

Imbalanced Learning for Gearbox Fault Detection via Attention-based Multireceptive Field Convolutional Neural Networks with An Adaptive Label Regulation Loss

Yadong Xu, Rui Shu, Sheng Li, Ke Feng, Member IEEE, Xiaolong Yang, Zhiheng Zhao, Senior Member IEEE, and George Q. Huang, Fellow IEEE

Abstract—Accurate gearbox fault identification is paramount for industrial production. In practice, gearboxes typically operate under normal conditions (rarely under faulty conditions), resulting in a long-tailed distribution of monitoring data. However, the majority of current algorithms are crafted based on the assumption of balanced sample distributions, which don't correspond with the prevalent conditions encountered in actual industrial settings. To cope with this challenge, an attention-based multireceptive field convolutional neural network (AMFCN) is established in this paper. This study's main contributions can be summarized as follows: 1) We introduce a global contextual attention module (GCAM) to instruct the model to focus on learning ample features; 2) We establish a hierarchical receptive field module (HRFM) to incorporate powerful multi-level learning capabilities into the AMFCN model; 3) We devise an adaptive label regulation loss (ALRL) to facilitate the model to obtain accurate fault identification results, particularly in situations with imbalanced data distributions. Two case studies show that the AMFCN model achieves 83.72% and 81.63% accuracy on two extremely imbalanced gearbox datasets, outperforming seven competitive algorithms.

Index Terms—Gearbox, fault identification, long-tailed distribution, global contextual attention module (GCAM),

hierarchical receptive field module (HRFM), adaptive label regulation loss (ALRL).

NOMENCLATURE

<i>ALRL</i>	Adaptive Label Regulation Loss
<i>AMFCN</i>	Attention-based multireceptive field convolutional neural network
<i>CAB</i>	Channel Attention Branch
<i>CNN</i>	Convolutional Neural Networks
<i>GCAM</i>	Global Contextual Attention Module
<i>HRFM</i>	Hierarchical Receptive Field Module
<i>KNN</i>	k-nearest Neighbors Algorithm
<i>MIXConv</i>	Mix Convolutional Layer
<i>MLP</i>	Multi-layer Perceptron
<i>PReLU</i>	Parametric Rectified Linear Unit
<i>SAB</i>	Spatial Attention Branch
<i>SMOTE</i>	Synthetic Minority Over-Sampling Technique
$t - SNE$	T-distributed stochastic neighbor embedding

I. INTRODUCTION

AS mechanical equipment continues to advance in precision, complexity, scale, and intelligence, the operational health of such machinery becomes pivotal for ensuring production process safety and product quality stability [1]. Among the essential components of rotating machinery, gearboxes stand out as the core support mechanism. The seamless and secure operation of machinery hinges on the optimal functioning of these gearboxes. Hence, in practical production settings, real-time dynamic monitoring, fault analysis, and predictive maintenance of gearboxes hold immense significance.

During gear operation, various factors such as fatigue, overload, and abrupt load changes can readily trigger surface peeling, leading to gear damage [2]. This peeling phenomenon arises from prolonged elastic deformation of the gear, subjected to repetitive twisting and bending stresses [3]. Furthermore, if the main shaft undergoes bending or sustains damage, the gearbox experiences recurring impact loads during operation, exacerbating potential damage. Moreover, inadequate lubrication or sealing of the gearbox may facilitate foreign particle intrusion, resulting in aggressive wear and subsequent

This work is supported by Natural Science Foundation of China (No. 52305557), Open Fund of State Key Laboratory of Intelligent Manufacturing Equipment and Technology (No. IMETKF2024022), Hong Kong RGC TRS Project(T32-707/22-N), Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011930), and Collaborative Research Fund (C7076-22GF). (Corresponding authors: Zhiheng Zhao).

Y. Xu, Z. Zhao, and G. Huang are with the Department of Industrial and Systems Engineering and also with Research Institute for Advanced Manufacturing, The Hong Kong Polytechnic University, Hong Kong, China

Z. Zhao is also with the state Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China

R. Shu is with School of Mechanical Engineering, Jiangsu University of Science and Technology, Zhenjiang, 212134, China

S. Li is with the College of Civil Engineering, Nanjing Forestry University, Nanjing, 210037, China.

Ke Feng is with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710054, China, and also with the School of Mechanical Engineering, Xi'an Jiaotong University, Shaanxi 710049, China.

X. Yang is with the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China.

peeling failures [4]. Hence, the foundation for ensuring the smooth operation of mechanical equipment lies in the effective condition monitoring and diagnosis of gear spalling faults.

The vibration signal collected during mechanical equipment operation serves as a crucial and dependable foundation for fault diagnosis [5]. Traditionally, fault diagnosis involves extracting pertinent fault characteristic information from the original vibration signal. This includes methods such as time domain analysis, frequency domain analysis, and time-frequency domain analysis, etc. Zhang et al. [6] put forward a total variational denoising-based model for critical feature extraction of bearings. Yu et al. [7] developed a time-frequency feature extraction approach to construct quantitative indicators for running status monitoring of bearings. Zhang et al. [8] devised a recursive frequency estimation algorithm, which utilized the Cholesky decomposition strategy for the fault frequency location of machines. Nikula et al. [9] designed a new time domain feature autocorrelation approach for locating the failure location of low-speed bearings. First, a filter was applied to locate the explicit span of the given fault signals; Then, some time-frequency indicators were calculated to strengthen the fault-related frequency; Finally, a novel search strategy was utilized to locate the exact fault frequency of the given signal. A new component separable synchroextraction transform approach was devised for extracting the critical fault-related components of the input signals [10]. First, a new ridge optimization scheme was utilized to extract potential fault information; Then, an intelligent time-varying filter was employed to locate the specific failure features; Finally, the concise fault-related information was reconstructed with high fidelity using the synchronous extraction method. The efficacy of the proposed approach is confirmed through the analysis of both simulated and experimental signals. While these methods are indeed effective, they heavily depend on the expertise of technicians. For less experienced personnel, achieving desired diagnostic outcomes with these methods can be challenging. Furthermore, the original vibration signals produced by modern mechanical equipment are voluminous and intricate. Given their complexity, traditional fault diagnosis methods struggle to accurately analyze and process such large amounts of complex data.

With the continuous accumulation of data and the improvement of computing power, the application of artificial intelligence technology in the industrial field has been greatly expanded and deepened. Many companies and organizations have begun to build large-scale models based on artificial intelligence to effectively respond to the complex challenges encountered in production operations. Looking around the historical monitoring data allows us to uncover patterns, trends, and anomalies that can signify various aspects of bearing health. In this context, deep learning-based diagnostic methods have begun to be used in fault diagnosis of mechanical equipment. Deep learning-based approaches can autonomously learn to extract fault characteristic information from extensive vibration data. The extracted insights facilitate precise fault classification, mitigating the influence of manual expertise on results and enhancing fault identification accuracy. A parallel deep learning-based model was put forward in [11] for running

status monitoring of drilling pumps. Specifically, an attention-enabled AlexNet and a signal transformer were combined for time-frequency feature exploration of the input signals. The on-site pump data was utilized for the experiment to validate the effectiveness of this approach. Xiao et al. [12] developed a new Bayesian signal transformer for machine health status recognition. First, a universal Bayesian model was constructed; Then, a new self-attention block was used to obtain the attention weights of the prior and posterior distributions; Finally, a newly designed optimization scheme was utilized to guide the model learning process. A new lightweight transformer was devised for machine condition monitoring [13]. First, a separable multilevel convolution module was employed to capture the global fault information; Then, a broadcast attention module was used for exploring multireceptive features. Two datasets of mechanical systems were used for the experimental validation. Xu et al. [14] put forward a multiscale threshold denoising CNN model for health monitoring of rotating machinery. First, a multilevel feature extraction module and a threshold denoising module were introduced for efficient feature exploration. Afterward, a novel knowledge distillation scheme was applied for critical information location. A cross-fusion CNN model was devised for fault diagnostics and prognostics of machines [15]. First, a new CNN backbone was designed for abundant feature learning. Then, a central fusion block was established for feature interaction learning. Finally, a novel learning scheme was devised to guide the learning process of the model. Xu et al. [16] introduced a novel methodology that integrates a CNN model with a multi-receptive field module and residual connections for intelligent feature extraction from vibration signals. To effectively capture multiscale features, they introduced a multireceptive field denoising block that dynamically learns from the vibration signals. Additionally, an efficient information fusion module was integrated to intelligently combine the extracted feature information, thereby bolstering the diagnostic efficacy of the model.

As modern industry evolves and sensing technology advances, acquiring sensor data for fault diagnosis has become increasingly accessible, unavoidably adding complexity to the diagnostic process. Data-driven diagnostic models rely on thorough analysis of sensory data. Enhanced accuracy in fault diagnosis is achieved with ample training data samples encompassing diverse fault types. However, real-world scenarios often entail long periods of normal equipment operation with relatively few fault occurrences. Despite continuous data collection by sensors during system operation, the majority of collected data pertains to the system's healthy state, while fault data may be sparse or absent for certain fault categories. Moreover, the complexity of working conditions yields a plethora of data with unknown fault categories. Although fault simulation test beds in laboratories or simulation software can be employed to mimic system faults, challenges arise due to the inability to replicate complex working conditions accurately, leading to significant disparities between simulated and actual data. Consequently, the imbalanced distribution of samples presents substantial challenges in training data-driven diagnostic models, necessitating urgent attention from

academia and industry. Utilizing unbalanced data directly for fault diagnosis often results in diminished diagnostic accuracy and poor generalization performance, thereby failing to meet the requirements of effective fault diagnosis.

In recent years, a plethora of fault diagnosis methods have emerged to tackle the challenge of class imbalance. A new imbalanced diagnostic algorithm was presented by Wang et al. [17] for machine health state monitoring. First, Wavelet transform was employed to transform the input signals into images; Afterwards, a modified CNN architecture was used for feature exploration; Finally, a new attention block and a new loss function were applied for fault classification. The experimental cases validated the approach was effective. A cost-sensitive transformer-based architecture was devised for long-tail fault recognition of machines [18]. First, the classic CNN model was utilized for the construction of token embedding. Afterward, an adaptive loss was introduced to redistribute weights. Extensive case studies were implemented to validate the method's efficacy. A multi-attention-empowered CNN model with a new loss function was devised for imbalanced fault classification [19]. First, a multiscale block and a feature aggregation block were introduced for feature exploration; Afterwards, a dual focal loss function was designed to help the model to focus on critical information. Extensive approaches showed that this approach was useful. Wu et al. [20] built an information-imbalance framework for health monitoring under imbalance cases. First, data-level and feature-level information was considered for useful feature exploration. Then, a Bayes-optimal classifier was constructed for imbalance fault identification. An intelligent augmentation algorithm for bearing fault prognostics was proposed by Li et al. [21]. First of all, a γ based thresholding algorithm was applied to set the centers of the cluster; Afterwards, an enhanced KNN was applied to classify the samples effectively. Finally, seven cutting-edge methods were introduced to compare with this approach on two industrial datasets. A wavelet similarity fusion framework was devised for fault identification of hydraulic pumps under imbalance cases [22]. First, a Wavelet packet decomposition algorithm was utilized to decompose the input signals into frequency maps; Then, a similarity fusion scheme was used to fuse important features; Finally, the features were augmented using the developed framework. Li et al. [23] put forward a reinforcement learning-based architecture for condition monitoring of bearings under data imbalance scenarios. First, the SMOTE technique was applied to augment the industrial dataset; Afterwards, a densely connected network combined with mixed attention blocks was used for crucial feature extraction; Finally, a parallel learning scheme was deployed to carry out the diagnostic task.

The above-mentioned algorithms can be summarized as data augmentation-based algorithms and cost-sensitive loss-based algorithms. Although these techniques improve diagnostic accuracy in imbalanced situations to a certain extent, they still retain certain inadequacies: 1) Some methods perform poorly when dealing with highly imbalanced datasets, making it difficult to fully extract feature information from minority class samples; 2) The data augmentation method may introduce noise or cause excessive distortion of samples, affecting

the model's generalization ability; 3) Cost-sensitive methods typically require manual setting of class weights, which may be difficult to determine, and adjustments for different classes may affect model performance. In our perspective, the crucial aspect for successfully tackling data imbalance rests in the model's strong feature extraction capability and the implementation of effective sensitivity learning strategies tailored for imbalanced data. The model's strong feature learning capability can assist in extracting ample and comprehensive feature information from imbalanced data, while effective sensitivity learning strategies can guide the model to focus more on valuable features within small-sample data. To tackle the challenges mentioned earlier and attain favorable diagnostic outcomes in the presence of data imbalance, we put forward an attention-based multireceptive field convolutional network with an adaptive label regulation loss. To begin with, we introduce a global contextual attention module (GCAM) to exploit discriminative features. GCAM owns a parallel structure, which can achieve a balance between feature extraction and discriminate feature selection during the back-propagation process. Second, we establish a hierarchical receptive field module (HRFM) to improve the receptive field of the convolutional layer, thus incorporating powerful multi-level learning capabilities into the AMFCN model. Finally, an adaptive label regulation loss is devised to facilitate precise fault recognition of the monitored machine.

To wrap up, the key innovations of this paper are outlined as follows:

- (1) A global contextual attention module (GCAM) and a mix convolutional (MIXConv) layer are utilized to instruct the AMFCN to concentrate on extracting ample features.
- (2) A hierarchical receptive field module (HRFM) is established to incorporate powerful multi-level learning capabilities into the AMFCN model.
- (3) An adaptive label regulation loss (ALRL) is devised to facilitate precise fault recognition, particularly in situations with imbalanced data distributions.

The arrangement of this remaining content is outlined below: Section II delineates the AMFCN model, while Sections III and IV carry out experiments to verify the effectiveness of the AMFCN on two gearbox datasets. Afterward, ablation studies are implemented to verify the improvements of this paper. Lastly, Section VI furnishes a thorough overview of the study's contributions.

II. PROPOSED ALGORITHM

Fig. 1 displays the AMFCN model, featuring several convolutional layer, multiple mix convolutional (MIXConv) layers, a few global contextual attention module (GCAM), several hierarchical receptive field module (HRFM), and a classification module. In alignment with Ref [24], we employ MIXConv to augment the AMFCN's feature extraction prowess without inflating the parameter count. The utilization of GCAM assists the model in prioritizing discriminative features, thereby amplifying its overall performance. Due to its parallel structure, these two modules not only independently optimize themselves but also collaborate to achieve a balance between feature

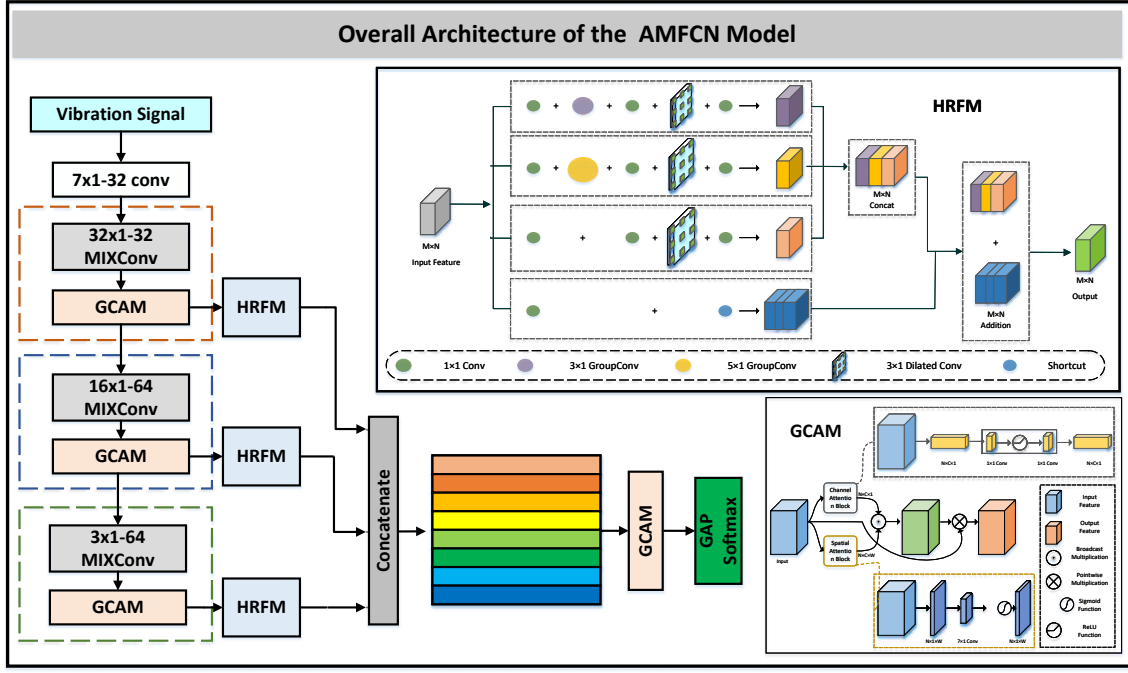


Fig. 1. The attention-based multireceptive field convolutional neural network.

extraction and discriminate feature selection during the back-propagation process. Subsequently, the side-out features are routed into the HRFM for further processing. The HRFM combines convolution kernels with different dilation rates in a parallel manner to increase the receptive field of the convolutional layer, thus injecting powerful multi-level learning capabilities into the AMFCN model. Afterward, the unique traits from various branches are efficiently concatenated. Eventually, an adaptive label regulation loss is devised to facilitate precise fault identification of the monitored machine.

A. Hierarchical Receptive Field Module

This section introduces the hierarchical receptive field module (HRFM) to augment the diversity of local receptive fields. The HRFM combines convolution kernels with different dilation rates in a parallel manner to increase the receptive field of the convolutional layer, thus injecting powerful multi-level learning capabilities into the AMFCN model. To be specific, the HRFM first deploys a 1×1 convolution to integrate features across channels. Afterwards, features collected from each layer of the encoder are then merged, ensuring that both original feature data and enhanced contextual details are preserved. Within the HRFM, individual branches process their groups independently through distinct convolutions, with receptive sizes being hierarchically adjusted. Based on the set of operations, the MRFM can extract rich multiscale features.

As illustrated in Fig. 1, the HRFM comprises three branches. Initially, a 1×1 convolution is utilized to the input feature for each branch. Subsequently, group convolutions with varying kernels and dilated convolutions with different dilation rates are executed to extract additional features. Then, feature maps from the different branches are concatenated along the

channel dimension. Finally, a shortcut connection scheme is employed to retain the original feature information.

B. Global Contextual Attention Module

As depicted in Fig. 1, we employ GCAM to aid the model in prioritizing discriminative features, consequently boosting its overall performance. The GCAM consists of two branches, namely the channel attention branch (CAB) and the spatial attention branch (SAB). The average channel attention branch is utilized to capture global information across the feature map and extract representative features, while the maxpool spatial attention branch further considers spatial information of the feature map to accurately capture important features. Due to its parallel structure, these two modules not only independently optimize themselves but also collaborate to achieve a balance between feature extraction and discriminate feature selection during the back-propagation process.

To commence, the input feature I is initially supplied. The avgpool operation is first utilized for discovering global channel information. Afterward, a multilayer perceptron (MLP), which consists of two 1×1 convolutions, was applied to further explore channel features. To amplify the nonlinearity within the MLP, PReLU is utilized to activate the output of MLP. Moreover, to mitigate the parameter count and computational complexity of the channel attention branch, a bottleneck architecture is adopted in the architecture. This architecture entails that the number of input channels for the initial convolutional layer is r times the number of output channels. Ultimately, the channel attention weights are derived through a sigmoid activation function. The generation process of the channel branch can be delineated as:

$$CAB(I) = \text{Sigmoid}[F_2^k(\text{PReLU}(F_1^k(\text{Avgpool}(I))))] \quad (1)$$

where $F_i^k, i = 1, 2$ denotes the i_{th} $k \times 1$ convolution.

On the other hand, the spatial attention branch explores crucial spatial information to accurately capture discriminative features. To begin with, the maxpool operation is applied to produce coefficients for the horizontal segment of the input feature map. The max pooling layer's output undergoes convolution with a substantial receptive field sized at $k \times 1$, followed by activation through a sigmoid function to yield spatial attention weights. This method adeptly fine-tunes parameter count, as illustrated in the subsequent expression:

$$SAB(I) = \text{Sigmoid}[F_3^k(\text{Maxpool}(I))] \quad (2)$$

where $SAB(I)$ denotes the output of the spatial attention branch.

Further, the global contextual attention weight is calculated via the point-wise multiplication of the channel attention weight and the global attention weight. At last, the global attention feature map is obtained via the broadcast multiplication of the global contextual attention weight and the input feature. The generation process of the global attention feature map can be delineated as:

$$GCAM(I) = SAB(I) \otimes CAB(I) \odot I \quad (3)$$

where $GCAM(I)$ represents the global contextual attention feature map.

C. Mix Convolutional Layer

Building upon the concepts introduced in [24], we incorporate the MIXConv operation to effectively explore abundant fault-related information, as displaced in Fig. 1. At the outset, the MIXConv utilizes depthwise convolution to explore spatial information. Moreover, it proficiently fosters cross-channel feature interaction via convolutional operations, ensuring the smooth combination of information across diverse channels. Following this, we introduce the skip connection strategy, which can avoid information loss. By appropriately increasing the network depth, MIXConv highlights important information extracted from the input signal. Moreover, it facilitates a more comprehensive integration of spatial information by employing larger convolution kernels.

D. Adaptive Label Regulation Loss

In many diagnostic tasks, regularization techniques seamlessly integrate calibration methods into the training process to empower deep learning-based models to deliver heightened diagnostic precision, even amidst data imbalance scenarios. Network calibration regularization techniques commonly incorporate an additional term alongside the cross-entropy loss function, aiming to alleviate issues of miscalibration, as given as follows:

$$L = L_{CE} + L_R \quad (4)$$

In this context, L_R stands for the regularization term. Then, the gradient of Eq. (4) can be obtained:

$$\frac{\partial L}{\partial z_i} = p_i - q_i + \frac{\partial L_R}{\partial z_i} \quad (5)$$

This can be restated as follows:

$$\frac{\partial L}{\partial z_i} = \begin{cases} p_i - (q_i - f(z_i)G(z_i)), & i = \hat{y}, \\ p_i - (q_i + f(z_i)G(z_i)), & i \neq \hat{y}, \end{cases} \quad (6)$$

In this context, p signifies the output generated by the model and q refers to the real target distribution; f stands for a smoothing function that governs the level of smoothing, while G acts as an indicator determining the application of smoothing. As in [25], the smoothing function is defined as:

$$f(z_i) = \begin{cases} \lambda_1(z_i - \min_k z_k - M), & i = \hat{y}, \\ \lambda_2(z_{\hat{y}} - z_i - M), & i \neq \hat{y}, \end{cases} \quad (7)$$

in which, λ_1 and λ_2 serve as hyperparameters to balance the two terms; M stands for a margin. The piecewise linearity within the smoothing function guarantees a continual decrease in the target label as the probability rises. In addition, the assessment indicator is designed as:

$$G(z_i) = \begin{cases} \mathbb{1}[z_i - \min_k z_k \geq M], & i = \hat{y}, \\ \mathbb{1}[z_{\hat{y}} - z_i \geq M], & i \neq \hat{y}, \end{cases} \quad (8)$$

in which, $\mathbb{1}[\cdot]$ represents a function that returns a value of 1 if the condition holds true, and 0 otherwise. By integrating the smoothing and evaluation functions described in Eq. (7) and Eq. (8) into the terms provided in Eq. (6), we can obtain:

$$\frac{\partial L_{ALRL}}{\partial z_i} = \begin{cases} \lambda_1 \text{ReLU}(Z_i - \min_k z_k - M), & i = \hat{y}, \\ \text{ReLU}(Z_{\hat{y}} - Z_i - M), & i \neq \hat{y}, \end{cases} \quad (9)$$

Through the integration, we derive the loss function, denoted as the adaptive label regulation loss (ALRL), as follows:

$$L_{ALRL} = \begin{cases} \lambda_1 (\text{ReLU}(Z_i - \min_k z_k - M))^2, & i = \hat{y}, \\ (\text{ReLU}(Z_{\hat{y}} - Z_i - M))^2, & i \neq \hat{y}, \end{cases} \quad (10)$$

We incorporate both fidelity and regularization terms in the model training process. The comprehensive loss function is presented as follows:

$$L = L_{CE} + L_{ALRL}. \quad (11)$$

The ALRL can help the AMFCN model to achieve heightened diagnostic precision, even in data imbalance scenarios.

E. AFMCN-based Framework for Gearbox Health Status Monitoring

We offer a detailed depiction of the fault diagnosis paradigm rooted in AFMCN. To encapsulate, the specific procedure can be delineated as:

- 1) **Data Preparation:** The process begins with the collection of mechanical signals using a signal acquisition system, followed by their segmentation into many samples.
- 2) **Model Training:** The AFMCN algorithm undergoes training using the provided training samples, with the process enhanced through the integration of adaptive label regulation loss.
- 3) **Model Validation:** Feed the testing data into the trained model to precisely evaluate the running state of the monitored gearbox.

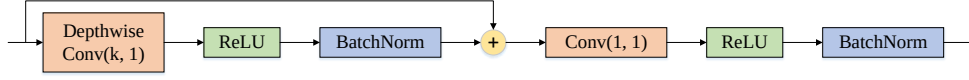


Fig. 2. Mix convolutional layer.

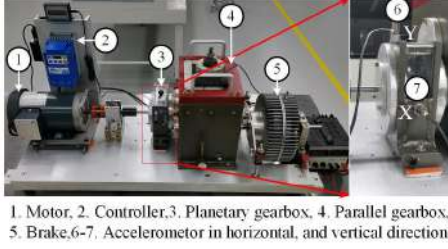


Fig. 3. The XJTU test rig designed for simulating gearbox faults [26].

TABLE I
COMPREHENSIVE OVERVIEW OF THE XJTU GEARBOX DATASET

Class	Detailed information	Data for training	Data for testing
0	Outer race failure	1000	1000
1	Ball failure	1000	1000
2	Missing teeth failure	1000	1000
3	Root cracks	1000	1000
4	Tooth surface wear	1000	1000
5	Normal status	1000	1000

III. CASE STUDY I: XJTU GEARBOX DATASET

A. Data Collection

Fig. 3 illustrates the experimental setup, comprising a driving motor, a controller, a planetary gearbox, a parallel gearbox, and a brake. One notable component is the motor, specifically a 3-phase, 3 HP motor, powered by a three-phase alternating current (230V, 60/50Hz). Two PCB352C04 accelerometers are strategically positioned on the X and Y axes of the planetary gearbox to capture vibration signals, with a focus on utilizing the signals from the Y-axis direction. The signals of six distinct condition type are collected for the experiment. In every health state, we painstakingly gather 2000 samples, splitting them into 1000 for training and 1000 for testing. Each of these samples consists of 2048 measured data points. Throughout the experiments, the motor maintains a speed of 1800 rotations per minute (rpm), while the sampling frequency is configured at 20480 Hertz (Hz). Detailed information regarding the XJTU gearbox dataset is outlined in Table I.

B. Deployment Configurations

All experiments are carried out on a laptop featuring a GTX 3060 Ti GPU. Seven cut-edging approaches are used to compare with the AMFCN model.

- (1) *MA1DCNN*: MA1DCNN, as introduced in [27], is a CNN model utilized for interpretable fault diagnosis, capable of operating effectively under varying speed conditions.

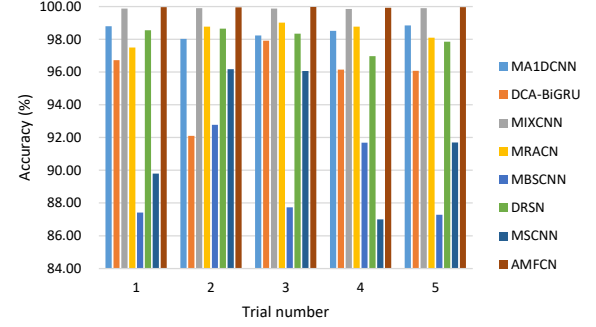


Fig. 4. Accuracy of diagnostics performed on the XJTU gearbox dataset.

- (2) *DCA-BiGRU*: DCA-BiGRU [28] is a recently devised approach that merges CNN and LSTM models to improve bearing fault diagnostics.
- (3) *MIXCNN*: MIXCNN, outlined in [29], is a newly devised lightweight algorithm tailored for fault classification.
- (4) *MRACN*: MRACN, detailed in the study by [30], is an innovative residual attention network crafted to excel in fault recognition amidst nonstationary conditions.
- (5) *MBSCNN*: MBSCNN, proposed in [31], is a CNN network with multiple branches designed to extract multiscale features from monitored signals.
- (6) *DRSN*: DRSN, proposed in [32], is a CNN algorithm highly regarded for its proficiency in machinery fault diagnosis
- (7) *MSCNN*: MSCNN [33] is a fault diagnosis model designed to handle gearbox issues amid fluctuating speed conditions.

C. Compared with State-of-the-Art Techniques

Fig. 4 and Table II display the average accuracy (Ave-Acc) and average F_1 score for different methods on the XJTU dataset. Additionally, the tables also show the standard deviations derived from 5 trials. As shown in Table II, the proposed AMFCN method demonstrated an impressive performance, achieving Ave-Acc of 99.96%. The result notably surpassed all comparative methods. Moreover, AMFCN stands out for its exceptional performance in terms of the average F_1 score. Specifically, AMFCN (99.96%) outperforms MA1DCNN (98.49%), DCA-BiGRU (95.80%), MIXCNN (99.88%), MRACN (98.44%), MBSCNN (88.70%), DRSN (98.08%), and MSCNN (92.06%) by margins of 1.47%, 4.16%, 0.08%, 1.52%, 11.26%, 1.88%, and 7.9%, respectively. Furthermore, the proposed method showed the least variability among all techniques in terms of standard deviation. These results confirm the efficacy and durability of the proposed method in recognizing the failure of the gearbox.

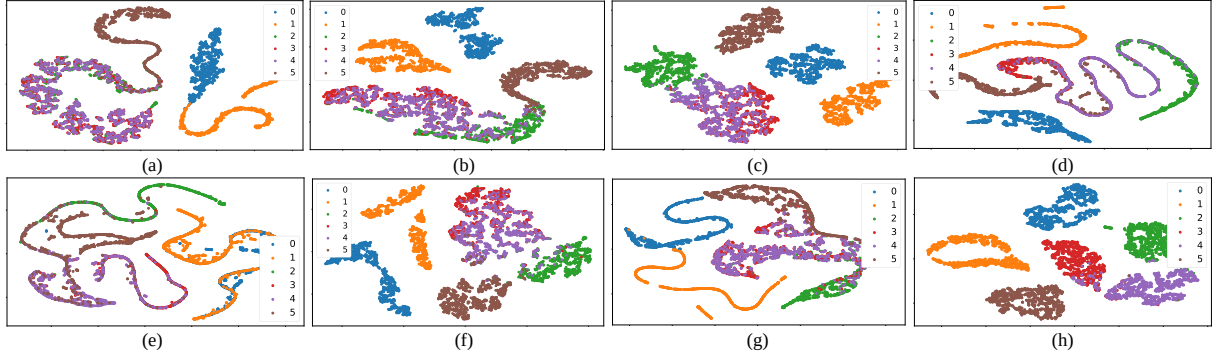


Fig. 5. Employing t-SNE for the Visualization of Features on XJTU dataset: (a) MA1DCNN; (b) DCA-BiGRU; (c) MIXCNN; (d) MRACN; (e) MBSCNN; (f) DRSN; (g) MSCNN; and (h) AMFCN.

TABLE II
OUTCOMES DERIVED FROM THE DIAGNOSTIC EXAMINATION OF THE XJTU GEARBOX DATASET.

Algorithms	Max-acc [%]	Min-acc [%]	Avg-acc [%]	Avg- F_1 [%]
MA1DCNN	98.85	98.03	98.49 ± 0.36	98.49 ± 0.35
DCA-BiGRU	97.92	92.10	95.80 ± 2.19	95.80 ± 2.20
MIXCNN	99.90	99.85	99.88 ± 0.02	99.88 ± 0.02
MRACN	99.02	97.50	98.43 ± 0.62	98.44 ± 0.62
MBSCNN	92.78	87.28	89.38 ± 2.64	88.70 ± 3.03
DRSN	98.65	96.97	98.07 ± 0.69	98.08 ± 0.69
MSCNN	96.17	87.00	92.15 ± 3.99	92.06 ± 4.02
AMFCN	99.98	99.93	99.96 ± 0.02	99.96 ± 0.02

TABLE III
DETAILED DESCRIPTION OF THE FOUR NEWLY XJTU DATASETS.

Class	Training data			
	Training set A	Training set B	Training set C	Training set D
0	100%	100%	100%	100%
1	10%	5%	2%	1%
2	10%	5%	2%	1%
3	10%	5%	2%	1%
4	10%	5%	2%	1%
5	10%	5%	2%	1%

D. Resilience Towards Imbalanced Dataset

To confirm the diagnostic effectiveness of our proposed method on imbalanced datasets, we created four imbalanced datasets derived from the XJTU dataset. The specifics of these datasets are outlined in Table III. It can be observed that dataset D is highly imbalanced, with the number of fault category instances comprising only one percent of the total instances in the normal category. The results, as shown in Table IV, indicate that all methods are highly sensitive to the imbalance ratio of the data. More precisely, as the imbalance ratio escalates, there is a marked decrease in the diagnostic efficacy achieved by these methods. However, the findings unmistakably demonstrate that AMFCN consistently outshines the other seven algorithms across all imbalanced conditions, consistently attaining a notably better result. In particular, AMFCN demonstrates a significant performance

TABLE IV
OUTCOME OF THE DIAGNOSTIC EVALUATION CONDUCTED ON THE FOUR XJTU GEARBOX DATASET.

Methods	Dataset A [%]	Dataset B [%]	Dataset C [%]	Dataset D [%]
MA1DCNN	73.98 ± 4.09	69.74 ± 4.17	62.72 ± 3.91	54.03 ± 2.84
DCA-BiGRU	83.00 ± 3.52	79.51 ± 2.08	62.94 ± 4.17	52.17 ± 4.24
MIXCNN	96.75 ± 0.87	91.20 ± 2.81	71.12 ± 3.15	52.76 ± 2.56
MRACN	71.42 ± 7.86	75.93 ± 10.50	65.53 ± 9.43	68.25 ± 4.72
MBSCNN	71.57 ± 4.81	62.46 ± 2.38	54.65 ± 8.26	38.55 ± 6.89
DRSN	92.24 ± 0.74	86.77 ± 1.36	72.85 ± 2.62	57.19 ± 2.85
MSCNN	72.87 ± 4.91	59.71 ± 5.30	52.72 ± 5.75	40.76 ± 8.27
AMFCN	98.40 ± 0.35	94.95 ± 0.48	88.24 ± 0.56	83.72 ± 0.89

advantage over the other seven algorithms on dataset D. To be more specific, AMFCN achieves a notable performance superiority with an accuracy of 83.72%, respectively surpassing MA1DCNN (54.03%), DCA-BiGRU (52.17%), MIXCNN (52.76%), MRACN (68.25%), MBSCNN (38.55%), DRSN (57.19%), and MSCNN (40.76%) by margins of 29.69%, 31.55%, 30.96%, 15.47%, 45.17%, 26.53%, and 42.96% on dataset D. Furthermore, AMFCN's robustness is evidenced by its standard deviation consistently registering lower values compared to the other seven state-of-the-art algorithms across all imbalanced datasets.

For a deeper insight into feature distribution, we apply the t-SNE technique [34] to visualize the features learned from Dataset B. As depicted in Fig. 5, AMFCN demonstrates exceptional capability in discerning class differences in comparison with other algorithms. Particularly noteworthy are the feature maps produced by AMFCN, showcasing clear clustering of similar categories while minimizing overlaps across different classes.

IV. CASE STUDY II: CUMT GEARBOX DATASET

A. Data Description

Fig. 6 showcases the CUMT gearbox test bench utilized for experimental validation. This setup comprises components such as an induction motor and a control cabinet. In the signal acquisition stage, the rotational speed was controlled from 0 to 1200 rpm, while the sampling frequency is configured at 20480 Hertz (Hz). The signals of eight distinct condition types

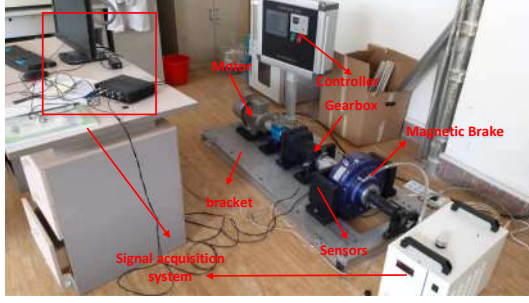


Fig. 6. The CUMT test rig is designed for simulating gearbox faults.

TABLE V
COMPREHENSIVE OVERVIEW OF THE CUMT GEARBOX DATASET

Category	Detailed description
0	Normal
1	Root crack and teeth wear
2	broken teeth
3	Cage failure
4	pitting defect
5	ball defect
6	Inner race failure
7	Gear wear and pitting

TABLE VI
OUTCOME OF THE DIAGNOSTIC EVALUATION CONDUCTED ON THE CUMT GEARBOX DATASET.

Methods	Max-acc [%]	Min-acc [%]	Avg-acc [%]	Avg- F_1 [%]
MA1DCNN	99.20	91.98	96.86 ± 2.93	96.87 ± 2.91
DCA-BiGRU	94.45	83.30	88.95 ± 4.48	89.05 ± 4.38
MIXCNN	98.48	97.33	98.15 ± 0.47	98.16 ± 0.47
MRACN	97.85	95.85	96.74 ± 0.83	96.75 ± 0.83
MBSCNN	96.95	92.43	95.71 ± 1.88	95.72 ± 1.88
DRSN	98.77	98.12	98.40 ± 0.29	98.41 ± 0.29
MSCNN	88.15	74.67	80.75 ± 4.98	80.54 ± 5.05
AMFCN	99.40	99.13	99.27 ± 0.13	99.27 ± 0.13

are collected for the experiment. In every health state, we painstakingly gather 1000 samples, splitting them into 500 for training and 500 for testing. Each of these samples consists of 2,048 measured data points. Detailed information regarding the CUMT gearbox dataset is outlined in Table V.

B. Compared with State-of-the-Art Techniques

Fig. 7 and Table VI display the average accuracy (Ave-Acc) and average F_1 score for different methods on the CUMT dataset. Additionally, the tables also show the standard deviations derived from 5 trials. As shown in Table VI, the proposed AMFCN method demonstrated an impressive performance, achieving Ave-Acc of 99.27%. The result notably surpassed all comparative methods. Moreover, AMFCN stands out for its exceptional performance in terms of the average F_1 score. Specifically, AMFCN (99.27%) outperforms MA1DCNN (96.87%), DCA-BiGRU (89.05%),

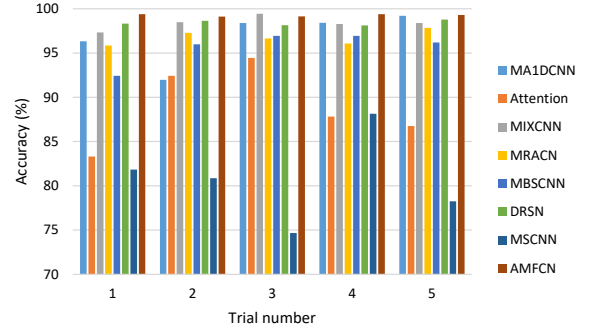


Fig. 7. Accuracy of diagnostics performed on the CUMT gearbox dataset.

TABLE VII
DETAILED DESCRIPTION OF THE FIVE NEWLY CUMT DATASETS.

Class	Training data				
	Training set I	Training set II	Training set III	Training set IV	Training set V
0	100%	100%	100%	100%	100%
1	40%	30%	20%	10%	6%
2	40%	30%	20%	10%	6%
3	40%	30%	20%	10%	6%
4	40%	30%	20%	10%	6%
5	40%	30%	20%	10%	6%
6	40%	30%	20%	10%	6%
7	40%	30%	20%	10%	6%

MIXCNN (98.16%), MRACN (96.75%), MBSCNN (95.72%), DRSN (98.41%), and MSCNN (80.54%) by margins of 2.4%, 10.22%, 1.11%, 2.52%, 3.55%, 0.86%, and 18.73%, respectively. Furthermore, the proposed method showed the least variability among all techniques in terms of standard deviation. These results confirm the efficacy and durability of the proposed method in recognizing the failure of the gearbox.

C. Resilience Towards Imbalanced Dataset

To confirm the diagnostic effectiveness of our proposed method on imbalanced datasets, we created five imbalanced datasets derived from the CUMT dataset. The specifics of these datasets are outlined in Table VII. It can be easily observed that dataset V is highly imbalanced, with the number of fault category instances comprising only one percent of the total instances in the normal category. The results, as shown in Table VIII, indicate that all methods are highly sensitive to the imbalance ratio of the data. More precisely, as the imbalance ratio escalates, there is a marked decrease in the diagnostic efficacy achieved by these methods. However, the findings unmistakably demonstrate that AMFCN consistently outshines the other seven algorithms across all imbalanced conditions, consistently attaining a notably better result. In particular, AMFCN demonstrates a significant performance advantage over the other seven algorithms on dataset V. To be more specific, AMFCN achieves a notable performance superiority with an accuracy of 81.63%, respectively surpassing MA1DCNN (61.69%), DCA-BiGRU (38.44%), MIXCNN (51.31%), MRACN (42.57%), MBSCNN (49.50%), DRSN (56.35%), and MSCNN (32.07%) by margins of 19.94%, 43.19%, 30.32%, 39.06%, 32.13%, 25.28%, and 49.56% on

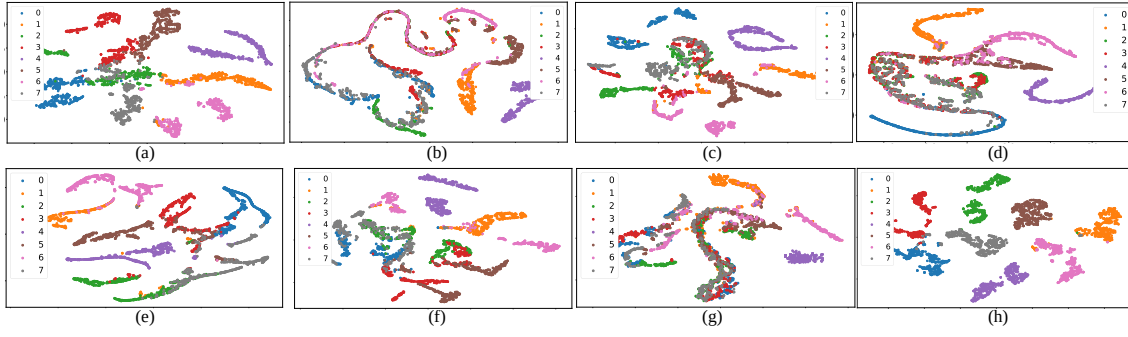


Fig. 8. Employing t-SNE for the Visualization of Features on CUMT dataset: (a) MA1DCNN; (b) DCA-BiGRU; (c) MIXCNN; (d) MRACN; (e) MBSCNN; (f) DRSN; (g) MSCNN; and (h) AMFCN.

TABLE VIII
OUTCOME OF THE DIAGNOSTIC EVALUATION CONDUCTED ON THE FIVE CUMT GEARBOX DATASET.

Algorithms	Dataset I [%]	Dataset II [%]	Dataset III [%]	Dataset IV [%]	Dataset V [%]
MA1DCNN	97.03 $\pm$ 0.40	92.76 $\pm$ 0.99	89.76 $\pm$ 1.34	79.68 $\pm$ 1.95	61.69 $\pm$ 5.08
DCA-BiGRU	81.04 $\pm$ 3.16	59.02 $\pm$ 5.35	47.55 $\pm$ 7.17	41.30 $\pm$ 6.17	38.44 $\pm$ 5.45
MIXCNN	96.51 $\pm$ 0.85	93.17 $\pm$ 1.44	91.30 $\pm$ 1.34	66.90 $\pm$ 0.89	51.31 $\pm$ 3.06
MRACN	87.90 $\pm$ 4.78	71.47 $\pm$ 6.04	79.04 $\pm$ 9.61	69.38 $\pm$ 8.57	42.57 $\pm$ 4.31
MBSCNN	89.36 $\pm$ 1.66	88.43 $\pm$ 1.52	83.32 $\pm$ 3.38	70.60 $\pm$ 4.86	49.50 $\pm$ 5.79
DRSN	91.98 $\pm$ 1.33	89.30 $\pm$ 0.39	83.56 $\pm$ 1.67	64.68 $\pm$ 3.82	56.35 $\pm$ 3.45
MSCNN	68.77 $\pm$ 3.12	56.33 $\pm$ 9.10	54.83 $\pm$ 6.38	48.37 $\pm$ 4.76	32.07 $\pm$ 3.84
AMFCN	99.23 $\pm$ 0.17	99.00 $\pm$ 0.16	97.87 $\pm$ 0.29	94.21 $\pm$ 0.75	81.63 $\pm$ 1.63

TABLE IX
RESULTS FROM THE DIAGNOSTIC ANALYSIS OF THE FOUR ALGORITHMS.

Methods	Dataset C [%]	Dataset D [%]	Dataset IV [%]	Dataset V [%]
AMFCN	88.24 $\pm$ 0.56	83.72 $\pm$ 0.89	94.21 $\pm$ 0.75	81.63 $\pm$ 1.63
AMFCN-nALRL	85.33 $\pm$ 2.41	79.50 $\pm$ 1.98	90.47 $\pm$ 3.06	77.40 $\pm$ 2.91
AMFCN-nGCAM	86.08 $\pm$ 1.26	80.32 $\pm$ 1.69	90.63 $\pm$ 0.74	77.10 $\pm$ 1.73
AMFCN-nHRFM	84.46 $\pm$ 1.91	79.79 $\pm$ 3.01	88.83 $\pm$ 2.67	74.55 $\pm$ 1.74

dataset V. Furthermore, AMFCN's robustness is evidenced by its standard deviation consistently registering lower values compared to the other seven state-of-the-art algorithms across all imbalanced datasets.

For a deeper insight into feature distribution, we apply the t-SNE technique [34] to visualize the features learned from Dataset IV. As depicted in Fig. 8, AMFCN demonstrates exceptional capability in discerning class differences in comparison with other algorithms. Particularly noteworthy are the feature maps produced by AMFCN, showcasing clear clustering of similar categories while minimizing overlaps across different categories.

V. ABLATION STUDIES

A. Efficacy of the ALRL

Within this segment, we delve into experiments aimed at gauging the effectiveness of the ALRL. To provide a baseline for comparison, we introduce a model named AMFCN-nALRL, which mirrors the architecture of AMFCN but excludes ALRL. The core difference between AMFCN and

AMFCN-nALRL is that the latter uses cross-entropy loss instead of ALRL in the training stage. Summarizing the experimental findings in Table IX, we observe that AMFCN outperforms AMFCN-nALRL in terms of average accuracy, showcasing improvements of 2.91%, 4.22%, 3.74% and 4.23% across four distinct diagnostic scenarios. This enhancement is ascribed to the ALRL's adeptness in Facilitating the AMFCN to achieve accurate fault identification results, particularly in situations with imbalanced data distributions.

B. Efficacy of the GCAM

Within this segment, we delve into experiments aimed at gauging the effectiveness of the GCAM. To provide a baseline for comparison, we introduce a model named AMFCN-nGCAM, which mirrors the architecture of AMFCN but excludes GCAM. The core disparity between AMFCN and AMFCN-nGCAM lies in the latter's omission of GCAMs. Summarizing the experimental findings in Table IX, we observe that AMFCN outperforms AMFCN-nGCAM in terms of average accuracy, showcasing improvements of 2.16%, 3.4%, 3.58% and 4.53% across four distinct diagnostic scenarios. This enhancement is ascribed to the GCAM's adeptness in harnessing ample features, thereby augmenting the model's capacity to glean robust feature representations.

C. Efficacy of the HRFM

Within this segment, we delve into experiments aimed at gauging the effectiveness of the HRFM. To provide a baseline for comparison, we introduce a model named AMFCN-nHRFM, which mirrors the architecture of AMFCN but excludes HRFM. The core disparity between AMFCN and AMFCN-nHRFM lies in the latter's omission of HRFMs. Summarizing the experimental findings in Table IX, we observe that AMFCN outperforms AMFCN-nHRFM in terms of average accuracy, showcasing improvements of 3.78%, 3.93%, 5.38% and 7.08% across four distinct diagnostic scenarios. This enhancement is ascribed to the HRFM's adeptness in capturing multilevel discriminate features.

D. Feature Visualization

In this section, we present a comprehensive analysis of the enhancements for interpretability. Specifically, we apply the t-SNE technique to visualize the features learned from Dataset

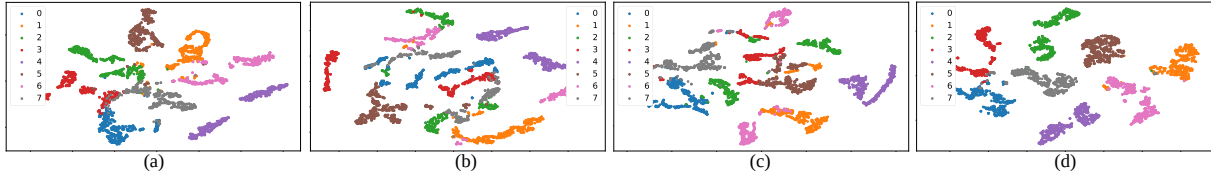


Fig. 9. Employing t-SNE for the Visualization of Features on CUMT dataset: (a) AMFCN-nALRL; (b) AMFCN-nGCAM; (c) AMFCN-nHRFM; (d) AMFCN.

IV. As depicted in Fig. 8, AMFCN demonstrates exceptional capability in discerning class differences in comparison with the other three algorithms. Particularly noteworthy are the feature maps produced by AMFCN, showcasing clear clustering of similar categories while minimizing overlaps across different categories.

VI. CONCLUSION

Accurate gearbox fault identification applied to imbalanced data scenarios is crucial for industrial production. From our perspective, the key to effectively addressing data imbalance lies in the model's robust feature extraction capabilities and the integration of sensitive learning strategies for unbalanced datasets. To attain favorable diagnostic outcomes in the presence of data imbalance, this study introduces an attention-based multireceptive field convolutional network with an adaptive label regulation loss. Firstly, we introduce a global contextual attention module (GCAM) designed to leverage discriminative features. GCAM adopts a parallel structure, enabling a harmonious interplay between feature extraction and selective feature discrimination during the back-propagation process. Secondly, we introduce a hierarchical receptive field module (HRFM) aimed at enhancing the receptive field of the convolutional layer, thereby integrating robust multi-level learning capabilities into the AMFCN model. Lastly, we devise an adaptive label regulation loss (ALRL) to refine predictions and ensure accurate classification of the operational states of the target machine. The experimental results indisputably indicate that AMFCN outperforms seven other cutting-edge methods, particularly exhibiting superiority in addressing imbalanced scenarios. Future work will focus on the implementation of the AMFCN in real industrial scenarios.

REFERENCES

- [1] W. Li, H. Lan, J. Chen, K. Feng, and R. Huang, "Wavcapsnet: An interpretable intelligent compound fault diagnosis method by backward tracking," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.
- [2] Y. Xu, K. Feng, X. Yan, X. Sheng, B. Sun, Z. Liu, and R. Yan, "Cross-modal fusion convolutional neural networks with online soft label training strategy for mechanical fault diagnosis," *IEEE Transactions on Industrial Informatics*, 2023.
- [3] Y. Xu, X. Yan, K. Feng, Y. Zhang, X. Zhao, B. Sun, and Z. Liu, "Global contextual multiscale fusion networks for machine health state identification under noisy and imbalanced conditions," *Reliability Engineering & System Safety*, vol. 231, p. 108972, 2023.
- [4] Y. Zhang, J. Ding, Y. Li, Z. Ren, and K. Feng, "Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating conditions," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108236, 2024.
- [5] R. Huang, J. Li, Y. Liao, J. Chen, Z. Wang, and W. Li, "Deep adversarial capsule network for compound fault diagnosis of machinery toward multidomain generalization task," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2020.
- [6] Z. Wan, Y. Xiaonan, and J. Minping, "Sparse enhancement based on the total variational denoising for fault feature extraction of rolling element bearings," *Measurement*, vol. 195, p. 111163, 2022.
- [7] G. Yu, "A concentrated time-frequency analysis tool for bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, DOI 10.1109/TIM.2019.2901514, no. 2, pp. 371–381, 2020.
- [8] X. Zhang, H. Luo, K. Li, and O. Kaynak, "Time-domain frequency estimation with application to fault diagnosis of the unmanned aerial vehiclesâ blade damage," *IEEE Transactions on Industrial Electronics*, vol. 69, DOI 10.1109/TIE.2021.3084177, no. 5, pp. 5257–5266, 2022.
- [9] R.-P. Nikula, K. Karioja, M. Pylv  n  inen, and K. Leivisk  , "Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features," *Mechanical Systems and Signal Processing*, vol. 138, p. 106572, 2020.
- [10] L. Cui, J. Chen, D. Liu, and D. Zhen, "Fault diagnosis of offshore wind turbines based on component separable synchroextracting transform," *Ocean Engineering*, vol. 291, p. 116275, 2024.
- [11] J. Guo, Y. Yang, H. Li, L. Dai, and B. Huang, "A parallel deep neural network for intelligent fault diagnosis of drilling pumps," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108071, 2024.
- [12] Y. Xiao, H. Shao, J. Wang, S. Yan, and B. Liu, "Bayesian variational transformer: A generalizable model for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 207, p. 110936, 2024.
- [13] S. Yan, H. Shao, J. Wang, X. Zheng, and B. Liu, "Liconvformer: A lightweight fault diagnosis framework using separable multiscale convolution and broadcast self-attention," *Expert Systems with Applications*, vol. 237, p. 121338, 2024.
- [14] Y. Xu, X. Yan, B. Sun, K. Feng, L. Kou, Y. Chen, Y. Li, H. Chen, E. Tian, Q. Ni *et al.*, "Online knowledge distillation based multiscale threshold denoising networks for fault diagnosis of transmission systems," *IEEE Transactions on Transportation Electrification*, 2023.
- [15] Y. Xu, K. Feng, X. Yan, R. Yan, Q. Ni, B. Sun, Z. Lei, Y. Zhang, and Z. Liu, "Cfcnn: A novel convolutional fusion framework for collaborative fault identification of rotating machinery," *Information Fusion*, vol. 95, pp. 1–16, 2023.
- [16] Y. Xu, X. Yan, B. Sun, J. Zhai, and Z. Liu, "Multireceptive field denoising residual convolutional networks for fault diagnosis," *IEEE Transactions on Industrial Electronics*, DOI 10.1109/TIE.2021.3125666, pp. 1–1, 2021.
- [17] S. Wang, J. Tian, P. Liang, X. Xu, Z. Yu, S. Liu, and D. Zhang, "Single and simultaneous fault diagnosis of gearbox via wavelet transform and improved deep residual network under imbalanced data," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108146, 2024.
- [18] S. Chang, L. Wang, M. Shi, J. Zhang, L. Yang, and L. Cui, "Extended attention signal transformer with adaptive class imbalance loss for long-tailed intelligent fault diagnosis of rotating machinery," *Advanced Engineering Informatics*, vol. 60, p. 102436, 2024.
- [19] Y. Xu, S. Li, X. Yan, J. He, Q. Ni, Y. Sun, and Y. Wang, "Multiattention-based feature aggregation convolutional networks with dual focal loss for fault diagnosis of rotating machinery under data imbalance conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, DOI 10.1109/TIM.2023.3346532, pp. 1–11, 2024.
- [20] J. Wu, H. Shang, C. Sun, R. Yan, and X. Chen, "Information-imbalance learning with hazard-sensitive loss for machine fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, DOI 10.1109/TIM.2024.3369146, pp. 1–1, 2024.
- [21] S. Li, Y. Peng, Y. Shen, S. Zhao, H. Shao, G. Bin, Y. Guo, X. Yang, and C. Fan, "Rolling bearing fault diagnosis under data imbalance and

variable speed based on adaptive clustering weighted oversampling,” *Reliability Engineering & System Safety*, vol. 244, p. 109938, 2024.

- [22] S. Fu, L. Lin, Y. Wang, M. Zhao, F. Guo, S. Zhong, and Y. Liu, “High imbalance fault diagnosis of aviation hydraulic pump based on data augmentation via local wavelet similarity fusion,” *Mechanical Systems and Signal Processing*, vol. 209, p. 111115, 2024.
- [23] Y. Li, Y. Wang, X. Zhao, and Z. Chen, “A deep reinforcement learning-based intelligent fault diagnosis framework for rolling bearings under imbalanced datasets,” *Control Engineering Practice*, vol. 145, p. 105845, 2024.
- [24] Z. Zhao and Y. Jiao, “A fault diagnosis method for rotating machinery based on cnn with mixed information,” *IEEE Transactions on Industrial Informatics*, DOI 10.1109/TII.2022.3224979, pp. 1–11, 2022.
- [25] H. Park, J. Noh, Y. Oh, D. Baek, and B. Ham, “Acls: Adaptive and conditional label smoothing for network calibration,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3936–3945, 2023.
- [26] T. Li, Z. Zhou, S. Li, C. Sun, R. Yan, and X. Chen, “The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study,” *Mechanical Systems and Signal Processing*, vol. 168, p. 108653, 2022.
- [27] H. Wang, Z. Liu, D. Peng, and Y. Qin, “Understanding and learning discriminant features based on multiattention 1dcnn for wheelset bearing fault diagnosis,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5735–5745, 2020.
- [28] X. Zhang, C. He, Y. Lu, B. Chen, L. Zhu, and L. Zhang, “Fault diagnosis for small samples based on attention mechanism,” *Measurement*, vol. 187, p. 110242, 2022.
- [29] X. Dong, H. Gao, L. Guo, K. Li, and A. Duan, “Deep cost adaptive convolutional network: A classification method for imbalanced mechanical data,” *IEEE Access*, vol. 8, pp. 71 486–71 496, 2020.
- [30] L. Jia, T. W. Chow, Y. Wang, and Y. Yuan, “Multiscale residual attention convolutional neural network for bearing fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.
- [31] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo, and J. Chen, “Multibranch and multiscale cnn for fault diagnosis of wheelset bearings under strong noise and variable load condition,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4949–4960, 2020.
- [32] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, “Deep residual shrinkage networks for fault diagnosis,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4681–4690, 2020.
- [33] G. Jiang, H. He, J. Yan, and P. Xie, “Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox,” *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196–3207, 2018.
- [34] L. Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, no. 2605, pp. 2579–2605, 2008.



Yadong Xu received a PhD degree in Mechanical Engineering from Southeast University, Nanjing, China, in 2023. He is currently a Postdoctoral Researcher at The Hong Kong Polytechnic University, Hong Kong, China. His current research interests include intelligent fault diagnosis, digital image processing, and pattern recognition.



Sheng Li received a PhD degree in Industrial Engineering from Hohai University, Nanjing, China in 2024. He was a visiting scholar at the University of British Columbia from 2022 to 2023. He is currently an Associate Professor at the College of Civil Engineering, Nanjing Forestry University, Nanjing, China. Dr. Li's current research interests include intelligent fault diagnosis, digital image processing, and pattern recognition.



Ke Feng is a Full Professor at Xi'an Jiaotong University, China. He is a Marie Curie Fellow (Imperial College London & Brunel University London). He received a Ph.D. degree from the University of New South Wales, Australia, in 2021. He worked at the University of British Columbia and the National University of Singapore in 2022 and 2023, respectively. His main research interests include digital twins, vibration analysis, structural health monitoring, dynamics, tribology, signal processing, and machine learning. He is recognized as the Emerging Leader (2023) by the Measurement Science and Technology journal. He is the Associate Editor and Guest Editor of several journals, including MSSP, JIM, SHM, IEEE TIM, IEEE TICPS, Neurocomputing, etc.



Xiaolong Yang received the Ph.D. degrees in mechatronic engineering in 2018 from Nanjing University of Aeronautics and Astronautics. From 2018 to 2019, he was a Postdoctoral Fellow and Mechatronics Laboratory Director with the Lab of Biomechanics and Intelligent Robotics in the Mechanical Engineering Department at City University of New York, City College, US. Currently, he is an associate professor with the School of Mechanical Engineering, Nanjing University of Science and Technology. His research interests include robotics, exoskeletons, and parallel mechanisms.



Zhiheng Zhao is currently a Research Assistant Professor at The Hong Kong Polytechnic University. Dr. Zhao obtained his Ph.D. in Industrial Engineering from The University of Hong Kong in 2017. His main research areas are the analysis and optimization of Industry 4.0 manufacturing and spatial-temporal tracking for sustainable logistics. In the past five years, he has published over 30 academic papers and has secured several research funds, including National Natural Science Foundation of China and the Guangdong Provincial Natural Science Foundation, amounting to approximately 5 million HKD. Dr. Zhao's research achievements have also been applied in the industrial sector, securing over 3 million HKD in corporate collaboration projects.



George Q. Huang received a B.E. degree from Southeast University, Nanjing, China, in 1983, and a Ph.D. degree from Cardiff University, Cardiff, U.K., in 1991, both in mechanical engineering. He joined the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, in December 2022 as Chair Professor of Smart Manufacturing. Prior to this appointment, he was Chair Professor of Industrial and Systems Engineering and Head of Department with the Department of Industrial and Manufacturing Systems Engineering at the University of Hong Kong.

Prof. Huang has published extensively and his works have been highly cited by research communities. He has conducted research projects in areas of Smart Manufacturing, Logistics, and Construction through IoT-enabled Cyber-Physical Internet and Systems Analytics. His research has been supported with substantial government and industrial grants. Prof. Huang is the Associate Editor and Editorial Member for several international journals. He is a Chartered Engineer (CEng), Fellow of IEEE, ASME, CILT, HKIE, IET, and IISE.