4. Therefore, this branch has the capability to correct uncertain decisions made by the coarse-grained branch.

From a functional perspective, the Mamba branch mainly addresses two issues. First, it provides stronger temporal representation capability for complex samples, enabling the model to capture the evolution patterns of fault characteristics over a longer temporal range. Second, by processing only the filtered samples, it concentrates the high-complexity modeling capacity on the samples that require refined discrimination, thereby avoiding redundant modeling of all samples. This design not only improves the classification capability for minority classes under imbalanced data conditions but also helps maintain the efficiency of the overall inference process.

3.4 Confidence-Guided Dynamic Sample Screening Mechanism

As discussed above, in the imbalanced five-class classification task, samples from different categories exhibit substantial differences in both class frequency and discrimination difficulty. For majority classes with relatively sufficient samples and clear decision boundaries, the model can usually produce reliable predictions through the coarse-grained branch. In contrast, samples belonging to minority classes, boundary regions, or categories with ambiguous inter-class relationships often require further fine-grained discrimination. To this end, a confidence-guided dynamic sample screening mechanism is introduced to adaptively determine the inference path of each input sample according to the predicted coarse category and the corresponding confidence score.

Let the original label space of the five-class task be denoted as

$$\mathcal{Y} = \{0, 1, 2, 3, 4\}. \tag{4}$$

The two majority classes are mapped to two independent coarse-grained categories, while the remaining categories are merged into the merged complex category, denoted by C_{other}. Therefore, the coarse-grained label space can be expressed as

$$\mathcal{C}^{s1} = \{C_1, C_2, C_{\text{other}}\} \tag{5}$$

where the number of coarse-grained categories is $K_c = 3$.

For an input sample X_i, the coarse-grained branch first outputs a three-dimensional logit vector:

$$\mathbf{z}_i^{s1} = [z_{i,1}^{s1}, z_{i,2}^{s1}, z_{i,3}^{s1}] \tag{6}$$

where the third dimension corresponds to the merged category C_{other}. The corresponding coarse-grained prediction probability distribution is obtained by the Softmax function:

$$\mathbf{p}_i^{s1} = \text{Softmax}(\mathbf{z}_i^{s1}) \tag{7}$$

The maximum predicted probability is used as the confidence score of sample X_i in the coarse-grained branch:

$$\gamma_i = \max_k p_{i,k}^{s1} \tag{8}$$

Accordingly, the predicted coarse-grained category is defined as

$$\hat{c}_i = \arg\max_k p_{i,k}^{s1} \tag{9}$$

Based on the coarse prediction and confidence score, a routing indicator m_i is defined to determine whether sample X_i should be forwarded to the fine-grained branch:

$$m_i = \mathbb{I}[(\hat{c}_i = C_{\text{other}}) \vee (\gamma_i < \tau)] \tag{10}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, $\vee$ represents the logical OR operation, and τ is the confidence threshold. When $m_i = 1$, the sample is either predicted as the complex category set C_{other} or assigned to a majority coarse category with insufficient confidence. Such samples are regarded as difficult or uncertain samples and are therefore forwarded to the Mamba-based fine-grained branch for further discrimination. When $m_i = 0$, the sample is considered to be confidently classified by the coarse-grained branch and can be directly mapped back to the original label space through a predefined mapping function.

The final prediction of sample X_i is formulated as

$$\hat{y}_i = \begin{cases} g(\hat{c}_i), & m_i = 0, \\ \arg\max\limits_{k \in \{1,\ldots,5\}} z_{i,k}^{s2}, & m_i = 1, \end{cases} \tag{11}$$

where $g(\cdot)$ denotes the mapping function from the majority coarse-grained categories to the original label space, and $\mathbf{z}_i^{s2}$ represents the five-class logits produced by the Mamba-based fine-grained branch. It should be noted that $g(\cdot)$ is only applied to the two independent majority coarse categories. Samples predicted as C_{other} are not directly mapped to the original label space, but are instead forwarded to the fine-grained branch for re-identification.

The threshold τ controls the proportion of samples activated by the fine-grained branch. A larger τ forwards more low-confidence samples to the fine-grained branch, which may improve the classification capability for complex and boundary samples but also increases computational cost. Conversely, a smaller τ reduces the number of samples processed by the fine-grained branch and improves inference efficiency, but may ignore some uncertain samples. Therefore, τ should be selected to balance fine-grained recognition performance and computational efficiency. Through this dynamic screening mechanism, the proposed model can adaptively allocate samples to different inference paths according to their discrimination difficulty, thereby enabling rapid recognition of majority-class samples while improving the fine-grained identification capability for complex and low-confidence samples.

3.5 Imbalance-Aware Optimization Strategy

To enable stable optimization of the hierarchical collaborative network under imbalanced data distributions, task-specific loss functions are designed for the coarse-grained and fine-grained branches, respectively. The two branches are then optimized jointly in an end-to-end manner. Since the coarse-grained and fine-grained branches undertake different classification objectives, the imbalance patterns encountered by them are not identical. The primary objective of the coarse-grained branch is to distinguish the two majority categories from the complex category set C_{other}, whereas the fine-grained branch aims to enhance the detailed classification capability for complex and minority-class samples. Therefore, using an identical loss design for both branches may fail to sufficiently reflect their different optimization priorities.

For the five-class task considered in this study, the original label space is defined as

$$\mathcal{Y} = \{0, 1, 2, 3, 4\} \tag{12}$$

The two majority classes are mapped to two independent coarse-grained categories, while the remaining categories are merged into C_{other}. Let $h(\cdot)$ denote the label mapping function from the original label space to the coarse-grained label space. Then,

$$h(y_i) \in \{1, 2, C_{\text{other}}\} \tag{13}$$

The number of coarse-grained categories is therefore $K_c = 3$.

Let n_c^{s1} denote the number of training samples belonging to the coarse-grained category c. To mitigate the dominance of majority categories during optimization, the class weight of the coarse-grained branch is defined according to the inverse square-root of the class frequency:

$$\tilde{w}_c^{s1} = \frac{1}{\sqrt{n_c^{s1} + \delta}} \tag{14}$$

$$w_c^{s1} = \frac{K_c \tilde{w}_c^{s1}}{\sum\limits_{j=1}^{K_c} \tilde{w}_j^{s1}} \tag{15}$$

where δ is a small smoothing constant used to avoid division by zero. The coarse-grained branch is optimized using a class-weighted cross-entropy loss:

$$\mathcal{L}_{s1} = -\frac{1}{N}\sum_{i=1}^{N} w_{h(y_i)}^{s1} \log \frac{\exp\left(z_{i,h(y_i)}^{s1}\right)}{\sum_{k=1}^{3} \exp\left(z_{i,k}^{s1}\right)} \tag{16}$$

Here, $\mathbf{z}_i^{s1}$ denotes the logits generated by the coarse-grained branch, $h(y_i)$ is the mapped coarse-grained label of sample $X_i, w_{h(y_i)}^{s1}$ denotes the corresponding class weight, and \$N\$ is the number of samples in the mini-batch. Since C_{other} contains multiple original classes and may include minority or more difficult samples, assigning an appropriate weight to this category helps reduce the risk that it is overwhelmed by the majority categories during training.

For the fine-grained branch, a class-weighted label smoothing loss is adopted. Since the fine-grained branch receives dynamically screened samples, its input distribution may differ from the original training distribution. Directly applying the standard cross-entropy loss may lead the model to overfit hard labels, especially when the number of routed samples is limited. Therefore, label smoothing is introduced to transform the original one-hot target distribution into a softened label distribution. For a sample with the ground-truth label y_i, the smoothed target distribution is defined as

$$q_{i,k} = \begin{cases} 1-\epsilon, & k = y_i, \\ \dfrac{\epsilon}{K-1}, & k \neq y_i, \end{cases} \tag{17}$$

where $K = 5$ is the number of original fine-grained categories and ϵ is the label smoothing coefficient.

Let n_k denote the number of training samples in the original fine-grained category k. The class weight of the fine-grained branch is defined as

$$\tilde{w}_k^{s2} = \frac{1}{\sqrt{n_k + \delta}} \tag{18}$$

$$w_k^{s2} = \frac{K\tilde{w}_k^{s2}}{\sum_{j=1}^{K} \tilde{w}_j^{s2}} \tag{19}$$

Let $\mathcal{M}$ denote the set of samples routed to the fine-grained branch in the current mini-batch. Let $p_{i,k}^{s2}$ be the predicted probability of class k generated by the Mamba-based fine-grained branch. The loss of the fine-grained branch is formulated as

$$\mathcal{L}_{s2} = -\frac{1}{|\mathcal{M}|}\sum_{i\in\mathcal{M}} w_{y_i}^{s2} \sum_{k=1}^{5} q_{i,k} \log p_{i,k}^{s2}. \tag{20}$$

If no sample in the current mini-batch is routed to the fine-grained branch, $\mathcal{L}_{s2}$ is set to zero, and only the coarse-grained branch is updated. Otherwise, both branches participate in the joint optimization.

The overall training objective is defined as

$$\mathcal{L} = \mathcal{L}_{s1} + \beta\mathcal{L}_{s2} \tag{21}$$

where β controls the contribution of the fine-grained branch loss to the overall objective. In this study, β is set to 2.0 to emphasize the learning of complex and minority-class samples. Since the fine-grained branch processes only the screened samples, the number of samples available to this branch is usually

smaller than that of the coarse-grained branch. Without increasing its loss contribution, the overall optimization may be dominated by the coarse-grained task, thereby weakening the learning effectiveness of the fine-grained branch.

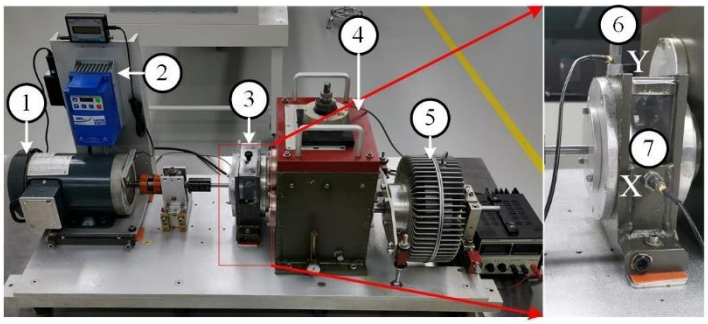
During training, the two branches are optimized using the AdamW optimizer with independent learning-rate scheduling strategies, so that different modules can converge stably according to their own optimization dynamics. In addition, to alleviate gradient fluctuations caused by complex and low-confidence samples, gradient norm clipping is applied after backpropagation. By constraining the overall gradient norm within a predefined threshold, this operation suppresses excessively large parameter updates caused by abnormal mini-batches and improves the numerical stability of the joint optimization process.

4 Experimental verification and result analysis

In this section, the XJTU Gearbox dataset from Xi'an Jiaotong University and the MCC-THU Gearbox dataset from Tsinghua University are used to evaluate the fault diagnosis generalization capability of the proposed algorithm for gearboxes under constant-speed, variable-speed, and variable-load operating conditions.

4.1 Case 1:Fault Diagnosis Experiment on a Gearbox under Constant-Speed Conditions

To verify the performance of CGH-RMNet in gearbox fault diagnosis under constant-speed conditions, the XJTU Gearbox dataset is adopted in this study. As shown in Fig.2, the gearbox test rig consists of a driving motor, a controller, a planetary gearbox, a parallel gearbox, and a brake. Two one-dimensional accelerometers, with the model number PCB352C04, are installed on the planetary gearbox in the X and Y directions, respectively(Li et al. 2022). In this experiment, four types of planetary gear faults are artificially introduced into the planetary gearbox, including tooth surface wear, missing tooth, tooth root crack, and broken tooth. The sampling frequency is 20.48kHz, and the driving motor operates at a rotational speed of 1800r/min.



1. Motor, 2. Controller,3. Planetary gearbox, 4. Parallel gearbox,
5. Brake,6-7. Accelerometor in horizontal, and vertical direction.

Fig.2 Test rig for the constant-speed gearbox fault diagnosis experiment

To evaluate the fault recognition accuracy of CGH-RMNet under class-imbalanced conditions, the normal state and the missing-tooth state in the gearbox test rig are regarded as majority-class samples, whereas the other three health states are regarded as minority-class samples. The number of training samples for each majority class is set to 300. For the minority classes, four different imbalance ratios (IRs), namely 5, 10, 15, and 30, are considered in this study. Accordingly, the numbers of training samples for each minority class are 60, 30, 20, and 10, respectively. In the test set, however, each class contains 100 samples to ensure a balanced and fair evaluation. The description of the constant-speed gearbox fault diagnosis dataset is presented in Table 1.

Table 1 Description of the constant-speed gearbox fault diagnosis dataset.

Label	Health state	Training samples	Test samples
0	Normal state	300	
1	Broken tooth	60/30/20/10	
2	Missing tooth	300	100
3	Tooth root crack	60/30/20/10	
4	Tooth wear	60/30/20/10	

To further verify the superiority of the proposed method, six commonly used diagnostic algorithms for class-imbalanced learning are selected as comparison methods, namely Oversampling ResNet, Reweight ResNet, Multi Attention 1D Convolutional Neural Network (MA1DCNN), One Dimensional Dilated Convolutional Neural Network (ODC-Net), signal feature-based support vector machine (SVM-features), and Transformer.

After setting the random seeds, each experiment is repeated five times. All models are trained for 50 epochs with a batch size of 64. The experiments are implemented based on the PyTorch framework. The operating environment is Windows 11, equipped with an Intel Core i5-13600KF CPU at 3.50GHz, 32GB RAM, and an NVIDIA GeForce RTX 4060Ti GPU with 24GB memory.

The experimental accuracy results are presented in Table 2. Under the imbalance ratio of 5:1, both the proposed method and the comparison methods achieve relatively high classification accuracy. However, as the imbalance ratio increases, the accuracy of all algorithms decreases to different degrees. Since the comparison methods are also optimized for class-imbalanced learning, they exhibit satisfactory performance in imbalanced sample classification. Nevertheless, when the IR reaches 30:1, the accuracy of all comparison methods drops below 96%, whereas CGH-RMNet maintains a stable classification accuracy above 98%. These results demonstrate that CGH-RMNet can effectively improve the fault diagnosis accuracy under constant-speed conditions by partitioning majority and minority classes into simple and complex samples and decomposing the original single-stage classification task into two collaborative classification subtasks.

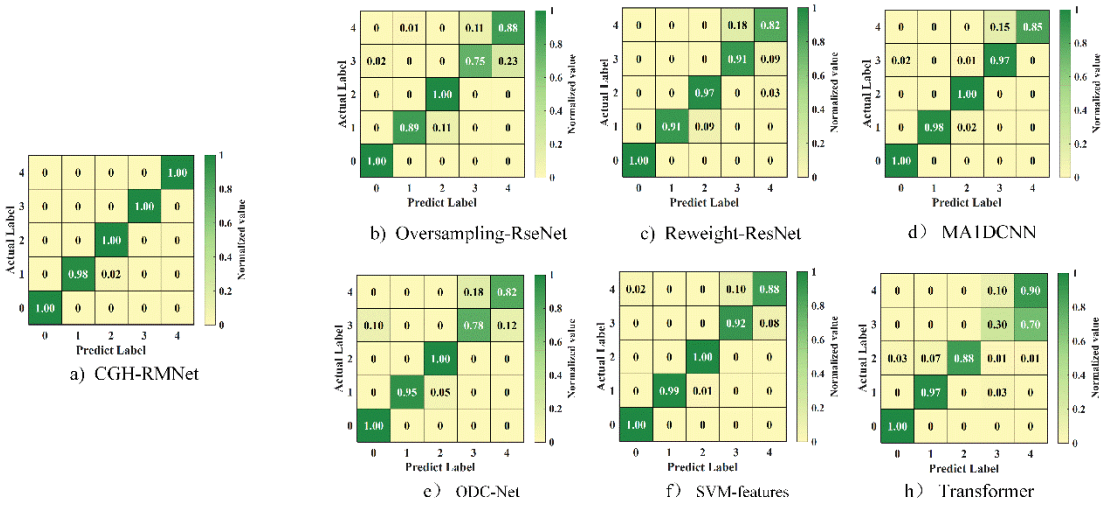


Fig.3 Confusion matrices of constant-speed gearbox fault diagnosis under four imbalance ratios.

Table 2 Comparison results of constant-speed gearbox fault diagnosis under four imbalance ratios.

Compared method	IRs (%)			
	5:1	10:1	15:1	30:1
CGH-RMNet	99.90±0.10	99.70±0.30	99.40±0.20	98.90±0.70
Oversampling-	99.80±0.20	99.40±0.20	98.60±1.00	89.90±4.50

ResNet				
Reweight-ResNet	99.40±0.60	99.60±0.20	98.50±0.70	92.90±3.30
MA1DCNN	99.50±0.30	99.10±0.50	98.40±1.40	93.40±1.80
ODC-Net	99.80±0.20	99.60±0.20	99.30±0.30	93.10±2.50
SVM-features	98.60±0.00	98.10±0.30	97.60±0.40	95.30±0.50
Transformer	97.70±1.50	95.50±3.70	91.40±3.20	80.70±2.50

Fig.3 presents the confusion matrices of different methods at IR = 30, visualizing their fault recognition results. As observed from Fig.3, the proposed method misclassifies only one test sample of missing-tooth samples as broken-tooth samples, while achieving correct classification for almost all gearbox fault samples. In contrast, all comparison methods produce more than two misclassified samples. These results further demonstrate that the proposed method exhibits superior accuracy and robustness in gearbox fault diagnosis under constant-speed conditions.

To investigate the characteristics of feature maps extracted from different layers of deep neural networks, t-distributed stochastic neighbor embedding (t-SNE) is commonly used to project high-dimensional feature representations of test samples into a two-dimensional space. The t-SNE visualization results of all algorithms in this experiment are shown in Fig.4. As illustrated in Fig.4, the proposed method can extract feature representations with clear distributions and well-defined decision boundaries. Samples with different labels form compact and distinguishable clusters in the feature space, indicating high inter-class separability. In contrast, although some comparison methods achieve relatively good separation between certain classes, class-overlapping phenomena still exist. The t-SNE visualization further demonstrates that the proposed method can effectively distinguish the latent characteristics among different gearbox faults and exhibits favorable accuracy and robustness under severely imbalanced scenarios.

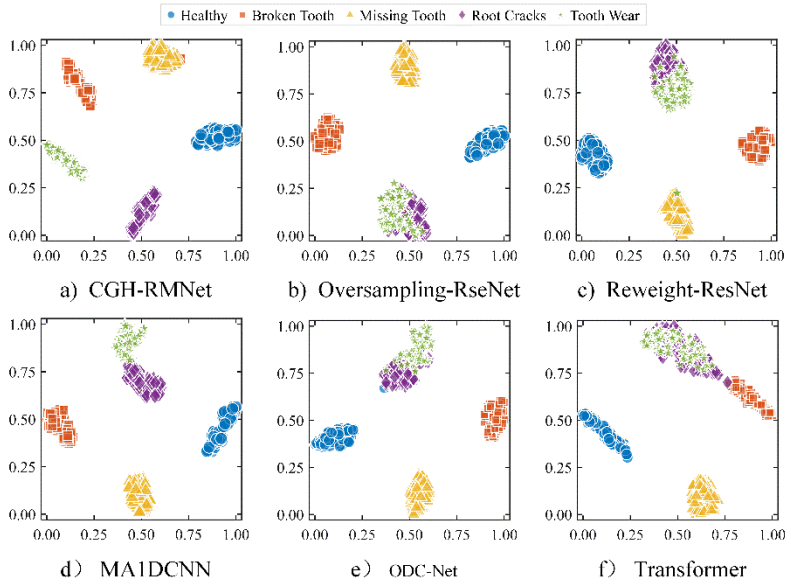


Fig.4 Feature visualization results of different methods for constant-speed gearbox fault diagnosis

To quantitatively evaluate the feature clustering and separability performance of different methods, the within-class scatter matrix S_w and the between-class scatter matrix S_b are introduced. Specifically, S_w characterizes the compactness of samples within the same class, whereas S_b describes the separability among samples from different classes. Therefore, S_w and S_b are defined as follows:

$$S_w = \sum_{r=1}^{l} \sum_{v_k \in L_r} (v_k - m'_r)(v_k - m'_r)^T, \quad k = 1, 2, ..., d,$$
$$S_b = \sum_{r=1}^{l} (m'_r - \bar{m}_v)(m'_r - \bar{m}_v)^T. \tag{22}$$

where m'_r denotes the mean feature vector of the r-th class $r = 1, 2, ..., l$, and $\bar{m}_v = \frac{1}{l}\sum_{r=1}^{l} m'_r$ denotes the mean feature vector of all classes. In general, larger values of S_b/S_w and $(S_b + S_w)/S_w$ indicate better feature extraction performance. In other words, these two evaluation metrics can be used to quantitatively describe and assess the feature representation capability of the network, which are defined as follows:

$$J_1 = \frac{|S_b|}{|S_w|},$$
$$J_2 = \frac{|S_w + S_b|}{|S_w|}. \tag{23}$$

Therefore, the calculated between-class and within-class evaluation metrics of the six diagnostic methods are presented in Table 3. According to the evaluation criterion, a larger value of J_i indicates better classification performance. The feature representations learned by CGH-RMNet exhibit stronger discriminability and clustering performance than those learned by the other methods. This demonstrates that the proposed method can effectively learn minority-class features under the interference of majority-class samples.

Table 3 Quantitative evaluation of feature separability based on t-SNE.

Methods	Feature evaluation metrics	
	J_1	J_2
CGH-RMNet	16.86	17.86
Oversampling-ResNet	11.74	12.81
Reweight-ResNet	9.68	10.68
MA1DCNN	14.55	15.55
ODC-Net	13.24	14.24
Transformer	13.62	14.62

4.2 Case 2: Fault Diagnosis Experiment on a Gearbox under Variable-Speed Conditions

Case 1 has demonstrated the effectiveness of CGH-RMNet in gearbox fault diagnosis under constant-speed conditions. However, in practical industrial applications, mechanical equipment often operates under more complex non-stationary working conditions. Compared with the stationary vibration signals collected from gearboxes operating at constant speed, vibration signals under variable-speed conditions exhibit non-stationary, time-varying, and frequency-modulated characteristics, leading to more severe feature coupling. These characteristics pose significant challenges to the accurate extraction of fault-related features. Therefore, to further verify the generalization capability and robustness of the proposed model under complex dynamic operating conditions, Case 2 is introduced in this section to conduct gearbox fault diagnosis experiments under variable-speed conditions.

The rotating machinery fault diagnosis dataset released by the Mechanical Control and Condition Monitoring Laboratory of Tsinghua University(MCC-THU Gearbox dataset) is adopted for experimental validation(Chen et al. 2024). The experimental system mainly consists of a three-phase asynchronous motor, a torque sensor, a two-stage parallel gearbox, a magnetic powder brake, and a measurement system. The sampling frequency used in this experiment is 12.8kHz, and the experimental setup is shown

in Fig.5.

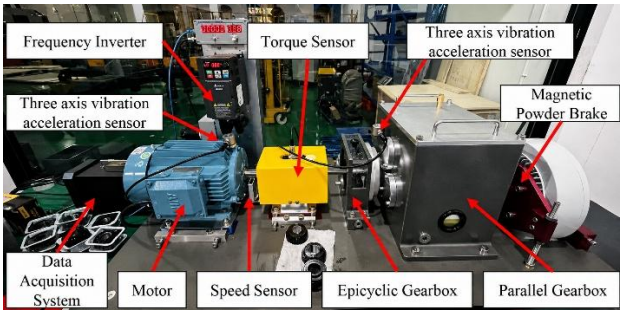


Fig.5 Test rig for the variable-speed gearbox fault diagnosis experiment.

During the experiment, single-gear faults and compound faults involving gears and bearings are introduced by laser etching. The single-gear faults include missing tooth, gear wear, gear pitting, tooth crack, and broken tooth. Meanwhile, the compound faults include the broken-tooth and bearing inner-race compound fault, as well as the broken-tooth and bearing outer-race compound fault. Thus, a total of seven fault states are considered. Similarly, the seven fault states and the healthy state are assigned labels from 0 to 7. The normal state, gear pitting, and the broken-tooth and bearing inner-race compound fault are regarded as majority classes. The performance of different methods is also investigated under four different imbalance ratios (IRs) on this dataset. The description of the dataset is presented in Table 4.

Table 4 Description of the MCC-THU Gearbox dataset

Label	Health condition	Training samples	Test samples
0	Normal condition	600	300
1	Missing tooth	120/60/40/20	
2	Gear pitting	600	
3	Gear wear	120/60/40/20	
4	Tooth-root crack	120/60/40/20	
5	Broken tooth	120/60/40/20	
6	Broken tooth and inner-race fault	600	
7	Broken tooth and outer-race fault	120/60/40/20	

In this section, the comparison algorithms are the same as those used in Case 1, and the implementation platform is also kept consistent. Since the operating conditions of the variable-speed gearbox are more complex and the corresponding vibration signals are more difficult to identify, the number of training epochs is set to 100 in this experiment, while the remaining parameters are kept identical to those in Case 1. The results obtained by repeating each algorithm five times are presented in Table 5 and Fig.6.

Table 5 Comparison results of variable-speed gearbox fault diagnosis under four imbalance ratios.

Compared methods	IRs (%)			
	5:1	10:1	15:1	30:1
CGH-RMNet	96.23±0.27	92.22±1.10	91.82±0.65	80.98±0.69
Oversampling-ResNet	95.25±0.80	89.35±2.51	83.19±1.66	72.92±1.44
Reweight-ResNet	95.22±1.26	91.33±1.37	87.03±0.66	74.53±1.63
MA1DCNN	92.60±1.74	88.53±2.51	76.26±4.43	62.32±3.21
ODC-Net	91.85±1.65	83.06±4.71	75.84±1.48	61.95±1.79
SVM-features	90.40±0.23	78.32±0.22	71.30±0.14	58.42±0.61
Transformer	79.61±1.38	65.35±1.08	58.84±1.62	50.33±1.10

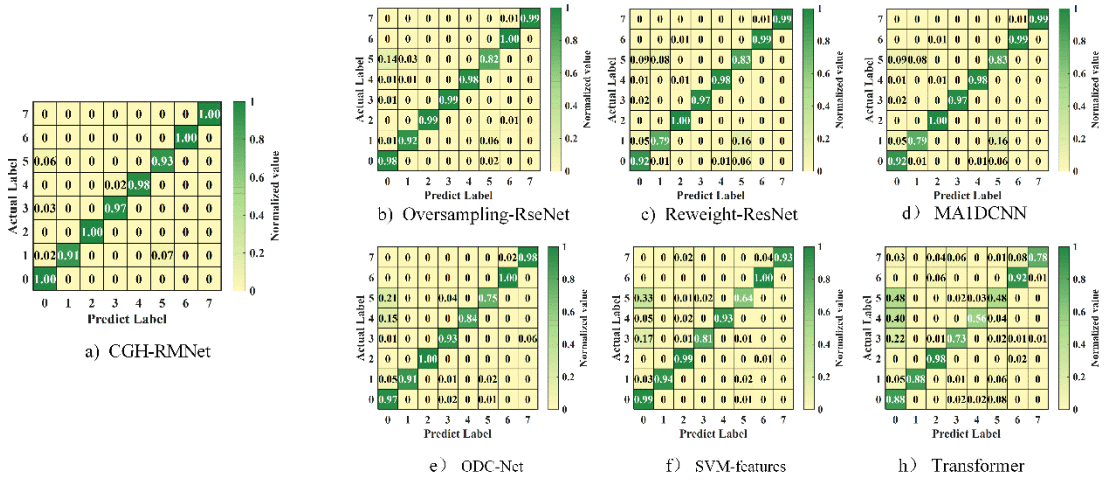


Fig.6 Confusion matrices of different methods for variable-speed gearbox fault diagnosis

As shown in Table 5, the complex variable-speed operating conditions have a significant influence on the diagnostic performance of all algorithms. As the imbalance ratio increases, the diagnostic accuracies of all methods decrease to varying degrees, indicating that the non-stationary vibration characteristics under variable-speed conditions and class imbalance jointly increase the difficulty of fault identification. In contrast, CGH-RMNet consistently achieves the highest diagnostic accuracy under all imbalance ratios. In particular, when the imbalance ratio reaches 30:1, CGH-RMNet still achieves an accuracy of 80.98%, whereas the accuracies of all comparison methods are lower than 75%. Moreover, in terms of accuracy degradation, CGH-RMNet decreases by only 15.25 percentage points from 5:1 to 30:1, which is notably smaller than the decreases observed in Oversampling-ResNet, Reweight-ResNet, and MA1DCNN. These results indicate that the proposed method has stronger stability and resistance to class imbalance under severe imbalance conditions. This advantage can be attributed to the collaborative mechanism of majority-class coarse screening and difficult-sample fine-grained identification, which effectively mitigates the negative impact of class imbalance on diagnostic performance under complex variable-speed conditions.

Fig.6 presents the normalized confusion matrices of different methods for variable-speed gearbox fault diagnosis. It can be observed that the predictions of CGH-RMNet are mainly concentrated along the diagonal, indicating that the proposed method achieves high recognition rates for different health states. In contrast, the comparison methods exhibit more evident off-diagonal elements, suggesting that they are more likely to confuse certain fault categories under complex variable-speed conditions. Specifically, CGH-RMNet shows stable recognition performance for both majority classes and fault categories with relatively clear decision boundaries, with only slight confusion among a few similar classes. By comparison, although Oversampling-ResNet, Reweight-ResNet, MA1DCNN, and ODC-Net can correctly identify some categories, they still suffer from evident misclassification for minority and similar fault classes. The off-diagonal distributions of SVM-features and Transformer are more dispersed, indicating their limited adaptability to non-stationary vibration characteristics and class imbalance under variable-speed conditions.

Fig.7 and Table 6 present the feature visualization and quantitative separability evaluation results of different methods based on t-SNE. As shown in Fig.7, the features extracted by CGH-RMNet exhibit more compact intra-class distributions and clearer inter-class boundaries than those of the comparison methods. Samples belonging to different health states are well separated in the two-dimensional feature space, indicating that the proposed method can learn more discriminative fault representations under complex gearbox fault conditions. In contrast, the feature distributions obtained by Oversampling-

ResNet, Reweight-ResNet, MA1DCNN, ODC-Net, and Transformer show varying degrees of overlap, especially among similar fault categories, which may lead to increased misclassification. The quantitative results in Table 6 further support this observation. CGH-RMNet achieves the highest J_1 and J_2 values, reaching 5.42 and 6.42, respectively, which are significantly higher than those of the other methods. These results demonstrate that CGH-RMNet has stronger feature clustering and class-separability capabilities, thereby improving the reliability of gearbox fault diagnosis.

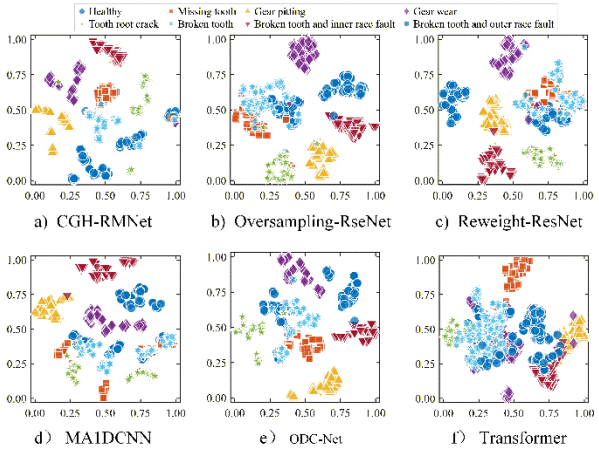


Fig.7 Feature visualization results of different methods for variable-speed gearbox fault diagnosis

Table 6 Quantitative evaluation of feature separability based on t-SNE.

Methods	Feature evaluation metrics	
	J1	J2
CGH-RMNet	5.42	6.42
Oversampling-ResNet	2.54	3.54
Reweight-ResNet	2.95	3.95
MA1DCNN	1.01	2.01
ODC-Net	1.68	2.68
Transformer	1.76	2.76

These results demonstrate that frequency modulation, feature coupling, and class imbalance under variable-speed operating conditions substantially increase the difficulty of fault identification. Benefiting from the coarse screening of reliable majority-class samples by the ResNet branch and the fine-grained refinement of complex and low-confidence samples by the Mamba expert branch, CGH-RMNet effectively reduces the misclassification risk of minority and boundary samples. Therefore, the proposed method exhibits stronger class-discriminative capability and diagnostic robustness under complex variable-speed conditions.

4.3 Case 3: Fault Diagnosis Experiment on a Gearbox under Variable-load Conditions

In practical service conditions, gearboxes are subjected not only to speed fluctuations but also to frequent variations in working load. Cases 1 and 2 have verified the superior performance of CGH-RMNet under constant-speed and variable-speed conditions, respectively. However, compared with the frequency modulation caused by speed variations, variable-load conditions directly lead to severe amplitude fluctuations and nonlinear evolution of impact energy in vibration signals, and may even induce pronounced multimodal feature overlap. To comprehensively evaluate the environmental adaptability and diagnostic reliability of the proposed model under multidimensional complex operating conditions, Case 3 is introduced in this section to further conduct gearbox fault diagnosis experiments

under variable-load conditions.

The MCC-THU Gearbox dataset dataset is still adopted in this experiment. The experimental platform, fault categories, sampling frequency, comparison methods, and model parameter settings are kept consistent with those in Case 2. Different from Case 2, this section focuses on evaluating the fault diagnosis performance of gearboxes under variable-load conditions, where the diagnostic capability of different methods is examined under alternating load levels. To ensure a fair comparison, the same imbalance ratio settings and training strategy are used in this case, and each method is independently repeated five times.The final results are reported in Table 7 in terms of the mean accuracy and standard deviation.

Table 7 Comparison results of variable-load gearbox fault diagnosis under four imbalance ratios.

Compared methods	IRs（%）			
	5:1	10:1	15:1	30:1
CGH-RMNet	99.67±0.10	98.31±0.19	97.57±0.29	93.13±0.35
Oversampling-ResNet	99.04±0.13	97.67±0.78	95.67±0.89	91.63±0.70
Reweight-ResNet	99.33±0.30	96.88±0.81	96.17±0.97	91.54±4.66
MA1DCNN	99.50±0.15	97.92±1.03	92.04±2.61	88.96±1.84
ODC-Net	99.63±0.12	97.54±0.87	95.42±1.69	87.96±1.37
SVM-features	98.46±0.08	95.79±0.35	91.33±0.05	79.29±0.60
Transformer	92.88±0.92	86.75±1.06	80.21±1.13	70.46±1.07

As reported in Table 7, the proposed CGH-RMNet achieves the best diagnostic performance under all variable-load imbalance settings. Different from the variable-speed case, load fluctuations mainly alter the amplitude distribution and impact energy of vibration responses, which may weaken the consistency of fault features and increase the overlap among different health states. Under such challenging conditions, most comparison methods show a clear performance deterioration as the imbalance ratio increases, especially SVM-features and Transformer, whose accuracies decrease markedly at the imbalance ratio of 30:1. By contrast, CGH-RMNet maintains accuracies of 99.67%, 98.31%, 97.57%, and 93.13% from 5:1 to 30:1, respectively, consistently outperforming the other methods. Notably, under the most severe imbalance ratio of 30:1, CGH-RMNet exceeds Oversampling-ResNet and Reweight-ResNet by 1.50 and 1.59 percentage points, respectively, and shows a much larger advantage over MA1DCNN, ODC-Net, SVM-features, and Transformer. Moreover, the accuracy of CGH-RMNet decreases by only 6.54 percentage points from 5:1 to 30:1, indicating that its performance is less sensitive to the reduction of minority-class samples. This superiority suggests that the proposed hierarchical diagnosis strategy can effectively preserve reliable majority-class decisions while assigning difficult and low-confidence samples to the Mamba-based refinement branch, thereby improving the robustness of gearbox fault diagnosis under variable-load and class-imbalanced conditions.

To further investigate whether the proposed method can maintain stable recognition performance across different fault categories, the normalized confusion matrices of all methods under variable-load conditions are presented in Fig.8. This analysis provides a more detailed view of class-wise classification performance and reveals the specific misclassification patterns among similar fault states.

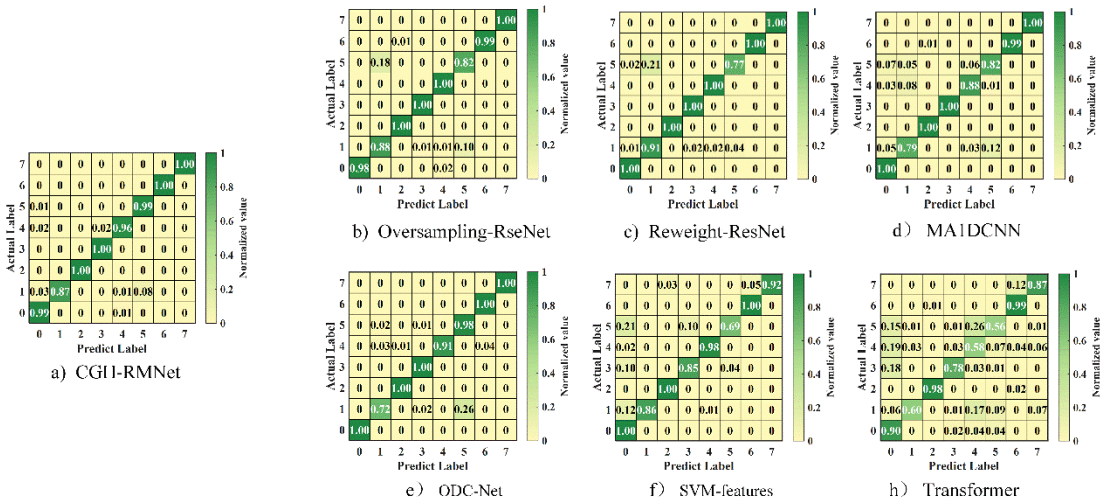


Fig.8 Confusion matrices of different methods for variable-load gearbox fault diagnosis

As shown in Fig.8, CGH-RMNet exhibits a more concentrated diagonal distribution than the comparison methods, indicating that most health states can be correctly identified under variable-load conditions. Only slight confusion occurs between a few similar fault categories, while the majority of classes maintain high recognition rates. In contrast, the comparison methods show more evident off-diagonal elements, especially for minority and compound fault classes, suggesting that load variations intensify feature overlap and increase the risk of misclassification. Overall, the confusion matrix results further confirm that CGH-RMNet has stronger class-wise discriminative capability and diagnostic robustness under variable-load and imbalanced conditions.

Table 8 Quantitative evaluation of feature separability based on t-SNE.

Methods	Feature evaluation metrics	
	J1	J2
CGH-RMNet	6.43	7.43
Oversampling-ResNet	3.04	4.04
Reweight-ResNet	3.36	4.36
MA1DCNN	1.56	2.56
ODC-Net	2.48	3.48
Transformer	2.02	3.02

Fig.9 and Table 8 present the t-SNE feature visualization and quantitative feature separability results of different methods. As shown in Fig.9, the feature distributions learned by CGH-RMNet exhibit more compact intra-class clustering and clearer inter-class boundaries than those obtained by the comparison methods. Different health states are well separated in the two-dimensional feature space, indicating that the proposed method can extract more discriminative fault representations under complex operating conditions. In contrast, the features extracted by Oversampling-ResNet, Reweight-ResNet, MA1DCNN, ODC-Net, and Transformer show different degrees of class overlap and scattered distributions, especially for similar fault categories, which may increase the risk of misclassification.

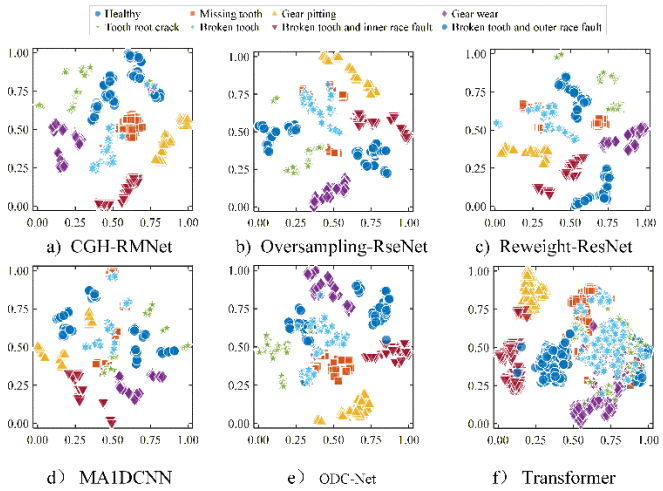


Fig.9 Feature visualization results of different methods for variable-load gearbox fault diagnosis

The quantitative results in Table 8 further support the above observation. CGH-RMNet achieves the highest feature separability scores, with J_1 and J_2 reaching 6.43 and 7.43, respectively. These values are significantly higher than those of the comparison methods, indicating stronger inter-class separability and intra-class compactness. In particular, compared with Reweight-ResNet, which obtains the second-best results, CGH-RMNet still improves J_1 and J_2 by 3.07 and 3.07, respectively. This demonstrates that the proposed hierarchical ResNet–Mamba framework can effectively enhance the representation of minority and difficult samples, thereby improving the overall discriminative capability of the model.

In summary, the experimental results under variable-load conditions demonstrate the effectiveness and robustness of CGH-RMNet in handling complex operating environments and class-imbalanced data. Compared with constant-speed and variable-speed conditions, variable-load conditions introduce more significant amplitude fluctuations and impact energy variations, which increase feature overlap among different fault categories. Nevertheless, CGH-RMNet consistently achieves the best diagnostic accuracy under all imbalance ratios and maintains clear advantages in both confusion matrix analysis and feature separability evaluation. The results show that the proposed hierarchical framework can effectively combine efficient majority-class screening with fine-grained refinement for difficult samples, thereby improving minority-class recognition and maintaining stable diagnostic performance under complex variable-load conditions.

4.4 Ablation study

To verify the effectiveness of each component in the proposed algorithm, three groups of ablation experiments are designed to investigate the contributions of the Mamba expert branch, the dynamic screening mechanism, and the imbalance-aware learning strategy. The experiments are conducted on the MCC-THU Gearbox dataset dataset under variable-speed gearbox fault diagnosis conditions. To ensure a fair comparison, the other network structures and parameter settings are kept unchanged in all ablation experiments, and only the investigated module is modified.

Specifically, three ablation models are constructed as follows. Model A replaces Mamba with Transformer to evaluate the contribution of the Mamba-based sequence modeling module in Stage 2. Model B removes the weighted loss term in the imbalance-aware learning strategy and adopts the standard cross-entropy loss for two-stage iterative optimization. Model C removes the low-confidence routing criterion and retains only the C_{other}-based routing strategy, aiming to evaluate the contribution of selective refinement for low-confidence samples. The experimental results are presented in Table 9.

Table 9 Ablation results of variable-speed gearbox fault diagnosis

Compared methods	IRs (%)			
	5:1	10:1	15:1	30:1
CGH-RMNet	96.23±0.27	92.22±1.10	91.82±0.65	80.98±0.69
Model A	87.37±4.45	74.73±1.10	71.30±1.41	58.73±0.92
Model B	92.47±6.21	91.39±0.46	87.88±1.17	74.79±0.72
Model C	93.74±1.31	83.70±0.64	76.28±0.52	59.23±0.99

As shown in Table 9, the complete CGH-RMNet consistently achieves the best performance under all imbalance ratios, demonstrating the effectiveness of the proposed collaborative design. When Mamba is replaced by Transformer in Model A, the diagnostic accuracy decreases significantly from 96.23%, 92.22%, 91.82%, and 80.98% to 87.37%, 74.73%, 71.30%, and 58.73%, respectively. This indicates that the Mamba-based expert branch is more effective in modeling long-range temporal dependencies and refining difficult samples under variable-speed conditions. Compared with Model A, Model B exhibits a relatively smaller but still noticeable performance degradation, especially under the severe imbalance ratio of 30:1, where the accuracy drops from 80.98% to 74.79%. This result verifies that the imbalance-aware weighted loss contributes to improving the learning of minority and hard samples. In addition, Model C shows a sharp decrease in accuracy when the low-confidence routing criterion is removed, particularly at imbalance ratios of 15:1 and 30:1, where the accuracies decrease by 15.54 and 21.75 percentage points, respectively. This demonstrates that relying only on the C_{other}-based routing strategy is insufficient, because some ambiguous majority-class samples and boundary samples may be incorrectly accepted by the first-stage classifier without further refinement. Overall, the ablation results confirm that the Mamba expert branch, imbalance-aware optimization, and low-confidence dynamic routing mechanism are all necessary for the proposed framework, among which the Mamba-based refinement branch and low-confidence sample routing play particularly important roles under severe class-imbalanced conditions.

4.5 Visualization analysis

To further analyze the selective routing behavior of CGH-RMNet, the number of samples routed to the Stage-2 expert branch during training is visualized in Fig. 10. It can be observed that, although the number of routed samples is relatively large at the beginning of training, it gradually decreases as the model optimization proceeds and finally converges to a relatively stable range. This trend indicates that the coarse ResNet branch progressively improves its ability to identify reliable majority-class samples, while the routing mechanism increasingly focuses the expert branch on a controlled subset of difficult samples.

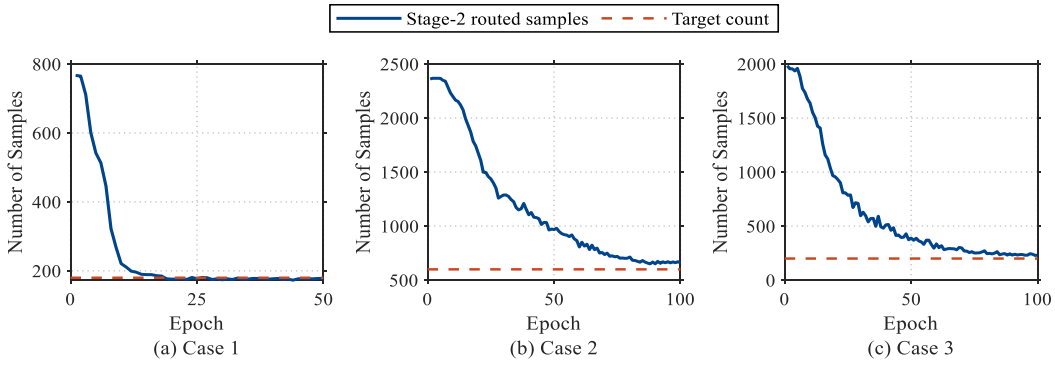


Fig.10 Evolution of the number of samples routed to the Stage-2 expert branch during training

The convergence of the routed-sample count toward the target count demonstrates that the proposed confidence-guided routing strategy can effectively regulate the activation scale of the Mamba expert branch. Instead of forwarding excessive samples throughout the entire training process, CGH-RMNet gradually reduces unnecessary expert activation after the coarse classifier becomes stable. Therefore, the visualization results confirm that the proposed framework can maintain a balance between diagnostic refinement and computational efficiency by adaptively controlling the number of samples entering Stage 2.

5 Conclusion and future research

To effectively address the class imbalance problem in gearbox fault diagnosis, this study proposes a confidence-guided hierarchical ResNet–Mamba network, termed CGH-RMNet. The proposed framework consists of several key components. First, a lightweight ResNet branch is designed for majority-class coarse screening, enabling high-confidence majority-class samples to be identified efficiently. Second, a Mamba-based expert branch is introduced to perform refined identification of complex and low-confidence samples by capturing long-range temporal dependencies in vibration signals. Finally, a confidence-guided routing mechanism and an imbalance-aware optimization strategy are developed to adaptively allocate diagnostic resources and enhance the learning of minority-class and hard samples. The effectiveness of CGH-RMNet is verified on the XJTU and MCC-THU gearbox datasets under constant-speed, variable-speed, and variable-load conditions. Experimental results demonstrate that the proposed method achieves superior performance over several representative comparison methods in terms of diagnostic accuracy, confusion matrix distribution, and feature separability.

Although the proposed method improves the performance of imbalanced gearbox fault diagnosis, some limitations remain. All experiments in this study are conducted under laboratory conditions, and validation using real industrial field data is still insufficient. Future work will further evaluate the diagnostic capability of the proposed method under practical field conditions and improve its applicability in real industrial environments.

Funding

This research was funded by the Graduate Innovation Fund Project of Hebei Province (Project Number: CXZZBS2025183), the National Natural Science Foundation of China (Project Number: 52675142), and the Basic Research Funds for Central Universities (Project Number: 2025MS133).

CRediT authorship contribution statement

Wei Zhou: Methodology, Formal analysis, Software, Writing - original draft. Guiji Tang: Data curation, Validation, Writing - review & editing, Funding acquisition. Xiaolong Wang: Methodology, Validation, Writing - review & editing. Zhenli Xu: Validation, Writing - review & editing. Kai Sun: Validation, Writing - review & editing.Lihong Cheng: Validation, Writing - review & editing.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmad H, Cheng W, Xing J, Wang W, Du S, Li L, Zhang R, Chen X, Lu J (2024) Deep learning-based fault diagnosis of planetary gearbox: A systematic review. Journal of Manufacturing Systems 77:730–745. https://doi.org/10.1016/j.jmsy.2024.10.004

Arika AM, Huda ASN, Ahmad S, Islam A (2026) Vibration-based multi-class fault severity classification of wind turbine gearboxes using explainable machine learning. Energy Conversion and Management: X 30:101621. https://doi.org/10.1016/j.ecmx.2026.101621

Benattia MEB, Habbouche H, Benkedjouh T, Amirat Y, Benbouzid M (2026) Robust planetary gearbox fault diagnosis through time–frequency analysis and transfer learning. Engineering Applications of Artificial Intelligence 167:113844. https://doi.org/10.1016/j.engappai.2026.113844

Chen S, Liu Z, He X, Zou D, Zhou D (2024) Multi-mode fault diagnosis datasets of gearbox under variable working conditions. Data in Brief 54:110453. https://doi.org/https://doi.org/10.1016/j.dib.2024.110453

Dong Y, Jiang H, Jiang W, Xie L (2024) Dynamic normalization supervised contrastive network with multiscale compound attention mechanism for gearbox imbalanced fault diagnosis. Engineering Applications of Artificial Intelligence 133:108098. https://doi.org/10.1016/j.engappai.2024.108098

España S, Sánchez J, Puruncajas B, Castellani F, Tutivén C, Vidal Y (2026) Early detection of gearbox faults in wind turbines using a fine-tuned transformer encoder. Results in Engineering 30:109915. https://doi.org/10.1016/j.rineng.2026.109915

Gu A, Dao T (2024) Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv preprint arXiv:231200752

Gu A, Goel K, Ré C (2022) Efficiently Modeling Long Sequences with Structured State Spaces. arXiv preprint arXiv:211100396

Guo Z, Du W, Liu Z, Hu T, Yu Y, Li C (2025) Few-shot sample multi-class incremental fault diagnosis for gearbox based on convolutional-attention fusion network. Expert Systems with Applications 264:125918. https://doi.org/10.1016/j.eswa.2024.125918

Han S, Zheng J, Hu H, Xue M, Wu M (2026) A physics based data generation method using a dynamic model for wind turbine gearbox fault diagnosis in fault data scarce settings. Results in Engineering 30:109565. https://doi.org/10.1016/j.rineng.2026.109565

Lakshmaiya N (2026) Deep learning-enhanced intelligent gearbox health assessment framework integrating adaptive feature extraction and operator-guided classification for robust multi-speed load condition monitoring. Results in Engineering 30:111035. https://doi.org/10.1016/j.rineng.2026.111035

Li T, Zhou Z, Li S, Sun C, Yan R, Chen X (2022) The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. Mechanical Systems and Signal Processing 168:108653. https://doi.org/https://doi.org/10.1016/j.ymssp.2021.108653

Li Y, Zhu Y, Yu Y, Mao R, Ye L, Liu Y, Liu R, Lang T, Zhang J (2024) DACA: A domain adaptive fault diagnosis approach with class-aware based on cross-domain extreme imbalance data. Expert Systems with Applications 256:124944. https://doi.org/10.1016/j.eswa.2024.124944

Li Z, Zhang K, Zheng Q, Ding G, Du Q, Zhang H, Zhao C, Yuan Z (2026) Progressive calibrated feature association network for imbalanced class incremental fault diagnosis in bogie mechanical transmission systems. Advanced Engineering Informatics 69:104112. https://doi.org/10.1016/j.aei.2025.104112

Liu R, Yang S, Chen J, Xu H, Li C (2026a) Imbalanced fault diagnosis with diffusion-based augmentation and dual-stage quality gating. Meas Sci Technol 37(23). https://doi.org/10.1088/1361-6501/ae6f3e

Liu X, Qian K, Zhao Z, Yang F, Wei D, Bo L, Luo J (2026b) RL-guided domain-calibrated prototype network for gearbox open-set diagnosis. Pattern Recognition 180:114029. https://doi.org/10.1016/j.patcog.2026.114029

Liu X, Zhao Z, Wei D, Yang F, Bo L, Luo J (2026c) A fault diagnosis strategy for gearboxes based on reinforcement learning with balanced reward and cost-equilibrium sampling. Advanced Engineering Informatics 69:103946. https://doi.org/10.1016/j.aei.2025.103946

Long Y, Chen Z, Xu Z, Liang J, Wang Y (2026) A fault sample generation method for planetary gearbox based on GA-ICDM. Meas Sci Technol 37(15). https://doi.org/10.1088/1361-6501/ae5ac0

Pang B, Liu Q, Wang B, Xu Z, He C (2026) A physical knowledge-informed bearing fault feature extraction network based on Mamba. Mechanical Systems and Signal Processing 252:114272.

https://doi.org/10.1016/j.ymssp.2026.114272

Rashid H, Batunlu C, Belkhier Y, Benbouzid M (2026) Intelligent fault detection in wind turbine gearboxes: A machine learning approach using SCADA temperature data. Energy Conversion and Management: X 29:101573. https://doi.org/10.1016/j.ecmx.2026.101573

Sun Y, Xiao B, Yuan Y, Ma S (2025) TFDL-GAN for fault diagnosis of aero-engine accessory gearbox with imbalanced data. IFAC-PapersOnLine 59(20):1434–1439. https://doi.org/10.1016/j.ifacol.2025.11.358

Wang C, Tian X, Shao X, Wang R, Wang L (2025a) Data imbalanced fault diagnosis of gearbox transmission system under various speeds based on dynamic dual-scale normalized fusion network. Meas Sci Technol 36(1). https://doi.org/10.1088/1361-6501/ad99f0

Wang C, Wang Y, Zheng F, Li M, Xia R (2025b) SSTG: An interpretable spatio-temporal Selective State-Space Model for multi-sensor data fusion in intelligent diagnosis. Knowledge-Based Systems 316:113278. https://doi.org/10.1016/j.knosys.2025.113278

Wang S, Tian J, Liang P, Xu X, Yu Z, Liu S, Zhang D (2024) Single and simultaneous fault diagnosis of gearbox via wavelet transform and improved deep residual network under imbalanced data. Engineering Applications of Artificial Intelligence 133:108146. https://doi.org/10.1016/j.engappai.2024.108146

Wang S, Zhu R (2023) Research on dynamics and failure mechanism of herringbone planetary gearbox in wind turbine under gear surface pitting. Engineering Failure Analysis 146:107130. https://doi.org/10.1016/j.engfailanal.2023.107130

Wang X, Liang Z, Du Y (2026) Fault diagnosis method for wind turbine gearboxes based on EMHFDE-RFE-NGO-LSSVM. Measurement 280:121897. https://doi.org/10.1016/j.measurement.2026.121897

Xia E, Liu Y, Gong J, Dai X, Pan T, Wang S (2025) BMTM-net: A rotating machinery fault diagnosis network based on 2D-1D fusion with bidirectional multi-granularity transformer-mamba. Neurocomputing 656:131293. https://doi.org/10.1016/j.neucom.2025.131293

Xiang W, Liu S, Li H, Cao X, Han B (2026) Cross-domain fault diagnosis of rotating machinery based on class-imbalanced joint domain adaptation network. Knowledge-Based Systems 341:115823. https://doi.org/10.1016/j.knosys.2026.115823

Xie Z, Chen J, Zhang K, Wang ZJ (2026) Decoupling compound faults under data imbalance: A multi-task dynamic balanced learning network for gearbox fault diagnosis. Mechanical Systems and Signal Processing 256:114530. https://doi.org/10.1016/j.ymssp.2026.114530

Yin J, Meng X, Cheng S (2025) Failure modeling and multidimensional optimization of gearbox journal bearings to prevent wear failures and reduce friction in wind turbines. Engineering Failure Analysis 178:109748. https://doi.org/10.1016/j.engfailanal.2025.109748

Zhang W, Vatn J, Rasheed A (2025) Gearbox pump failure prognostics in offshore wind turbine by an integrated data-driven model. Applied Energy 380:124829. https://doi.org/10.1016/j.apenergy.2024.124829

Zhu P, Tang B, Deng L, Wei J, Li Z, Li Q (2026) Vibration mechanism driven discrete wavelet hybrid attention weighted transfer network for partial domain fault diagnosis of gearboxes under data scarcity. Engineering Applications of Artificial Intelligence 163:112994. https://doi.org/10.1016/j.engappai.2025.112994