

Article

Intelligent Fault Diagnosis of Variable-Condition Motors Using a Dual-Mode Fusion Attention Residual

Fengyun Xie ^{1,2,*}, Gang Li ¹, Wang Hu ¹, Qiuyang Fan ¹ and Shengtong Zhou ¹

¹ School of Mechanical Electrical and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China; liganggrant@163.com (G.L.); gugu1199@163.com (W.H.); fgy18532671715@163.com (Q.F.); zhoust@ecjtu.edu.cn (S.Z.)

² State Key Laboratory of Performance Monitoring Protecting of Rail Transit Infrastructure, East China Jiaotong University, Nanchang 330013, China

* Correspondence: xiefyun@163.com

Abstract: Electric motors play a crucial role in ship systems. Detecting potential issues with electric motors is a critical aspect of ship fault diagnosis. Fault diagnosis in motors is often challenging due to limited and noisy vibration signals. Existing deep learning methods struggle to extract the underlying correlation between samples while being susceptible to noise interference during the feature extraction process. To overcome these issues, this study proposes an intelligent bimodal fusion attention residual model. Firstly, the vibration signal to be encoded undergoes demodulation and is divided into high and low frequencies using the IEEMD (Improved Ensemble Empirical Mode Decomposition) composed of the EEMD (Ensemble Empirical Mode Decomposition) and the MASM (the Mean of the Standardized Accumulated Modes). Subsequently, the high-frequency component is effectively denoised using the wavelet packet threshold method. Secondly, current data and vibration signals are transformed into two-dimensional images using the Gramian Angular Summation Field (GASF) and aggregated into a bimodal Gramian Angle Field diagram. Finally, the proposed model incorporates the Self-Attention Squeeze-and-Excitation Networks (SE) mechanism with the Swish activation function and utilizes the ResNeXt architecture with a Dropout layer to identify and diagnose faults in the multi-mode fusion dataset of motors under various working conditions. Based on the experimental results, a comprehensive discussion and analysis were conducted to evaluate the performance of the proposed intelligent bimodal fusion attention residual model. The results demonstrated that, in comparison to traditional methods and other deep learning models, the proposed model effectively utilized multimodal data, thereby enhancing the accuracy and robustness of fault diagnosis. The introduction of attention mechanisms and residual learning enable the model to focus more effectively on crucial modal data and learn the correlations between modalities, thus improving the overall performance of fault diagnosis.

Keywords: bimodal fusion; SE attention mechanism; residual network; variable operating condition



Citation: Xie, F.; Li, G.; Hu, W.; Fan, Q.; Zhou, S. Intelligent Fault Diagnosis of Variable-Condition Motors Using a Dual-Mode Fusion Attention Residual. *J. Mar. Sci. Eng.* **2023**, *11*, 1385. <https://doi.org/10.3390/jmse11071385>

Academic Editor: Yassine Amirat

Received: 11 June 2023

Revised: 4 July 2023

Accepted: 5 July 2023

Published: 7 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electric motors are extensively utilized in equipment, such as ship propulsion systems [1] and deep-sea robots [2]. Accurately identifying and diagnosing different types of faults in motor components is of utmost importance in ensuring the safety of maintenance operations. According to a major engine damage study conducted by the Swedish club, ship mechanical failures account for 47% of the total ship damage claims, resulting in economic losses of nearly 384 million US dollars. Specifically, 28% of these claims stem from marine diesel engine failures, causing losses of approximately \$13 million [3]. Research indicates that electrical faults in ships not only compromise navigation safety, but also contribute to severe maritime accidents, leading to significant personal and property losses [4]. Therefore, the development of motor fault diagnosis algorithms holds great significance for the safe operation and maintenance of ships and marine machinery equipment.

Ship electric motors, as a form of rotating machinery, operate in demanding conditions within real-world equipment scenarios. It is essential to consider that such equipment typically operates under various conditions, and the sensor-collected data often exhibits characteristics, such as indistinct features and noise interference. The term “condition” refers to the operational state of the equipment under conditions directly associated with its functioning. In addressing noise-related challenges, Wang S et al. [5] proposed a sparse gradient denoising optimization method for neural network models used in diagnosing rolling bearing faults in ship propulsion systems. This approach achieved sparse denoising by leveraging the influence values of network nodes. Sun H et al. [6] introduced a data-driven multi-wavelet denoising technology, successfully applied to extract weak features of minor faults in bearing inner rings. These findings highlight the significance of ensuring model robustness in high-noise environments [7]. However, due to the complexity of existing denoising models, an adaptive denoising method is proposed for signal processing. Regarding multiple operating conditions, Sharma S et al. [8] put forward a feature extraction method based on a weighted multi-scale fluctuation dispersion entropy, enabling the diagnosis of faults in the target system. Consequently, signal denoising and multi-operating condition feature extraction play a critical role in equipment fault diagnosis [9].

In the realm of feature extraction, attention mechanisms have emerged as a significant breakthrough, particularly in domains like computer vision [10]. Neural networks, when aided by attention mechanisms, possess the ability to focus on specific local positions or features within given information. By assigning corresponding weights, attention mechanisms can highlight crucial features, suppress relatively insignificant ones, and effectively enhance the overall performance of the model. Notably, Hu J et al. proposed the Squeeze and Extraction Network (SENet), which establishes an attention mechanism between feature channels to model their interdependencies and adaptively recalibrate the features [11]. Woo S et al. introduced the Convolutional Block Attention Module (CBAM) to optimize features in both channel and spatial dimensions, enhancing input features with constraints [12]. Similarly, Zhao T et al. designed the pyramid feature attention networks (PFANet), incorporating spatial and channel attention mechanisms. This approach leverages spatial attention to optimize lower-level features, employs a context perception pyramid method to extract higher-level features rich in semantics, and employs channel attention to further optimize them, thereby preserving more structural information [13]. Moreover, reference [14] adopts the Multi-scale Attention Convolutional Neural Network (MACNN), which employs channel attention to adjust the weights of different feature channels. It selectively learns effective fault features, reduces the impact of irrelevant features, and suppresses noise interference. Additionally, reference [15] proposes a time series attention module that performs temporal learning on fault features. By incorporating a temporal attention module after the network’s feature extraction stage, temporal feature dependencies between channels are established, facilitating the acquisition of fused channel temporal features. Furthermore, reference [16] constructs a fault diagnosis network based on one-dimensional convolutional neural networks (1DCNN), a gated recurrent unit (GRU), and an attention mechanism (Attention). This network addresses the issue of traditional fault diagnosis methods relying heavily on human expertise for feature extraction. Consequently, the utilization of spatial attention information holds significant importance in the context of mechanical fault diagnoses.

Furthermore, the intelligent diagnosis of three-phase motors using multi-sensor data fusion has garnered significant research attention. In addition to vibration signals, current and other signals can effectively characterize fault characteristics [17]. Scholars have focused on integrating multi-sensor data from different sources into a unified framework. Based on the level of information fusion, existing intelligent fault diagnosis methods for three-phase motors can be categorized into data-level fusion, feature-level fusion, and decision-level fusion [18]. Data-level fusion involves the direct integration of raw multi-sensor data. For instance, reference [19] proposes three data-level fusion methods that fuse vibration, speed, and load data. The paper explains the significance of different

fusion techniques from a physical perspective. Decision-level fusion, the most widely employed fusion method in current research, performs individual diagnoses using sensor data and, subsequently, fuses the diagnostic results. For example, in literature [20], the three-phase current signals of motors are fed into separate convolutional neural networks for the automatic feature extraction and classification. The fusion fault diagnosis is achieved by employing a supervised learning algorithm at the decision-making level. Similarly, reference [21] utilizes convolutional neural networks to extract features from vibration and current signals. Softmax classifiers are then used for pre-classification, followed by comprehensive fault diagnosis through the application of the D-S evidence theory. Feature-level fusion involves the fusion of information between input data and decision output. This process entails extracting fault features from various sensor data and fusing these features into a classifier for fault diagnosis. For instance, reference [22] employs an improved sparse filter to automatically extract fault features from vibration and current signals. The extracted features are then fused using head-to-tail stitching. Finally, a multi-channel limit learning machine classifier is utilized to identify faults in the motor rotor system. Despite some exploration and research in intelligent fault diagnosis of three-phase motors through multi-sensor information fusion, existing methods still exhibit certain shortcomings. Firstly, data from different sensor sources contain varying degrees of fault information, making it challenging to effectively fuse multi-sensor information at the data level. Similarly, decision-level fusion fails to fully leverage the complementary nature of multi-sensor data, limiting the diagnostic accuracy and generalization of intelligent models. Secondly, feature-level fusion diagnosis methods increase the complexity of the fault information mining and feature-extraction fusion process, leading to difficulties in deploying the model effectively.

In recent years, fault diagnosis has witnessed remarkable advancements with the application of machine learning techniques [23]. Among them, deep learning, as a powerful technology within the field of machine learning, has exhibited tremendous potential. Convolutional Neural Networks (CNN) [24], Residual Neural Networks (ResNet) [25], and other deep learning models have yielded fruitful results in fault diagnosis. Wen L et al. [26] proposed a data-driven fault diagnosis method based on CNN, which converts raw mechanical signals into grayscale images and trains the CNN model using these images. Chen Z et al. [27] employed a single convolutional layer combined with a fully connected layer for gear fault diagnosis, while Chen P et al. [28] used a one-dimensional convolutional neural network to diagnose vibration signals in gearboxes, demonstrating its high precision. Patil et al. [29] employed a CNN architecture trained on vibration spectrograms for the purpose of health monitoring of milling tool inserts. These deep learning-based fault diagnosis methods effectively extract hidden features and have found widespread application [30].

In response to these limitations, a novel fault diagnosis method for motors under variable operating conditions is proposed, leveraging a modal fusion attention residual model. This method addresses the challenges by employing several innovative techniques. Firstly, precise denoising of vibration signals is achieved using an enhanced empirical mode decomposition (EEMD) combined with a modal analysis and signal matching (MSAM) technique called the Integrated Ensemble Empirical Mode Decomposition (IEEMD). This denoising step ensures accurate and clean vibration signals. Secondly, the one-dimensional vibration signals and current signals are encoded into two-dimensional feature images. By mapping each operating state of the motor into a two-dimensional feature space using the Gramian Angle Field, a comprehensive representation of the motor's state is obtained. These representations are then aggregated into a bimodal fused Gramian Angle Field graph, which serves as input to the neural network. Finally, the ResNeXt architecture, known for its strong performance, is selected as the network backbone. To further enhance the model's capability, a Squeeze-and-Excitation (SE) attention mechanism with a switch activation function and a Dropout layer are added. The main innovations of this study encompass three aspects: (1) High-frequency precise denoising of the vibration signal, which undergoes image encoding. This is achieved by combining denoising processing

with image encoding techniques. The process involves utilizing the EEMD for signal decomposition, the MSAM for high and low frequency division, wavelet packet threshold denoising for high-frequency signal processing, and concluding with image encoding. (2) Utilizing the dual mode fusion Gram angle field diagram to aggregate the current signal and vibration signal into a single picture. This approach enables a single training picture to encapsulate a greater amount of information. (3) Employing the SE attention mechanism with the switch activation function and the ISE-ResNeXt model with a Dropout layer. This combination enhances the feature extraction capability, mitigates overfitting, and facilitates fault diagnosis of the three-phase asynchronous motors under multiple working conditions, even in the presence of noise.

The rest of this article is organized as follows. The Section 2 discusses the principles of the related work. The Section 3 introduces these methods, including the bimodal fusion of signals based on Gramian Angle Field plots and the proposed ISE-ResNeXt method. The Section 4 introduces the test results of this method on the dataset. The research conclusions and future research results are presented in Section 5.

2. Principles

2.1. EEMD-MASD Denoising Model

The core of the EEMD is EMD decomposition, which is an adaptive analysis method that can decompose signals into a series of *IMFs* containing different frequency components [31]:

$$y(t) = \sum_{i=1}^N IMF_i(t) + r_N(t) \quad (1)$$

In the equation, $y(t)$ represents the measurement signal, $IMF_i (i = 1, 2, \dots, N)$ represents the i -th *IMF*, N represents the number of *IMF*, and $r_N(t)$ represents the residual.

Noise is distributed in the high-frequency range and features are distributed in the low-frequency range. In order to select *IMFs* with a strong correlation with fault modulation signals, a scale selection standard based on cumulative mean (MSAM) is used to separate high-frequency *IMFs* from all *IMFs*. MSAM is defined as [32]:

$$\hat{h}_m = mean[\sum_{i=1}^m [IMF_i(t) - \frac{mean(IMF_i(t))}{std(IMF_i(t))}], m \leq N \quad (2)$$

where \hat{h}_m is the cumulative mean, *mean* represents the mean function, and *std* is the standard deviation. If it deviates from zero, the scale m represents the marker for dividing high-frequency *IMFs* and low-frequency *IMFs*, and $m - 1$ represents the number of high-frequency *IMFs*.

To eliminate the noise of high-frequency *IMFs*, wavelet threshold denoising is used to enhance the signal-to-noise ratio. Then, combine the denoised high-frequency *IMFs* with low-frequency *IMFs* to obtain the reconstructed signal, which is represented as:

$$y(t) = \sum_{i=1}^{m-1} \overline{IMF}_i(t) + \sum_{i=m}^N IMF_i(t) + r_N(t) \quad (3)$$

In the formula, \overline{IMF}_i represents the high-frequency *IMFs* after wavelet threshold denoising and IMF_i represents the low-frequency *IMFs*.

In comparison to prior research, certain studies have predominantly concentrated on the analysis of individual or a few independent *IMFs* for the extraction of fault-related features, overlooking the varying effectiveness of different *IMFs* in revealing faults. Moreover, certain signal processing techniques, such as the Weighted Average EEMD (WEEMD) filtering, still retain random noise components that can adversely impact the accuracy of fault feature extraction. To address this issue, this article presents an EEMD analysis method that incorporates the cumulative mean (MSAM) and wavelet threshold denoising.

This approach not only effectively prevents the omission of *IMFs* containing crucial fault information, but also significantly reduces random noise within the *IMFs*.

2.2. Gramian Angular Summation Fields

The essence of the Gramian Angular Summation Fields (GASF) image encoding is to encode through the Gramian Angular Field (GAF). The GAF is the encoding of time series in a polar coordinate system [33]. The GAF image encoding converts the time and amplitude corresponding to a point in a one-dimensional time series to the radius and angle in a polar coordinate system, and converts the data from one-dimensional data to two-dimensional space. The GAF generates the Gramian Angular Summation Fields based on cosine functions. The specific process of generation is as follows, showing that the original time series is first normalized to be between 0 and 1, which is defined in Equation (4):

$$\tilde{x}_t = \frac{x(t) - \min(X)}{\max(X) - \min(X)} \quad (4)$$

Among them, x_t is the vibration signal at the original time, t , and the scaled signal, \tilde{x}_t , at that time. The minimum and maximum values in X are represented by $\min(X)$ and $\max(X)$, respectively. Therefore, one can use polar coordinates to represent the rescaled time series, with the timestamp being the radius:

$$\begin{cases} \varphi_i = \arccos(\tilde{x}_t), -1 \leq \tilde{x}_t \leq 1, \tilde{x}_t \in X \\ r_i = t_i / N, t_i \in N \end{cases} \quad (5)$$

In the formula, t_i is the timestamp code at the time of the point, and r_i is the polar axis, preserving the temporal relationship; φ_i represents the polar angle, preserving the numerical relationship. After converting the recalibrated time series into a polar coordinate system, one can use the perspective of angles to identify the time correlation between different time intervals through the triangular sum/difference between each point. Using the GASF encoding, the encoding method is defined by Equation (6):

$$GASF = \begin{bmatrix} \cos(\varphi_1 + \varphi_1) & \dots & \cos(\varphi_1 + \varphi_n) \\ \cos(\varphi_2 + \varphi_1) & \dots & \cos(\varphi_2 + \varphi_n) \\ \vdots & \cos(\varphi_i + \varphi_i) & \vdots \\ \cos(\varphi_n + \varphi_1) & \dots & \cos(\varphi_n + \varphi_n) \end{bmatrix} \quad (6)$$

In the equation, φ_i represents the angle value of the i -th sequence. It is evident that after encoding in this way, the order of time from the top left to bottom right on the two-dimensional image is preserved. The original information is retained at the diagonal position, while other regions express relationships between different time sequences. For a vibration signal with an original time series length of n , a numerical matrix with a size of $n \times n$ is obtained through GASF encoding.

2.3. Squeeze-and-Excitation Networks

The choice of activation function in a deep network has a significant impact on the performance of the model. The correction linear unit (ReLU) is the most widely used activation function, and the Swish activation function often performs well in deep networks.

The Swish activation function, also known as the self-gating activation function, was proposed by Google in 2017 [34]. It is verified that Swish can improve the accuracy of the model more than the ReLU activation function under the same conditions.

The expression of the Swish activation function is shown in Formula (1), where β can be a constant or a parameter obtained through training. When $\beta \rightarrow \infty$, Swish is a ReLU activation function, when $\beta = 0$, Swish becomes a linear function. Therefore, Swish can be regarded as a smooth activation function between the two [35]. When $X > 0$, there is no gradient disappearance, and when $X < 0$, neurons do not die like they do with the

ReLU. Compared with the ReLU, the derivative of Swish is not invariant, and Swish is differentiable and continuously smooth everywhere:

$$\text{Swish}(X) = x \cdot \text{sigmoid}(\beta X) \quad (7)$$

The essence of the attention mechanism is to imitate the human visual attention mechanism, ignoring the context-irrelevant information, which is widely used in the field of natural language processing. The attention mechanism in neural networks is commonly used in the channel dimension, filtering important information from a large amount of input information and assigning different weights to this information.

As shown in Figure 1, the SE module (SENet for short) mainly includes two operations: Squeeze and Excitation, which can be applied to any mapping; $F_{tr} : X \rightarrow U$ with the input $X \in R^{H \times W \times C}$ and output $U \in R^{H \times W \times C}$. The SENet proposes a squeezing operation, namely $F_{sq}(\cdot)$ in Figure 1, which encodes the entire spatial feature on a channel into a global feature. The calculation formula is:

$$Z_c = F_{sq}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (8)$$

In the formula, Z_c represents the weight of the extrusion operation, $c \in (1, C)$, where C is the number of feature channels; W and H represents the size of the feature map; $U_c(i, j)$ is the value in the i -th row and j -th column of the input feature value channel, c . The squeezing operation obtains global descriptive features and grasps the relationships between channels through another operation.

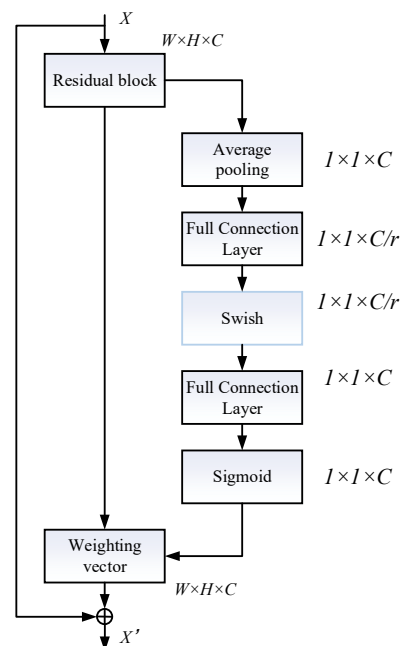


Figure 1. The SE attention model.

The purpose of the incentive operation is to obtain the correlation between feature channels, which needs to meet two criteria: firstly, the nonlinear relationship between each channel can be learned; secondly, the learning relationship is not mutually exclusive, as it allows for multi-channel features rather than one hot form.

The above entire operation can be seen as learning the weight coefficients of each channel, which makes the model more discriminative of the features of each channel, thus forming an adaptive attention mechanism.

2.4. The ResNeXt Network Model Structure

Traditional convolutional neural networks need to deepen or widen the network to improve the recognition accuracy of the model. However, as the number of layers in the network deepens, gradient explosion or vanishing phenomena may occur, and network degradation may also occur. The ResNeXt network structure [13] adopted in this paper can stack deep network models to improve the identification accuracy, while reducing the number of hyperparameters. It changes the idea of the traditional VGG network and ResNet network stacking, borrows the Inception network series division transformation aggregation strategy, and transforms the single path convolution into multiple convolutions of multiple branches, with the same topology structure, reducing the design of hyperparameter, and facilitating transplantation. Adding a short cut to the simplified Inception network and combining it with a residual neural network can result in a ResNeXt neural network. The ResNeXt network utilizes a Block structure with grouped convolutions, as illustrated in Figure 2. In this configuration, the input feature matrix undergoes dimension reduction by employing a 1×1 convolutional kernel, reducing the channel count to half of its original value. Next, 32 convolutional groups are applied, each composed of multiple 3×3 convolutional kernels, for the purpose of feature extraction. Subsequently, the extracted features are aggregated and concatenated. Finally, through 1×1 , after the convolutional kernel is dimensionalized and the input and output are added, the final output is obtained by activating the ReLU function. The formula for the aggregation transformation of ResNeXt network is as follows:

$$y = x + \sum_{i=1}^C T_i(x) \quad (9)$$

In the formula, y represents the output; x is the input; C is the base number; and $T_i(x)$ is the i -th point mapping function and can be in any form.

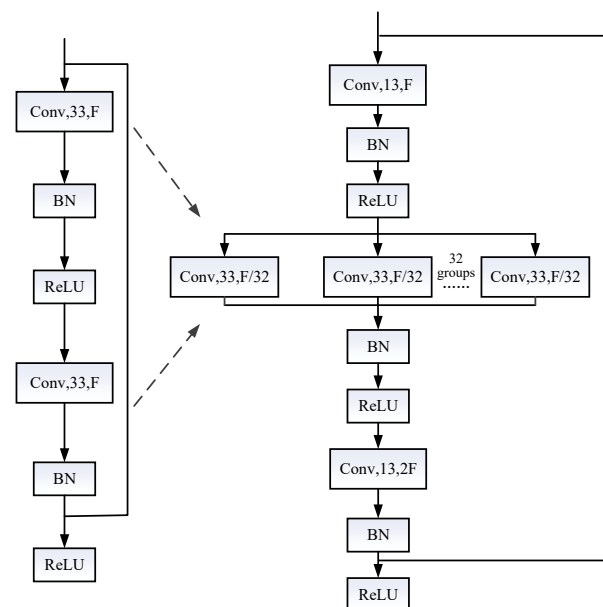


Figure 2. The ResNeXt cell structure.

3. The Intelligent Bimodal Fusion Attention Residual Model for Fault Diagnosis of Motors under Variable Operating Conditions

The fault diagnosis process for motors involves several steps, as illustrated in Figure 3. Firstly, the vibration and current signals of the rotating machine were obtained, followed by denoising of the vibration signal using the EEMD–MSAM. Secondly, a Gramian Angle Field graph was constructed for the vibration signals, incorporating motor characteristics through

the Gram matrix. Utilizing the diagonal symmetry of the Gramian Angle Field graph, a fault diagnosis graph for motors was generated. Subsequently, the proposed fault diagnosis method, named SE-ResNeXt, comprised of three main components. The first component included the input layer, convolution layer, and batch normalization layer. The second component was the core of the model, consisting of the ResNeXt unit structure (Figure 2) and the Squeeze and Extraction Networks (Figure 1). This combination is repeated four times. The third component consisted of a pooling layer and a fully connected layer, responsible for outputting the training results. Finally, the constructed image was fed into the proposed model to achieve fault diagnosis of motors through image classification.

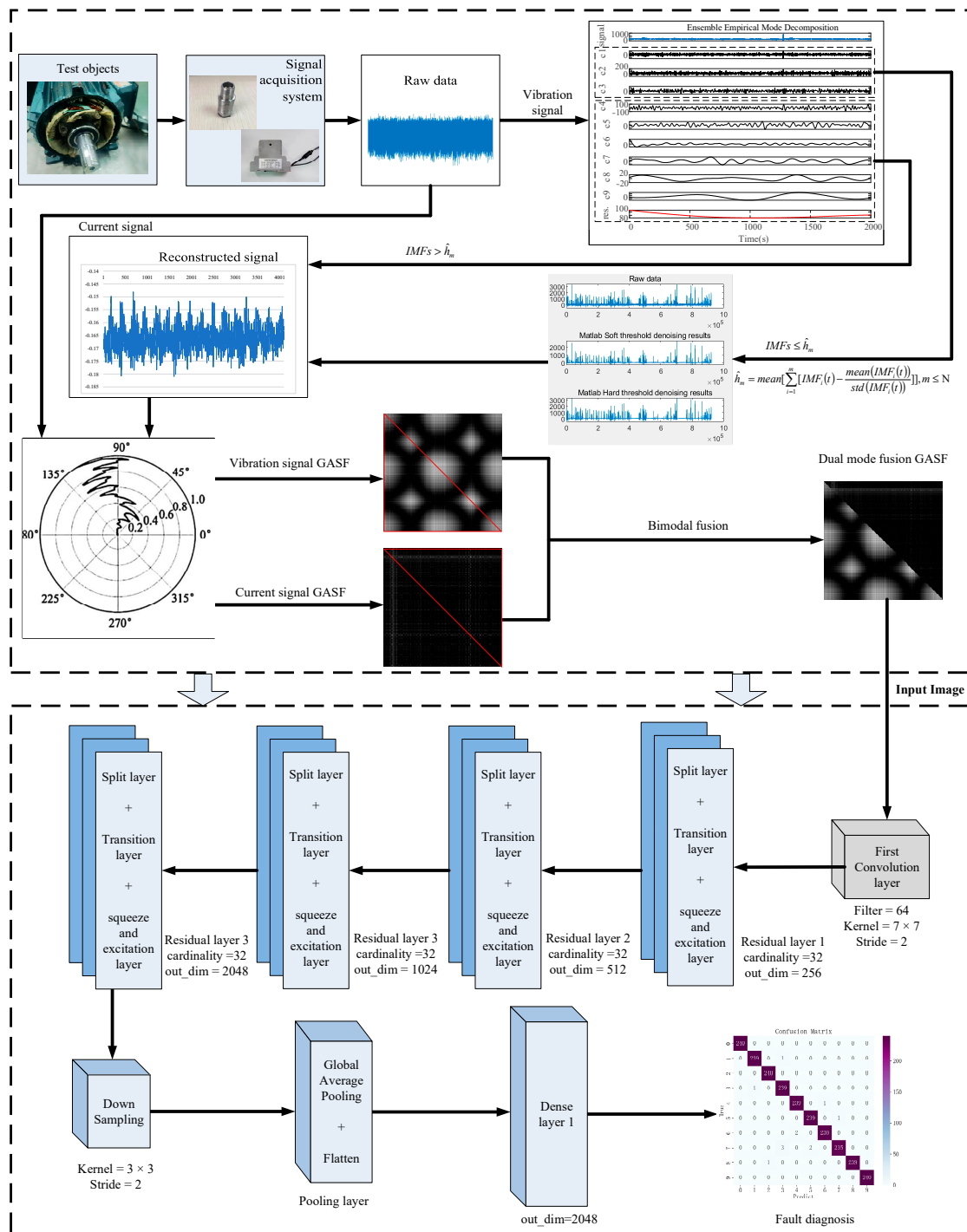


Figure 3. The intelligent bimodal fusion attention residual network for fault diagnosis of motors.

3.1. Data Acquisition and Signal Preprocessing

To develop a reliable fault diagnosis model, it is crucial to gather multimodal datasets. For this study, data from multiple sensors, including current and vibration, were carefully selected. Signal preprocessing plays a vital role in motor fault diagnosis as it aims to eliminate noise and interference from the signal, thereby enhancing the accuracy of subsequent feature extraction and fault diagnosis processes. In this study, the researchers employed an EEMD–MSAM-based method to process vibration signals. EEMD, a time-frequency analysis technique, decomposes signals into sub-signals with different frequency ranges. In vibration signals, valuable low-frequency features are typically concentrated in the low-frequency portion, while noise predominantly manifests as high-frequency components. By selectively choosing appropriate sub-signals through MSAM, one can effectively preserve low-frequency features while eliminating high-frequency noise. This method enables the distinction between high-frequency and low-frequency components within the vibration signal, ultimately improving the overall signal quality.

3.2. The Gramian Angle Field Diagram of Bimodal Fusion

To enhance the recognition accuracy of the conventional Gramian Angle Field graph model, it is often necessary to increase either the number of samples or the complexity of the model. However, as the network layers deepen, the model size grows, and the training time becomes extensive. In this article, the authors proposed a bimodal fusion approach using the Gramian Angle Field images to augment the information content in each image, thereby improving the recognition accuracy without escalating the neural network's complexity. Departing from the conventional practice of stacking training samples in neural networks, the researchers leveraged the diagonal symmetry of the Gramian Angle Field graph and integrated the two modes into a single image. This merged image retained the same length and width as the individual mode images, effectively reducing the number of training samples required and facilitating seamless transplantation of the approach.

3.3. The SE ResNext Based on Swish Activation Function

The complete architecture of the proposed method is depicted in Figure 3. In this study, the researchers introduced a deep learning-based model that leveraged the ResNext architecture and incorporated the Swish activation function within the SENet block. Initially, the data had undergone normalization using the mean standard deviation approach to alleviate the influence of outliers or extreme values. To mitigate issues, such as overfitting and bottleneck, the researchers employed four ResNeXt modules, with each module comprising of three sub-layers: the split layer, transition layer, and the squeeze–excitation layer. The split layer encompassed 32 branches, providing ample capacity for capturing diverse features. The transition layer combined a pooling layer with a convolutional layer, effectively reducing the dimensionality of the image. Moreover, this project adopted the squeezing–excitation network, replacing the original activation function with the more efficient Swish activation function. This substitution improved accuracy without incurring additional computational costs or processing time.

Subsequently, the features were fed into the generator layer, which performed down-sampling on the features. To combat overfitting, this study incorporated global average pooling and Dropout layers. The fully connected layer served as the classification layer, dividing the results into ten categories. In the ResNeXt architecture, the output dimension was set to 256 in the first layer, followed by increases to 512 in the second layer, 1024 in the third layer, and 2048 in the fourth layer.

4. Experimental Research and Results Analysis

4.1. The Experimental Platform and Data Processing

4.1.1. The Experimental Platform

To assess the efficacy of the proposed method, the SE–ResNeXt was employed to detect and classify various types of faults in a motor test bench, as depicted in Figure 4.

The test bench primarily comprised of a motor, a gearbox reducer, a frequency converter, and a magnetic powder brake. Moreover, the vibration signal acquisition system primarily consisted of a YE6231 acquisition card, a CAYD051V acceleration sensor, and acquisition software. During the experimental setup, each faulty component was installed in a motor, and vibration acceleration sensors were positioned at both the drive end and fan end. This arrangement facilitated the collection of vibration signals to evaluate the performance of the proposed method in fault identification.

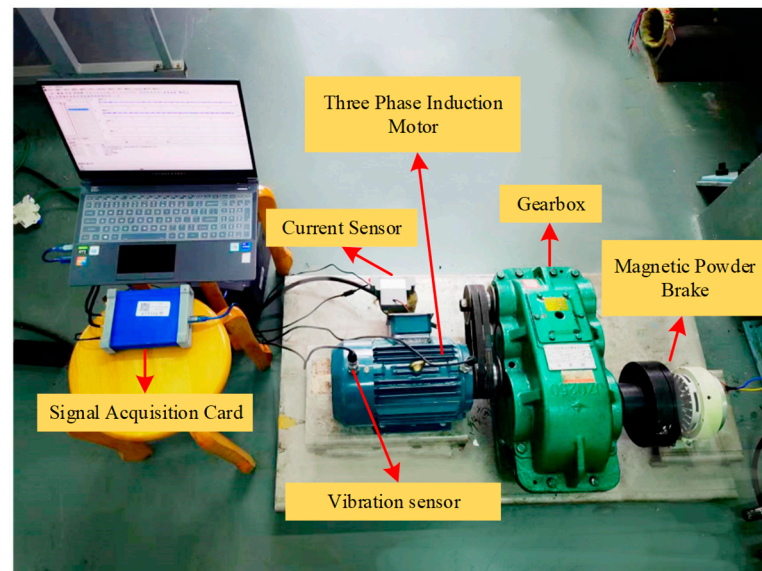


Figure 4. The Experimental Platform.

In this study, a dataset comprising of 10 different types of vibration signals from motor states was collected. During the experiment, a controller was employed to regulate the motor speed, gradually increasing it at a stable rate.

In the above figure, the theoretical maximum output speed of the YE2100L2-4 motor was 1500 r/min, while f_{\max} represents the maximum output frequency of the frequency converter at 50 Hz. When adjusting the motor speed to 600 r/min, 900 r/min, and 1200 r/min, the output frequency of the frequency converter was set to 20 Hz, 30 Hz, and 40 Hz, respectively.

To ensure accurate data collection, the collection mode was set to “timed trigger”. The sampling frequency (f_s) was configured at 12 KHz, the sampling duration was set to 8 s, and the sampling interval was established at 2 s. These parameters were selected to capture relevant information during the experimental process.

Vibration signals were collected under various health conditions, and the collected signals were divided into 800 samples. Each sample consisted of 800 data points, obtained by segmenting the original vibration signal with a sampling length of 640 K. The sampling frequency used in the experiment was 12,000 Hz.

During the experiment, the drive motor’s minimum rotation speed was set at 900 rpm. With a sampling length of 800 points, the corresponding acquisition time for each sample was 0.067 s, which corresponded to the minimum rotation period of the motor’s rotor. Thus, each sample captured the fault frequencies associated with bearings, rotors, and stators.

4.1.2. Experimental Data and Processing

This study aimed to perform diagnostic tasks on motors, specifically targeting bearing, rotor, stator, and bearing faults, as well as the overall health status. The diagnostic tasks were conducted under 3 different speed operating conditions, specifically at 30 Hz and 40 Hz. Figure 5 illustrates the 10 vibration signal types associated with motors.

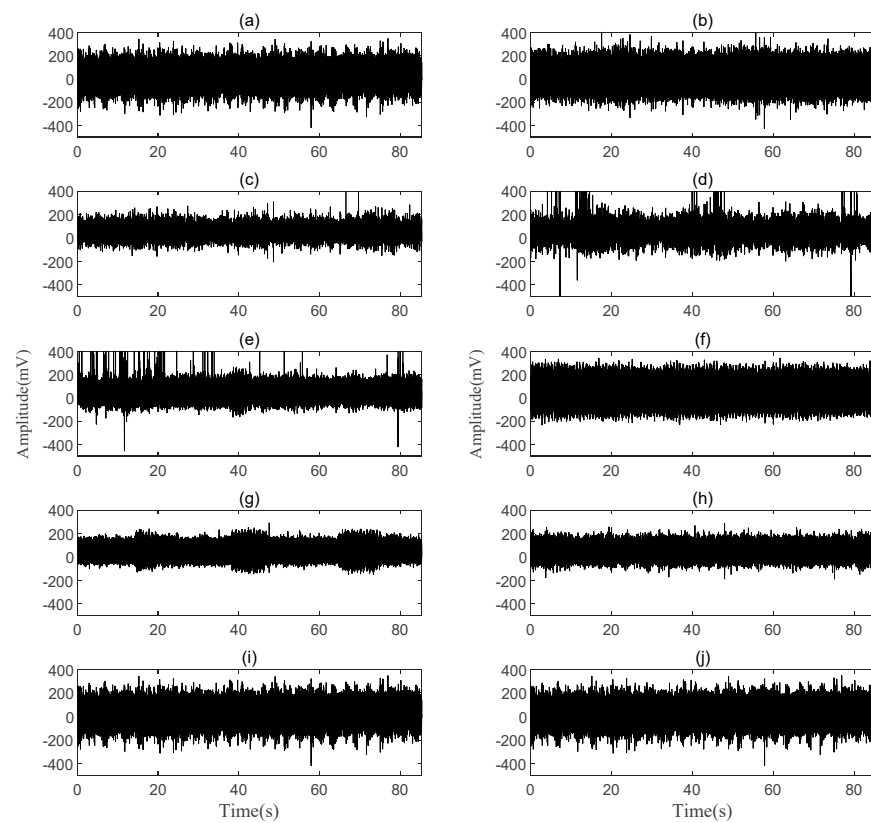


Figure 5. The ten vibration signals of motors: (a) end ring cracking 30 Hz; (b) end ring cracking 40 Hz; (c) broken rotor bar 30 Hz; (d) broken rotor bar 40 Hz; (e) health 30 Hz; (f) health 40 Hz; (g) turn-to-turn short circuit 30 Hz; (h) turn-to-turn short circuit 40 Hz; (i) bearing failure 30 Hz; (j) bearing failure 40 Hz.

- **Parameter Discussion of Denoising Methods:**

The denoising of vibration signals contaminated by noise is a well-known issue in vibration signal analysis, particularly in the case of asynchronous motors. The primary objective is to eliminate as much noise as possible while retaining crucial information within the signal. To evaluate and estimate the effectiveness of signal denoising, two commonly used indicators are the signal-to-noise ratio (SNR) and the root mean square error (RMSE).

The signal-to-noise ratio (SNR) is an essential metric used to quantify the quality of a denoised signal. It assesses the ratio of signal power to noise power and is typically expressed in decibels (dB). The formula to calculate SNR is as follows:

$$SNR = 10 \log_{10} \left(\frac{p_s}{p_n} \right)^2 \quad (10)$$

where $p_s = \frac{1}{n} \sum_n f^2(t)$ is the power of the useful components of the original signal, $p_n = \frac{1}{n} \sum_n [f(t) - \hat{f}(t)]^2$ is the power of the noise component, $f(t)$ represents the original signal, and $\hat{f}(t)$ represents the denoised signal.

The root mean square error (RMSE) is a commonly used metric for evaluating the performance of signal denoising methods. It quantifies the average difference between the denoised signal and the original signal. The expression for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{n=1}^N [f(t) - \hat{f}(t)]^2} \quad (11)$$

This section aims to compare the parameters of wavelet packet threshold signal preprocessing methods using the vibration signal of a motor as an example for processing and analysis. The selection of basis functions plays a crucial role in wavelet and wavelet packet denoising. It is essential to choose suitable basis functions that are specific to the analysis objectives; otherwise, signal denoising and filtering cannot be effectively achieved.

Currently, there are over ten series of wavelet basis functions, with the total number reaching into the hundreds. A literature review showed that the Db4 wavelet performs the best when applying wavelet denoising to vibration signals [14]. Additionally, Db10 and sym4 have also been utilized for vibration denoising. Therefore, for the decomposition of the vibration signal of the motor with noise into different basis functions, a combination of soft threshold denoising (STD) and hard threshold denoising (HTD) techniques was employed. The Db3–10 and Sym3–10 series wavelets were specifically used for this purpose.

The results of the analysis and comparison of the various wavelet functions are presented in Table 1.

Table 1. A comparison of denoising signal-to-noise ratio and root mean square error.

Function	Method	RMSE	SNR	Function	Method	RMSE	SNR
Sym3	HTD	1.1023	34.8392	Db3	HTD	1.1023	34.8392
Sym4	HTD	1.1170	34.7243	Db4	HTD	1.1220	34.6859
Sym5	HTD	1.1410	34.5395	Db5	HTD	1.1455	34.5054
Sym6	HTD	1.1588	34.4056	Db6	HTD	1.1537	34.4436
Sym7	HTD	1.1681	34.3360	Db7	HTD	1.1723	34.3047
Sym8	HTD	1.1841	34.2178	Db8	HTD	1.1822	34.2316
Sym9	HTD	1.1935	34.1493	Db9	HTD	1.1807	34.2428
Sym10	HTD	1.1973	34.1214	Db10	HTD	1.1947	34.1404
Sym3	STD	2.8546	26.5746	Db3	STD	2.8546	26.5746
Sym4	STD	2.8280	26.6558	Db4	STD	2.8217	26.6753
Sym5	STD	2.8066	26.7218	Db5	STD	2.8054	26.7257
Sym6	STD	2.7864	26.7848	Db6	STD	2.7884	26.7785
Sym7	STD	2.7660	26.8486	Db7	STD	2.7734	26.8253
Sym8	STD	2.7556	26.8812	Db8	STD	2.7571	26.8766
Sym9	STD	2.7444	26.9164	Db9	STD	2.7415	26.9257
Sym10	STD	2.7321	26.9556	Db10	STD	2.7265	26.9734

According to the findings in Table 1, it was evident that when using the db3 and sym3 basis functions, the SNR value was maximized, and the RMSE was minimized. However, considering the better performance of the subsequent Db series basis functions compared to the Sym series, the db3 basis function was chosen.

Furthermore, the denoising effect of the hard threshold method was superior to that of the soft threshold method. This was because soft threshold denoising tends to smooth the signal, potentially missing out on important features. Therefore, the hard threshold denoising method with the db3 basis function was adopted.

By selecting the db3 basis function and employing the hard threshold denoising approach, it was expected to achieve effective denoising of the vibration signal from the motor while preserving essential signal characteristics.

- Examples of denoising methods:

To provide a detailed demonstration of the denoising method used in this study, this project first processed a set of motor vibration signals using the EEMD. The EEMD adaptively decomposed the signal into 15 Intrinsic Mode Functions (IMFs), as depicted in Figure 6.

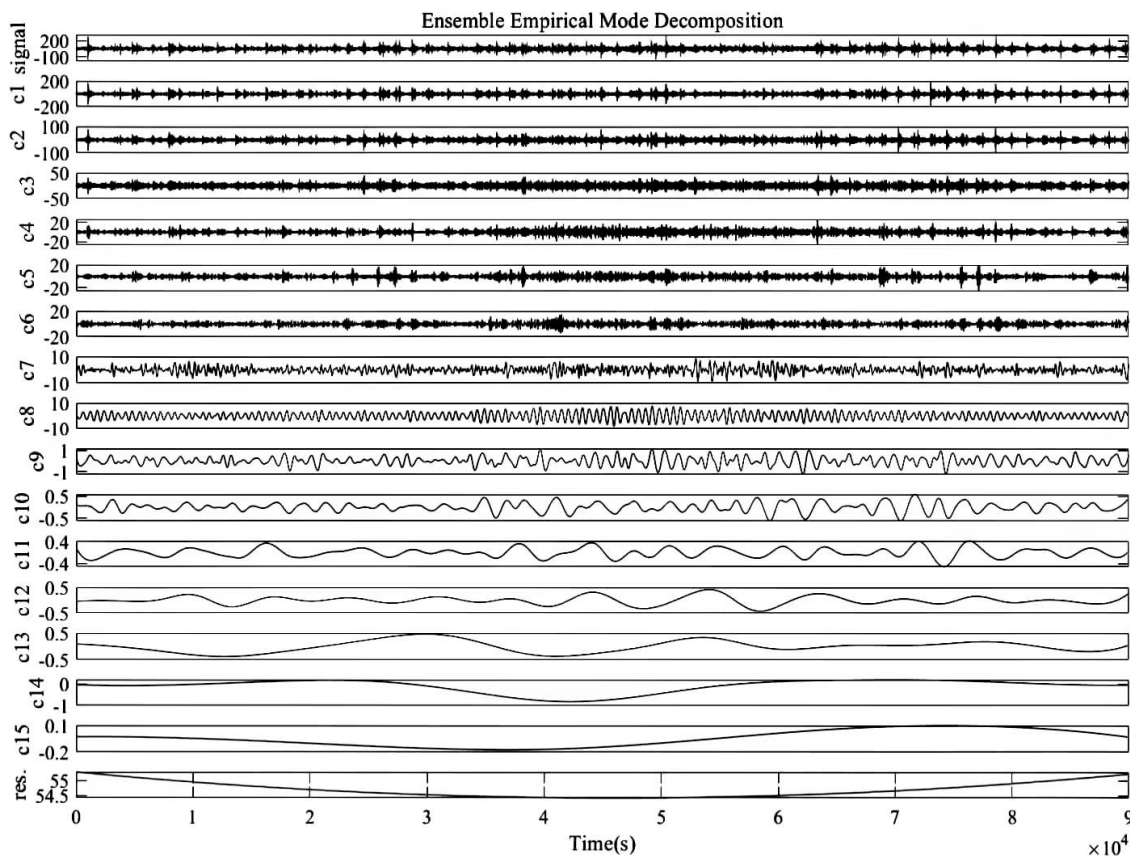


Figure 6. Decomposition results of the vibration signal.

Secondly, to effectively extract fault feature information, the mean of the standardized accumulated modes (MSAM) was employed to classify the decomposed *IMFs* into two categories: low-frequency *IMFs* and high-frequency *IMFs*. In accordance with Equation (2), a threshold of 12 was utilized to differentiate the high-frequency *IMFs* from the low-frequency *IMFs*. This differentiation is visually illustrated in Figure 7.

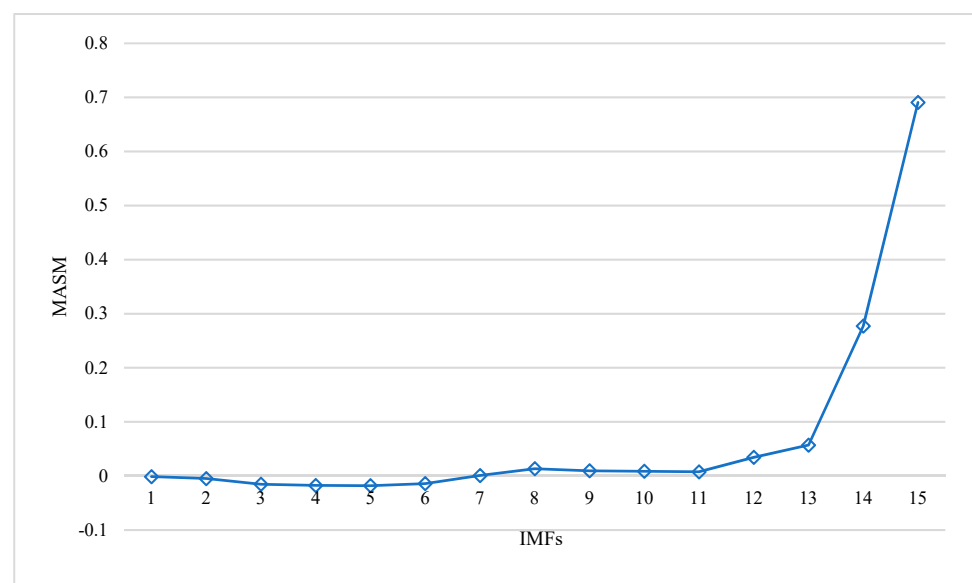


Figure 7. Values of the MSAM.

Finally, among the decomposed *IMFs*, the high-frequency *IMFs* (*IMF1* to *IMF11*) had undergone denoising processing using the wavelet threshold denoising algorithm. This denoising step aims to reduce noise and enhance the fault-related information contained in these high-frequency *IMFs*. Subsequently, the denoised high-frequency *IMFs* were reconstructed together with the low-frequency *IMFs* (*IMF12* to *IMF15*) to obtain fault-related reconstructed signals. The resulting reconstructed signals capture the relevant fault information, as shown in Figure 8.

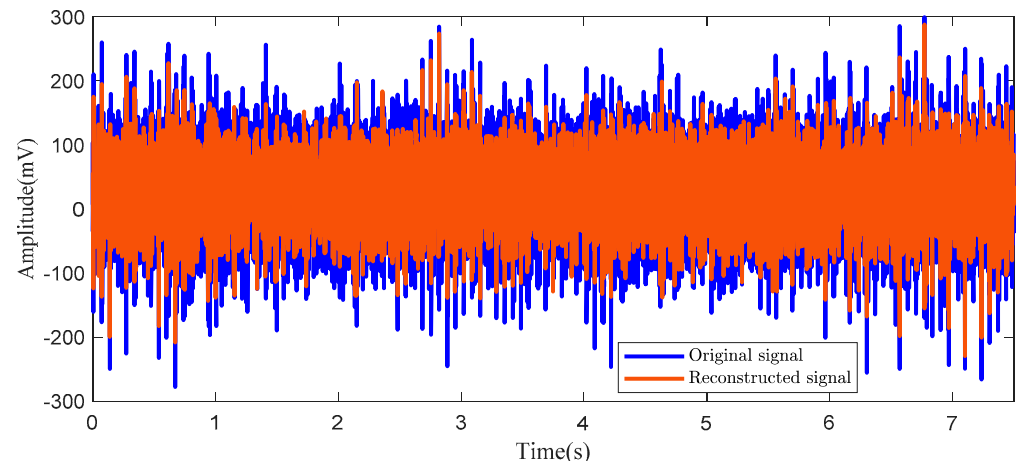


Figure 8. A Noise Reduction Comparison.

4.1.3. The Bimodal Fusion of Experimental Data in a Two-Dimensional Format

In the experimental setup, the first 400 samples of each type of fault vibration signal were utilized as the training set. Additionally, the validation set consisted of 160 samples, while the test set contained 240 samples. This experimental design aimed to enhance the similarity between the experimental results and the actual application effectiveness of the method. The reason for using this particular data distribution approach was to better align with practical engineering applications. In real-world fault diagnosis models, the training dataset typically consists of collected data and may not include a substantial amount of validation and testing data. Therefore, a ratio of 5:2:3 for the training, validation, and testing datasets was deemed more suitable than the traditional ratio of 7:2:1.

To comprehensively capture the health status information of the data, the Gramian Angle Field (GAF) diagram was employed. The GAF diagram was particularly advantageous for describing two-dimensional time-domain data, enabling a more comprehensive representation of the data's health condition. In the experiment, the one-dimensional vibration signal and current signal obtained from the sensor served as inputs to the network. These signals were transformed into a two-dimensional bimodal fused Gramian Angle Field map. Taking the data from the motor in this experiment as an example, the health characteristics, as well as the characteristics of rotor bar breakage, inter-turn short circuit, segment ring cracking, and bearing failure obtained from the sensors, are depicted in Figure 9.

Upon careful observation, it became evident that the operating states of the motor exhibited substantial differences under various conditions. Specifically, the fault state demonstrated distinct characteristics within the bimodal Gramian Angle Field (GAF) diagram, exhibiting varying degrees of intensity. This compellingly demonstrated the efficacy of the dual-mode fusion Gramian Angle Field graph method in accurately describing the diverse states of motors.

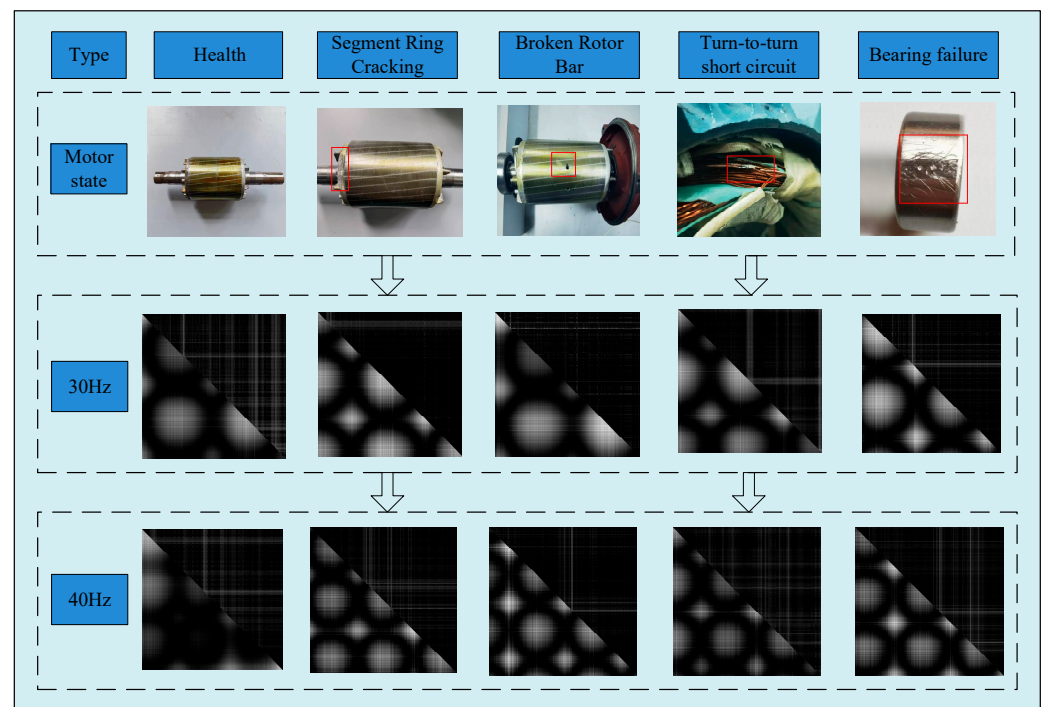


Figure 9. A dual-mode Gramian Angle Field diagram under various working conditions.

4.2. Performance Analysis of Intelligent Dual-Mode Fusion Attention Residual Models

To assess the superiority of the proposed method, this research conducted a comparative analysis between the ISE-ResNeXt (Improve SE-ResNeXt) and other models—namely, the ResNeXt, ResNet, and CNN—in terms of their performance for fault diagnosis of the motor testing platform dataset. For each model, this research considered both denoised and non-denoised Gramian Angle Summation Field (GASF) as input features. Comparisons were conducted using both single-mode GASF and multimodal GASF approaches. All models were optimized using the Adam optimizer, with a cross-entropy loss function. The learning rate was attenuated using a multi-cycle cosine annealing strategy, with an initial learning rate of 0.0001 and a minimum value of 1×10^{-8} . During training, this project set a predefined number of iterations and determined the optimal parameters by monitoring the loss values.

All networks were implemented using the TensorFlow 2.6.1 deep learning framework. The hardware setup comprised of an 11th Gen Intel Core i7-11800H @ 2.30 GHz CPU, Nvidia GeForce RTX 3070 GPU, and 16 GB RAM. The dataset consisted of a combination of 5 motor states and 2 working conditions, resulting in a total of 10 categories. The corresponding label codes for the motor states are provided in Table 2.

Table 2. Binary Label Encoding.

Numeral	Category	Numeral	Category
0	End ring cracking 30 Hz	5	Health 40 Hz
1	End ring cracking 40 Hz	6	Turn-to-turn short circuit 30 Hz
2	Broken Rotor Bar 30 Hz	7	Turn-to-turn short circuit 40 Hz
3	Broken Rotor Bar 40 Hz	8	Bearing failure 30 Hz
4	Health 30 Hz	9	Bearing failure 40 Hz

4.2.1. Comparing Fault Diagnosis Models for Motors: The SE ResNeSt vs. the ResNeXt in Single Modal Data (Non-Denoised and Denoised)

To effectively validate the efficacy of the denoising method for fault diagnosis types, this research utilized a confusion matrix to visualize the recognition performance under

different models and signal processing conditions in Experiment 1. The results of the calculations are presented in Figure 10.

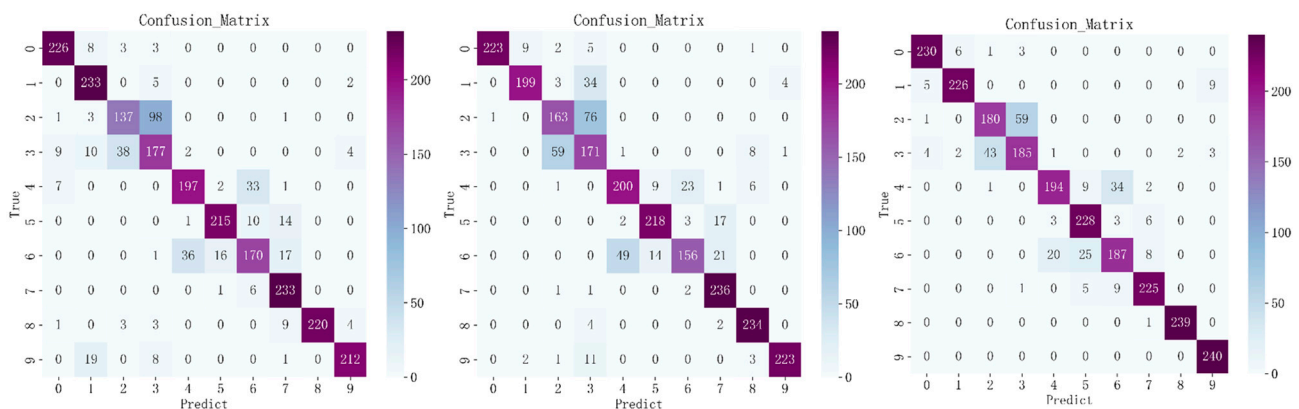


Figure 10. A confusion matrix of denoising results.

The figure illustrated the datasets consisting of single-modal data, with (a) and (b) representing the non-denoised data and (c) representing the denoised data. Regarding the models used, (a) represented the original ResNeXt model, while (b) and (c) utilized the SE-ResNeSt with a Dropout layer. The confusion matrix revealed that the accuracy rates for (a), (b), and (c) are 84.17%, 84.29%, and 88.92%, respectively. It was observed that, under the same model, the IEEMD denoising method enhanced the fault recognition rate by 4.63%, with improved recognition rates for various faults. The fault recognition performance was poorest without denoising, slightly better with the SE-ResNeSt model with a Dropout layer, and best when using denoised data combined with this model. However, despite these improvements, the final recognition rate was still not ideal. This can be attributed to two main factors. Firstly, the dataset encompassed multiple operating conditions, with minimal differences between each state, making identification challenging. Secondly, the limited information content of the Gramian Angle Field diagram used in the single-mode hampered the model's ability to aggregate fault features, resulting in unsatisfactory fault diagnosis accuracy for motors.

In summary, the combined denoising method of IEEMD and wavelet packet threshold denoising has shown effectiveness in suppressing noise. The improvement of the ResNeXt model had also resulted in an improved accuracy rate. However, despite these advancements, the overall recognition rate was still not at a satisfactory level. Therefore, further exploration and research are required to refine the motor fault diagnosis method and enhance its performance.

4.2.2. Comparison of Different Modes and Model Results

To assess the effectiveness of bimodal fusion, the model will undergo a comprehensive evaluation based on aspects such as the training process and accuracy. Figure 11 illustrates the training process of different modalities and neural networks. In this study, an analysis of various models was conducted through their training processes to showcase the superiority of the proposed model.

The training loss function and accuracy data showed that initially, all models exhibited a low training accuracy, but as training progressed, accuracy gradually improved. Throughout the entire training process, the training accuracy of each model showed a consistent upward trend, albeit with occasional decreases or fluctuations. In the final few training cycles, all models achieved relatively high training accuracy levels, approaching or surpassing 90%. Among the models, the DGASF + ISE-ResNeXt consistently maintained a high training accuracy throughout the entire training process and outperformed other models in the later stages, reaching the highest level of accuracy.

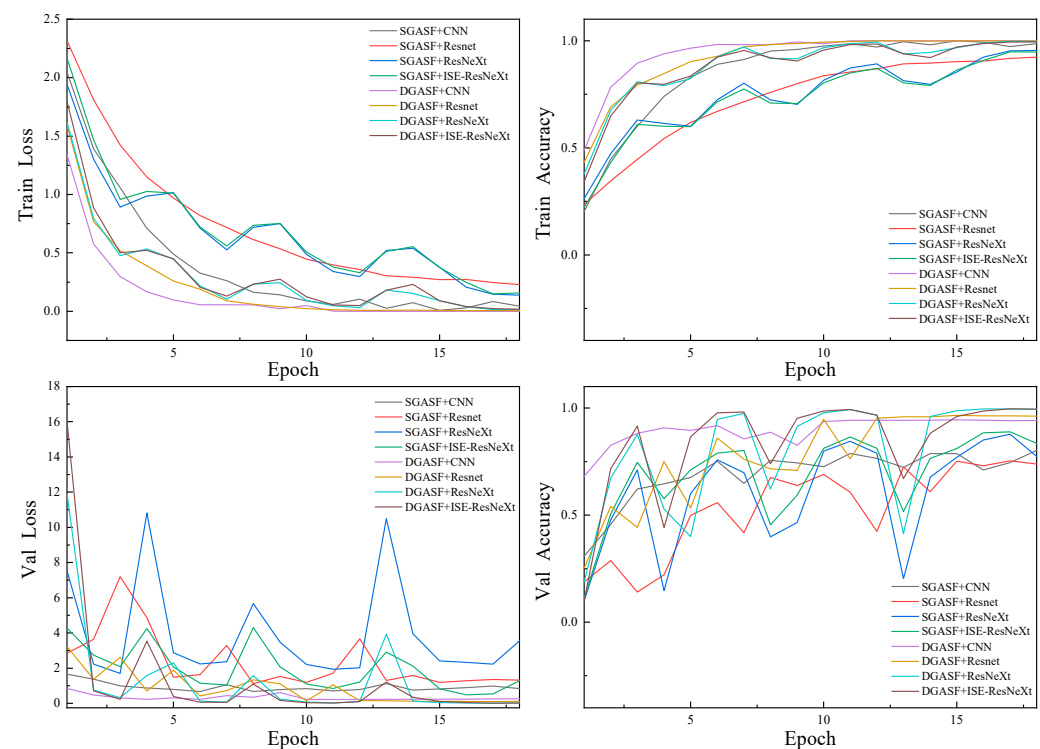


Figure 11. The Training Processes of the Different Methods.

Furthermore, based on the provided loss function and accuracy data, the DGASF + ISE-ResNeXt model demonstrated the best performance throughout the training process. It exhibited the lowest loss function, indicating a better fit to the training data. The graph also illustrated significant early-stage improvements in validation accuracy for all models, with the growth rate slowing down over time. Ultimately, the validation accuracy of each model gradually increased and stabilized in the later stages. Considering the validation set, the DGASF + ISE-ResNeXt model consistently performed the best throughout the entire training process, exhibiting the lowest loss function and achieving the highest validation accuracy of approximately 99%. These results indicated that the model effectively fits the training data.

It is important to note that higher training accuracy does not always guarantee good performance on the validation set. This can be observed in the comparison of training accuracy and validation accuracy. For instance, the SGASF + CNN model exhibited relatively high training accuracy on the training set but performs poorly on the validation set. Conversely, the DGASF + CNN model achieved a training accuracy of 0.4882 on the training set but demonstrated relatively good performance on the validation set.

However, it is worth mentioning that the chosen model in our study seemed to exhibit oscillations during the training process with convergence curves that may not appear as visually appealing as those of a regular CNN. This was due to the CNN models' susceptibility to getting stuck in local optima, which can produce more aesthetically pleasing curves. In contrast, the selected model emphasized strong feature extraction capabilities, which can be more sensitive to various factors. As a result, the training curve may not have possessed an aesthetically pleasing shape, but it achieved high accuracy and is less prone to falling into local optima.

Furthermore, the recognition effect of different models and modes in Experiment 2 was visualized using confusion matrices. The calculation results can be observed in Figure 12.

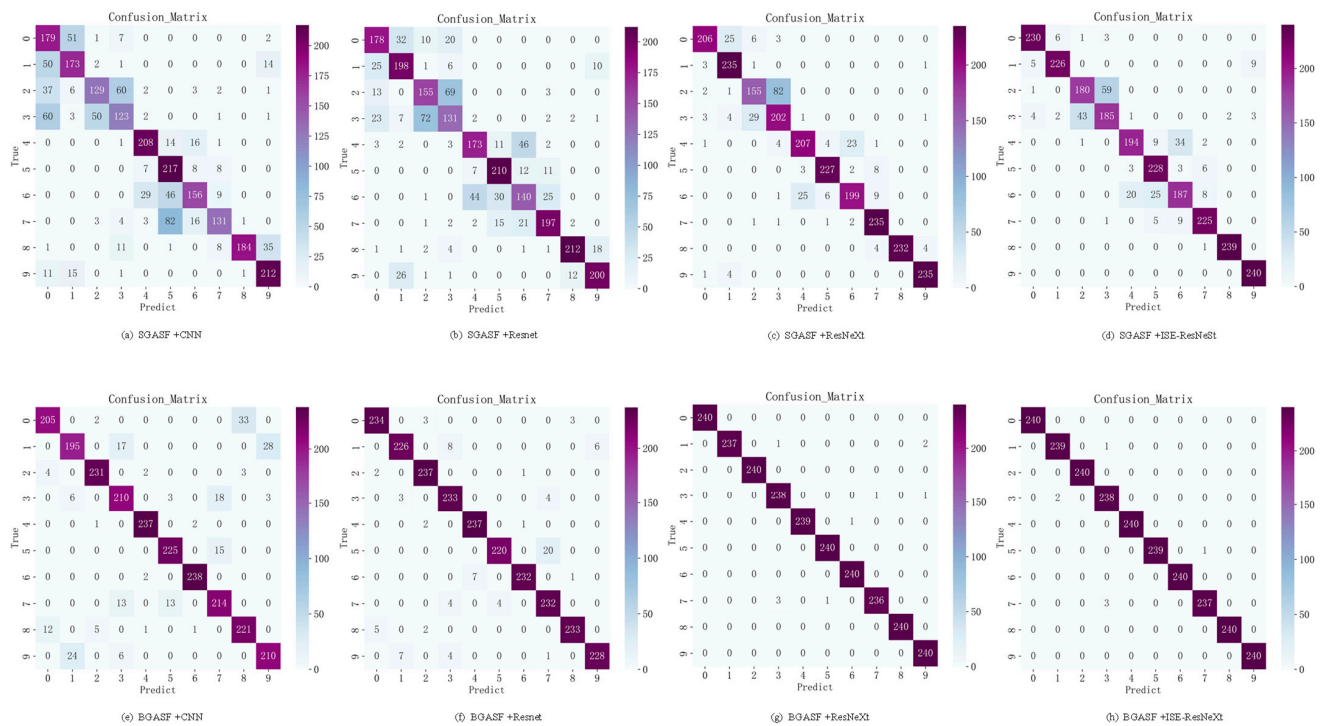


Figure 12. The model confusion matrices under different modes.

After analyzing the confusion matrices of the attention residual model for intelligent dual-mode fusion in Figure 12 and Table 3, the following conclusions can be drawn: the model demonstrated a high accuracy across most fault types, indicating its strong classification ability. Specifically, there was a 100% accuracy for the fault types, such as end ring cracking 30 Hz, broken rotor bar 30 Hz, health 30 Hz, bearing failure 30 Hz, and bearing failure 40 Hz. This indicated that the model's predictions for these fault types are highly accurate with no omissions. Similarly, for fault types, such as end ring cracking 40 Hz, broken rotor bar 40 Hz, health 40 Hz, and turn-to-turn short circuit 30 Hz, the model achieved a high accuracy, approaching or equaling 100%. This signified that the model can effectively identify these fault types and provide accurate predictions. However, in the turn-to-turn short circuit 40 Hz category, the model exhibited slightly lower accuracy, approximately 98.75%, with 3 recognition errors. Interestingly, all three errors involved misclassifying the turn-to-turn short circuit 40 Hz as the broken rotor bar 40 Hz. This suggested the presence of some false positives or omissions within this category. Further research and improvement can be carried out to enhance the model's performance and accuracy in this specific category. Overall, the accuracy evaluation revealed that the model possesses a strong classification ability for most fault types, although there is room for improvement in specific categories.

Table 3. The Accuracy of the Intelligent Dual-Mode Fusion Model in Each State.

Category	Accuracy	Category	Accuracy
End ring cracking 30 Hz	100%	Health 40 Hz	99.58%
End ring cracking 40 Hz	99.58%	Turn-to-turn short circuit 30 Hz	100%
Broken Rotor Bar 30 Hz	100%	Turn-to-turn short circuit 30 Hz	98.75%
Broken Rotor Bar 40 Hz	99.17%	Bearing failure 30 Hz	100%
Health 30 Hz	100%	Bearing failure 40 Hz	100%

These analytical findings hold great significance in evaluating the model's performance and guiding future research.

In Figure 12, the CNN, ResNet, ResNeXt, and ISE–ResNeXt models were utilized to detect faults in motors. Each model underwent single-mode and modal experiments under the IEEMD denoising conditions. The corresponding accuracies are presented in Table 4. Under dual-mode conditions, the fault diagnosis accuracies for the different models were 91.13%, 95.96%, 99.58%, and 99.71%, respectively. These values represented improvements of 19.75%, 21.17%, 10.71%, and 10.79% over the accuracies achieved by the CNN, Resnet, ResNeXt, and ISE–ResNeXt models under single-mode conditions, respectively. Notably, the CNN and Resnet exhibited higher enhancements in fault diagnosis rates compared to the ResNeXt and ISE–ResNeXt. This can be attributed to the stronger fault feature power and higher accuracy of the latter models. Additionally, due to the initially low fault rate of the CNN under single-mode conditions, the improvement observed in its fault diagnosis rate was relatively larger. Intuitively, it was evident that the fault identification accuracy was higher in dual-mode experiments across different models. This improvement was attributed to the bimodal Gramian Angle Field plot, which combined the characteristics of current signals and vibration signals. This method aggregated the maximum feature matrix during the fault feature extraction process, thereby increasing the amount of information and effectively improving the fault recognition accuracy of motors.

Table 4. The accuracy of different models.

Mode	Model	Accuracy	Mode	Model	Accuracy
Single mode	CNN	71.38%	Bimodal	CNN	91.13%
	Resnet	74.79%		Resnet	95.96%
	ResNeXt	88.87%		ResNeXt	99.58%
	ISE-ResNeXt	88.92%		ISE-ResNeXt	99.71%

5. Conclusions

In this article, the researchers have presented an IEEMD–GASF–ISE–ResNeXt approach for fault diagnosis of motors under varying operating and noise conditions. The proposed method combines the EEMD and MSAM to form an IEEMD denoising technique targeting high frequencies. The vibration and current signals were separately encoded into Gram matrices, which were then fused to create a bimodal Gramian Angle Field graph. Finally, the SE attention mechanism with a switch activation function, and the ISE–ResNeXt with a Dropout layer, were employed for the state identification and fault diagnosis of motors under diverse working and noise conditions.

The main conclusions of this study were as follows:

1. For the single-mode motor dataset, the accuracy rates of un-noised ResNeXt, denoised ResNeXt, and ISE–ResNeXt were 84.17%, 84.29%, and 88.92%, respectively. The IEEMD denoising method improved the fault recognition rate by 4.63% under the same model, leading to enhanced recognition rates for various faults. This demonstrated the effectiveness of the signal processing method that combined the EEMD–MSAM and wavelet packet threshold denoising in suppressing noise and improving accuracy. The utilization of the EEMD–MSAM enabled the signal to be separated into high and low frequency components, while the wavelet packet threshold denoising specifically targeted the high frequency signal for further processing. Consequently, the denoised signal exhibited an enhanced quality, facilitating more precise fault diagnosis and analysis. The integration of these techniques contributed to the overall improvement of the system’s performance by reducing noise interference and enhancing diagnostic accuracy.
2. Under denoising conditions, the respective fault diagnosis accuracy values for the dual-mode motor dataset were 91.13% (CNN), 95.96% (Resnet), 99.58% (ResNeXt), and 99.71% (ISE–ResNeXt). These values were 19.75%, 21.17%, 10.71%, and 10.79% higher than those of the CNN, Resnet, ResNeXt, and ISE–ResNeXt models under single-mode conditions, respectively. Furthermore, there was no significant difference

in training time based on the achieved recognition accuracy. This improvement in fault diagnosis accuracy can be attributed to the proposed IEEMD–GASF–ISE–ResNeXt approach, which effectively suppressed noise influence, enhanced feature extraction capabilities, and improved the fault diagnosis accuracy of motors.

3. The proposed intelligent bimodal fusion attention residual model proved effective in identifying motor datasets and met practical engineering fault diagnosis requirements. Future research will focus on further enhancing the fault diagnosis accuracy of the model by incorporating prior engineering knowledge with deep learning and exploring its applicability to small sample scenarios.

Author Contributions: Conceptualization, F.X. and G.L.; methodology, F.X. and W.H.; validation, F.X. and S.Z.; investigation, F.X. and Q.F.; writing—original draft preparation, F.X.; writing—re-view and editing, F.X. and G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (52265068, 52065022), the Natural Science Foundation of Jiangxi Province (20224BAB204050, 20224BAB204040), the Project of Jiangxi Provincial Department of Education (GJJ2200627), and the Jiangxi Provincial Graduate Innovation Special Fund Project (YC2022-s481).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ou, H.; Hu, Y.; Mao, Z.; Li, Y. A Method for Reducing Cogging Torque of Integrated Propulsion Motor. *J. Mar. Sci. Eng.* **2019**, *7*, 236. [CrossRef]
2. Rao, H.; Wang, N.; Du, R. Vibration Cascade Control for Motor-Driven Deep-Sea Robot Cable System with Actuator Fault. *J. Mar. Sci. Eng.* **2022**, *10*, 1772. [CrossRef]
3. The Swedish Club. Main Engine Damage Study. 2015. Available online: <https://www.swedishclub.com/mediaupload/files/Publications/Loss%20Prevention/Main%20Engine%20damage%202015%20The%20Swedish%20Club.pdf> (accessed on 9 April 2021).
4. Xu, X.; Yan, X.; Yang, K.; Zhao, J.; Sheng, C.; Yuan, C. Review of condition monitoring and fault diagnosis for marine power systems. *Transp. Saf. Environ.* **2021**, *3*, 85–102. [CrossRef]
5. Wang, S.; Zhang, Y.; Zhang, B.; Fei, Y.; He, Y.; Li, P.; Xu, M. On the Sparse Gradient Denoising Optimization of Neural Network Models for Rolling Bearing Fault Diagnosis Illustrated by a Ship Propulsion System. *J. Mar. Sci. Eng.* **2022**, *10*, 1376. [CrossRef]
6. Sun, H.; Zi, Y.; He, Z. Wind turbine fault detection using multiwavelet denoising with the data-driven block threshold. *Appl. Acoust.* **2014**, *77*, 122–129. [CrossRef]
7. Patange, A.D.; Pardeshi, S.S.; Jegadeeshwaran, R.; Zarkar, A.; Verma, K. Augmentation of decision tree model through hyper-parameters tuning for monitoring of cutting tool faults based on vibration signature. *J. Vib. Eng. Technol.* **2022**, 1–19. [CrossRef]
8. Sharma, S.; Tiwari, S.K.; Singh, S. Integrated approach based on flexible analytical wavelet transform and permutation entropy for fault detection in rotary machines. *Measurement* **2021**, *169*, 108389. [CrossRef]
9. Sharma, S.; Tiwari, S.K. A novel feature extraction method based on weighted multi-scale fluctuation based dispersion entropy and its application to the condition monitoring of rotary machines. *Mech. Syst. Signal Process.* **2022**, *171*, 108909. [CrossRef]
10. Ba, J.; Mnih, V.; Kavukcuoglu, K. Multiple object recognition with visual attention. *arXiv* **2014**, arXiv:1412.7755.
11. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
12. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19.
13. Zhao, T.; Wu, X.Q. Pyramid feature attention network for saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; IEEE Computer Society Press: Los Alamitos, CA, USA, 2019; pp. 3085–3094.
14. Zhang, X.J.; Shang, J.Y.; Yu, G.J.; Hao, J. Attention based multi-scale convolutional neural network bearing fault diagnosis. *J. Jilin Univ. (Eng. Ed.)* **2023**, 1–10.
15. Han, Y.; Li, C.; Huang, Q.Q.; Wen, R.; Zhang, Y. A gearbox fault diagnosis method based on temporal attention boundary enhancement prototype network under small samples. *J. Electron. Meas. Instrum.* **2023**, 1–8.
16. Shi, J.W.; Hou, L.Q. Bearing fault diagnosis based on one-dimensional convolutional attention gated cyclic network and transfer learning. *Shock. Vib.* **2023**, *42*, 159–164+173.

17. Sharma, S.; Tiwari, S.K. Residual signal-based condition monitoring of planetary gearbox using electrical signature analysis. *J. Vib. Control* **2023**, 10775463231178070. [\[CrossRef\]](#)
18. Wang, J.; Fu, P.; Zhang, L.; Gao, R.X.; Zhao, R. Multilevel information fusion for induction motor fault diagnosis. *IEEE/ASME Trans. Mechatron.* **2019**, *24*, 2139–2150. [\[CrossRef\]](#)
19. Zhang, T.; Li, Z.; Deng, Z.; Hu, B. Hybrid data fusion DBN for intelligent fault diagnosis of vehicle reducers. *Sensors* **2019**, *19*, 2504. [\[CrossRef\]](#)
20. Hoang, D.T.; Kang, H.J. A motor current signal-based bearing fault diagnosis using deep learning and information fusion. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 3325–3333. [\[CrossRef\]](#)
21. Jian, X.; Li, W.; Guo, X.; Wang, R. Fault diagnosis of motor bearings based on a one-dimensional fusion neural network. *Sensors* **2019**, *19*, 122. [\[CrossRef\]](#)
22. Zhao, X.; Jia, M.; Ding, P.; Yang, C.; She, D.; Liu, Z. Intelligent fault diagnosis of multichannel motor-rotor system based on multimanifold deep extreme learning machine. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2177–2187. [\[CrossRef\]](#)
23. Patange, A.D.; Jegadeeshwaran, R. A machine learning approach for vibration-based multipoint tool insert health prediction on vertical machining centre (VMC). *Measurement* **2021**, *173*, 108649. [\[CrossRef\]](#)
24. Kanwisher, N.; Gupta, P.; Dobs, K. CNNs Reveal the Computational Implausibility of the Expertise Hypothesis. *iScience* **2023**, *26*, 105976. [\[CrossRef\]](#)
25. Lin, M.; Cao, L.; Zhang, Y.; Shao, L.; Lin, C.-W.; Ji, R. Pruning networks with cross-layer ranking & k-reciprocal nearest filters. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–10.
26. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5990–5998. [\[CrossRef\]](#)
27. Chen, Z.Q.; Li, C.; Sanchez, R.V. Gearbox fault identification and classification with convolutional neural networks. *Shock. Vib.* **2015**, *2015*, 390134. [\[CrossRef\]](#)
28. Chen, P.; Li, Y.; Wang, K.; Zuo, M.J. An automatic speed adaption neural network model for planetary gearbox fault diagnosis. *Measurement* **2021**, *171*, 108784. [\[CrossRef\]](#)
29. Patil, S.S.; Pardeshi, S.S.; Patange, A.D. Patange. Health Monitoring of Milling Tool Inserts Using CNN Architectures Trained by Vibration Spectrograms. *CMES-Comput. Model. Eng. Sci.* **2023**, *136*, 177–199.
30. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Dong, S.; Pecht, M. Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis. *IEEE Trans. Ind. Electron.* **2020**, *68*, 2587–2597. [\[CrossRef\]](#)
31. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [\[CrossRef\]](#)
32. Chao, L.; Feng, Z.; Yan, L. GPS/Pseudolites technology based on EMD-wavelet in the complex field conditions of mine. *Procedia Earth Planet. Sci.* **2009**, *1*, 1293–1300. [\[CrossRef\]](#)
33. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In Proceedings of the 2017 International joint conference on neural networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1578–1585.
34. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941.
35. Zhang, H.; Zhang, Q.; Yu, J.Y. Review of the development of activation functions and its property analysis. *J. Xihua Univ. (Nat. Sci. Ed.)* **2021**, *40*, e16125.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.