applied sciences

MDPI

Article

Intelligent Fault Diagnosis for Rotating Machinery Utilizing Gramian Angular Field-Parallel Convolutional Neural Network and Gated Recurrent Unit Networks

Yuxiang Hu [1], Shengyi Cheng [1,2,*] and Xianjun Du [1,2,*]

[1] School of Automation and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China; lutyxhu@163.com
[2] School of Microelectronics Industry-Education Integration, Lanzhou University of Technology, Lanzhou 730050, China
* Correspondence: mayi_iris@163.com (S.C.); xdu@lut.edu.cn (X.D.)

Abstract

To address the limitations of traditional fault diagnosis methods for rotating machinery, which heavily rely on single-dimensional vibration data and fail to fully exploit the deep features of time-series data, this study proposes an innovative diagnostic model that integrates Gramian Angular Field-Parallel Convolutional Neural Network (GAF-PCNN) with Gated Recurrent Units (GRU). Specifically, one-dimensional vibration signals are first transformed into Gramian angular and difference fields as image representations using Gramian Angular Field (GAF). These two types of images are then input into parallel-configured PCNN modules for feature learning. The features extracted by the two CNN branches are weighted and fused to construct a combined feature sequence. This sequence is subsequently fed into the GRU network to capture temporal dependencies and perform deep feature extraction. In this process, an integrated self-attention mechanism is applied to dynamically select key features. The proposed method is validated using two publicly available datasets, including comparative and noise interference experiments. The results demonstrate that the proposed model excels in diagnostic accuracy, model generalization, and robustness against noise interference.

Keywords: rotating machinery; Gramian angular field; parallel convolutional neural network

check for updates

Received: 16 July 2025
Revised: 9 August 2025
Accepted: 20 August 2025
Published: 21 August 2025

Citation: Hu, Y.; Cheng, S.; Du, X. Intelligent Fault Diagnosis for Rotating Machinery Utilizing Gramian Angular Field-Parallel Convolutional Neural Network and Gated Recurrent Unit Networks. *Appl. Sci.* **2025**, *15*, 9217. https://doi.org/10.3390/app15169217

Copyright: © 2025 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

Rotating machinery constitutes critical infrastructure across industrial systems, widely employed in sectors such as energy production, manufacturing, and transportation [1–3]. However, due to prolonged high-load operation, harsh working environments, and structural aging, various faults—such as bearing damage, gear tooth breakage, and rotor eccentricity—are likely to occur. If not diagnosed promptly, these faults can lead to equipment shutdowns, production halts, or even safety incidents, posing significant threats to system stability and operational safety. Consequently, the development of efficient, accurate, and intelligent fault diagnosis techniques for rotating machinery has emerged as a prominent research focus in engineering and smart manufacturing domains [4–6].

The evolution of fault diagnosis techniques can be broadly categorized into three key stages: model-based approaches [7], knowledge-based methods [8], and, more recently, data-driven techniques [9,10]. The first stage relies on physical modeling of mechanical

systems, using formulations such as dynamic equations, finite element models, or transfer functions. While these approaches offer strong theoretical interpretability, their practical implementation is often hindered by the complex, nonlinear, and strongly coupled nature of rotating machinery, along with multi-source disturbances. These factors make accurate modeling challenging and limit the generalization capabilities of such models.

The second stage involves expert systems that integrate domain knowledge with rule-based reasoning, fuzzy logic, and manual inference to identify fault types. These methods offer a certain degree of practicality and interpretability. However, with the increasing complexity of industrial operations, such approaches face significant challenges, including slow knowledge updates, difficulty in model migration, and high maintenance costs due to their reliance on human expertise.

In recent years, the third stage—data-driven methods—has rapidly gained prominence, fueled by advancements in artificial intelligence, sensor technology, and big data analytics. These approaches eliminate the need for explicit physical modeling or manual feature engineering by learning directly from historical operational data. They offer notable advantages, such as modeling flexibility, adaptability, and scalability. Data-driven methods encompass both traditional machine learning algorithms—such as support vector machines (SVM), k-nearest neighbors (KNN), and decision trees—and deep learning models, including convolutional neural networks (CNN), deep belief network (DBN), long short-term memory (LSTM) networks, and gated recurrent units (GRU). Among these, deep learning has emerged as a dominant paradigm due to its powerful capabilities in automatic feature extraction and nonlinear representation learning. Extensive studies have demonstrated that deep learning models can effectively extract high-dimensional features directly from raw vibration signals, achieving excellent classification performance with minimal preprocessing. These methods utilize big data analytics and machine learning techniques, achieving breakthroughs not only in accuracy, real-time capability, and applicability but also demonstrating immense potential. Zhao et al. [11] developed a hybrid information convolutional architecture and used this model to extract features with more distinct rows and high recognition. Their superior performance was verified through a series of experiments. Ruan et al. [12] utilized the fault cycle and mechanical frequency (or shaft rotation frequency) information under different fault conditions to determine the input size for CNNs. Compared to traditional CNN models, this design improves diagnostic accuracy while reducing uncertainty. Additionally, Qu et al. [13] proposed a fault diagnosis method based on an improved DBN, which was validated using an authoritative open-source dataset. Zhou et al. [14] developed a two-stage joint denoising-driven intelligent diagnostic framework: first, coupling empirical mode decomposition with independent component analysis, followed by the use of stacked GRU as a deep classifier for fault classification. Kumar et al. [15] enhanced fault signals and applied long short-term memory (LSTM) networks combined with GRU for gear fault diagnosis. Their results show that this combined model offers superior computational efficiency, with a diagnostic accuracy of 99.6%. Moreover, the integration of attention mechanisms enables more flexible allocation of feature weights, allowing the model to emphasize critical features while suppressing redundant information [16].

Nevertheless, several challenges remain in existing methods. On one hand, the low-dimensional nature of raw one-dimensional (1D) vibration signals limits their capacity to represent complex fault patterns, particularly under strong noise or during early fault stages. On the other hand, single-model architectures often struggle to simultaneously capture both local and global patterns and generally exhibit limited robustness to varying operational conditions and signal interference.

To address these limitations, recent research has increasingly focused on feature reconstruction and multi-model fusion strategies. By transforming 1D vibration signals into two-dimensional (2D) representations—such as time–frequency images, phase space plots, and Gramian matrices—the expressive power of the input data can be significantly enhanced. In parallel, integrating diverse network architectures enables joint modeling of spatial, temporal, and semantic features. This hybrid modeling approach facilitates deeper fault pattern discovery and improves both classification accuracy and generalization.

Based on the above considerations, this study develops a hybrid framework for rotating machinery fault diagnosis, termed GAF-PCNN-GRU. Unlike existing approaches that typically combine a single-modality Gramian Angular Field (GAF) representation with a single CNN or a simple CNN-GRU structure, the proposed method innovatively integrates dual-modality GAF representations (GASF and GADF) with a parallel CNN (PCNN) architecture to achieve complementary spatial feature extraction. The extracted features are then modeled using a Gated Recurrent Unit (GRU) to capture temporal dependencies and further refined through a multi-head self-attention mechanism to emphasize critical fault-related features. This methodological design not only enhances feature diversity and strengthens spatial–temporal correlation modeling but also significantly improves robustness under noisy and variable operating conditions. Moreover, by organically combining multimodal feature fusion, temporal modeling, and attention mechanisms into a unified framework, the method achieves high accuracy while maintaining strong generalizability and scalability. It is not only applicable to rotating machinery fault diagnosis but also provides a generalizable paradigm for other time-series signal analysis tasks.

The main contributions of this work are summarized as follows:

(1) The GAF method is employed to convert raw one-dimensional vibration signals into two types of two-dimensional image representations: GASF and GADF, thereby constructing a dual-modality data input. This strategy fully exploits the relative phase and amplitude information embedded in the time-series signals, significantly enhancing the representational capacity and diversity of extracted features. The resulting high-dimensional complementary features effectively address the limitations of traditional approaches in capturing complex fault characteristics.

(2) A novel dual-branch parallel convolutional structure is designed at the front end of the network, in which GASF and GADF images are independently processed. This PCNN architecture enables the extraction of distinct spatial features from both modalities. Through hierarchical convolution operations, the network captures high-level semantic features such as edges, textures, and shapes, thereby improving the model's ability to perceive complex mechanical failures.

(3) The fused features obtained from PCNN are reshaped and fed into a GRU to model temporal dependencies. Additionally, a multi-head self-attention mechanism is incorporated to further identify critical segments within the time series, enhancing the model's sensitivity to localized and subtle fault patterns. This design significantly improves the efficiency of feature representation and the model's discriminative capability.

(4) Extensive experiments, including baseline comparisons, cross-dataset evaluations, and noise injection tests, are conducted on two benchmark datasets (CWRU and UC). The results demonstrate that the proposed GAF-PCNN-GRU model consistently achieves high diagnostic accuracy and robustness under various operating conditions and fault types, confirming its strong adaptability and practical potential for real-world industrial deployment.

Paper organization: Section 1 outlines the overall methodology proposed. Section 2 discusses three foundational models: GAF, CNN, and GRU. Section 3 constructs the fault diagnosis model. Section 4 presents model simulations and analyses. Section 6 concludes with discussions and future directions.

2. Overall Methodological Approach

Most existing fault diagnosis studies for rotating machinery adopt a strategy in which 1D vibration signals are first transformed into 2D images and then used as input to CNNs for feature extraction based on local textures or spectral representations. However, these approaches exhibit two main limitations: (1) the image representation is often based on a single-view transformation, resulting in limited utilization of signal information; (2) the reliance solely on CNNs neglects their inherent limitations in modeling temporal dependencies.

To address these challenges, a hybrid diagnostic architecture combining GAF-PCNN and GRU is proposed in this study. The method innovatively transforms one-dimensional vibration signals into two types of complementary two-dimensional image representations and integrates spatial feature extraction with temporal modeling to enhance both the accuracy and robustness of the diagnostic process.

The overall architecture of the proposed model is illustrated in Figure 1 and consists of four core functional stages. Initially, the raw vibration signal is segmented into fixed-length sub-sequences and converted into two different Gramian-based image types: GASF, which reflects the signal's amplitude trend, and GADF, which captures the relative relationships in signal variations. This dual-channel image representation effectively addresses the insufficient expressiveness of single-image approaches by enriching the data's semantic and structural content.

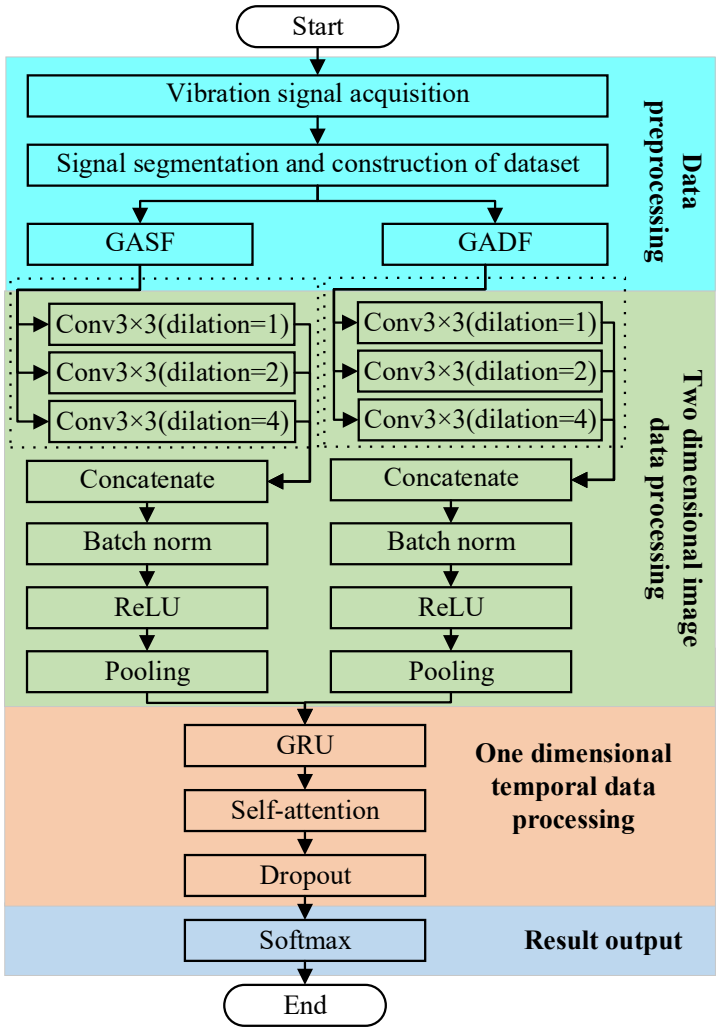


Figure 1. Overall scheme of the proposed model.

Next, the GASF and GADF images are separately fed into two parallel CNN branches with shared weights, which are responsible for extracting high-level spatial features. A layer-by-layer convolution and pooling structure is employed to improve the model's ability to perceive complex local textures, frequency shifts, and structural patterns embedded in the images.

After spatial features are extracted from the two branches, they are fused and reshaped into sequential vectors, which are then input into a GRU network to capture temporal dependencies. A multi-head self-attention mechanism is applied to the GRU outputs to dynamically reweigh the sequence features, thereby enhancing the representation of key patterns associated with local fault sensitivity. In addition, Dropout is employed to suppress overfitting.

Finally, a Softmax classifier is used to predict the fault category based on the refined feature vector. This architecture effectively integrates enhanced image representation, spatial feature extraction, temporal modeling, and attention-based feature refinement, resulting in a highly accurate, generalizable, and noise-resilient solution for complex fault diagnosis tasks.

3. Introduction to the Baseline Model

3.1. Gramian Angular Field

GAF [17] encodes time series into images while preserving temporal correlations through polar coordinate transformation, originally proposed for time-series classification using computer vision. The specific process is as follows:

Standardize X using Equation (1) so that each value is hidden within the same spatial scale:

$$\widetilde{x}_i = 2 \times \frac{X - \min(X)}{\max(X) - \min(X)} - 1 \tag{1}$$

The standardized X is denoted as $\widetilde{X}$, which is transformed into vectors through polar coordinate conversion of the time-series signal.

$$\begin{cases} \phi_i = \arccos(\widetilde{x}_i),\ -1 \le \widetilde{x}_i \le 1,\ \widetilde{x}_i \in \widetilde{X} \\ r_i = \frac{i}{n},\ i = 1, 2, \ldots, n \end{cases} \tag{2}$$

In Equation (2), ϕ_i represents the polar angle and r_i represents the polar radius of i.

Correlations across varying time intervals are characterized by trigonometric summation or difference operations between pairwise points, the specific visualization is depicted in Figure 2. GASF [18] and GADF [19] are respectively defined as follows:

$$G_s = \begin{bmatrix} \cos(\vartheta_1 + \vartheta_1) & \cos(\vartheta_1 + \vartheta_2) & \ldots \cos(\vartheta_1 + \vartheta_n) \\ \cos(\vartheta_2 + \vartheta_1) & \cos(\vartheta_2 + \vartheta_2) & \ldots \cos(\vartheta_2 + \vartheta_n) \\ \vdots & \vdots & \vdots \quad \vdots \\ \cos(\vartheta_n + \vartheta_1) & \cos(\vartheta_n + \vartheta_2) & \ldots \cos(\vartheta_n + \vartheta_n) \end{bmatrix} \tag{3}$$

$$G_d = \begin{bmatrix} \sin(\vartheta_1 - \vartheta_1) & \sin(\vartheta_1 - \vartheta_2) & \ldots \sin(\vartheta_1 - \vartheta_n) \\ \sin(\vartheta_2 - \vartheta_1) & \sin(\vartheta_2 - \vartheta_2) & \ldots \sin(\vartheta_2 - \vartheta_n) \\ \vdots & \vdots & \vdots \quad \vdots \\ \sin(\vartheta_n - \vartheta_1) & \sin(\vartheta_n - \vartheta_2) & \ldots \sin(\vartheta_n - \vartheta_n) \end{bmatrix} \tag{4}$$

where ϑ_i denotes the included angle between two vectors.

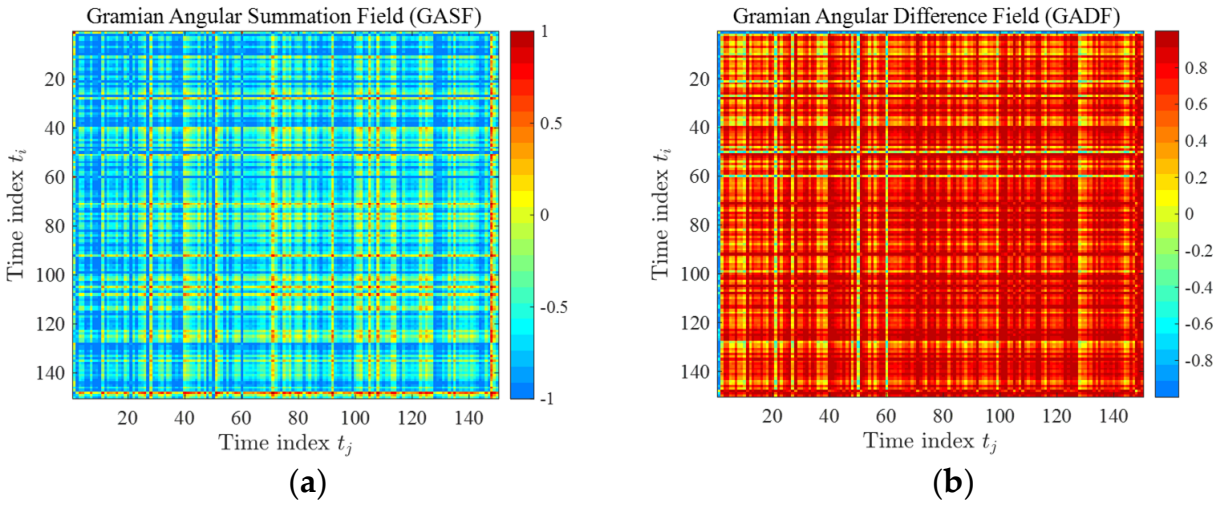


Figure 2. Gramian angular field image. (**a**) GASF; (**b**) GADF.

3.2. Convolutional Neural Network

CNN [20,21] employs sparse connections and weight-sharing strategies to locally perceive data through sliding windows, effectively extracting intricate local features. The basic structure of CNN, as illustrated in Figure 3, comprises input layers, convolutional layers, pooling layers, and fully connected layers. Convolutional layers capture local patterns and features of input data through filter convolution operations, while pooling layers reduce the dimensionality of data via down sampling to decrease computational load and retain significant features. Finally, global feature vectors undergo matrix transformation before propagation through fully connected layers for simultaneous classification and regression.

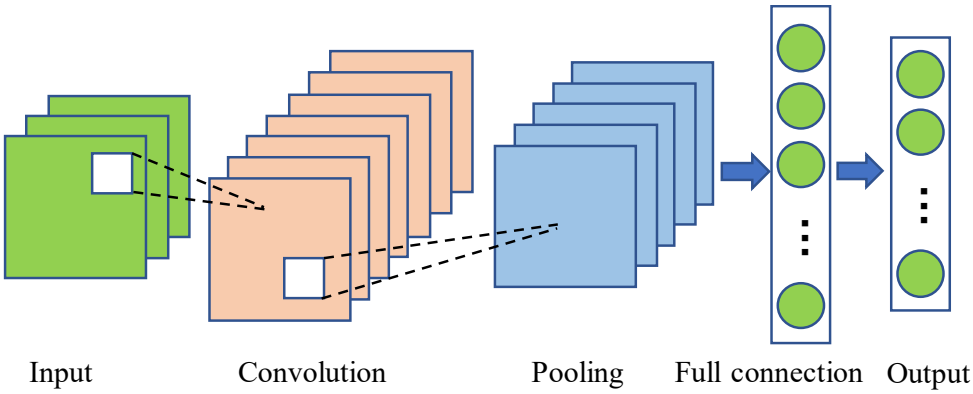


Figure 3. Diagram of the CNN structure.

CNN has achieved significant success in multiple industrial fields. To leverage the characteristics of time-series fault data, this work employs GAF transformation to convert such data into images, thereby harnessing CNN's powerful feature extraction capability to capture intrinsic patterns in fault signals.

3.3. Gated Recurrent Unit

GRU [22,23], a variant of LSTM [24], simplifies the structure while maintaining the effectiveness of LSTM, and offers faster computation and training speeds. As depicted in Figure 4, the internal structure and operational workflow of the GRU unit consist primarily of two components.

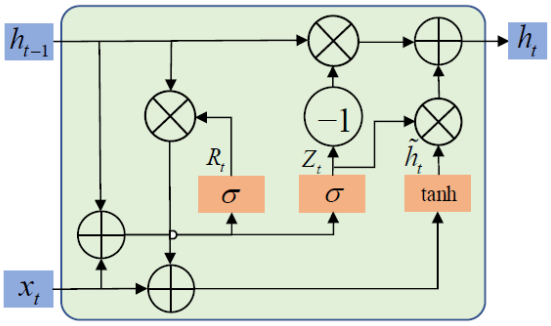


Figure 4. GRU detailed structure.

The computation of the GRU hidden layer is given by Equation (5):

$$\begin{cases} Z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \\ R_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \\ \widetilde{h}_t = \tanh(W \cdot [R_t h_{t-1}, x_t]) \\ h_t = (1 - Z_t) h_{t-1} + Z_t \widetilde{h}_t \end{cases} \tag{5}$$

In the equation, Z_t is the update gate with weight matrix W_z; x_t is the input at timestep t; σ is a sigmoid activation function; R_t is the reset gate with weight matrix W_r; W is a shared weight matrix; h_{t-1} is the hidden state at $t-1$; $\widetilde{h}_t$ is the candidate hidden state; and h_t is the output hidden state propagated to next timestep.

4. Construction of Fault Diagnosis Model

For the complex task of fault classification and recognition in rotating machinery, traditional diagnostic methods often suffer from two major limitations. First, directly inputting one-dimensional vibration signals into classification models makes it difficult to extract deep temporal features effectively. Second, while some approaches convert one-dimensional data into two-dimensional images for CNN-based processing, they typically extract only local spatial features and fail to capture the temporal dependencies inherent in vibration signals. To address these issues, a hybrid fault diagnosis model based on GAF-PCNN-GRU is proposed, as illustrated in Figure 5.

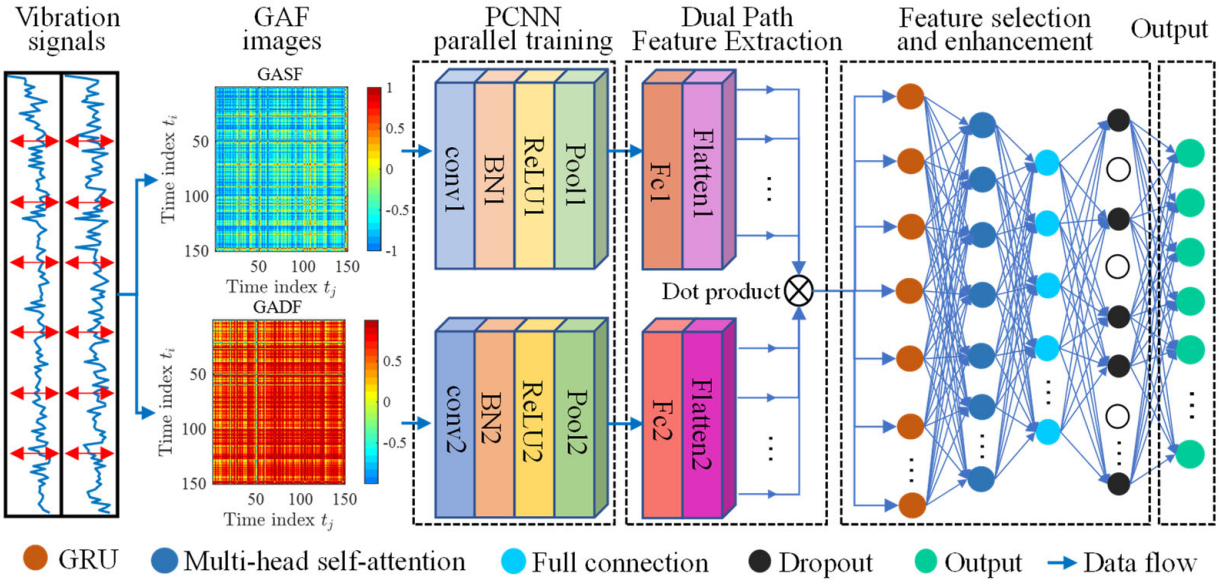


Figure 5. Schematic representation of the novel diagnostic model architecture.

Initially, the original vibration signals are transformed into two types of GAF images—GASF and GADF—which serve as inputs to two parallel CNN branches within the PCNN structure. The use of CNNs for image processing enables the effective extraction of high-level spatial features, such as edges, textures, and local structures. Through multiple layers of convolution and pooling, the receptive field is expanded while the feature dimensionality is significantly reduced, thereby lowering computational complexity and mitigating the risk of overfitting. Subsequently, the feature maps extracted from the two CNN branches are fused via element-wise multiplication to obtain a unified feature tensor.

Considering that CNNs primarily focus on spatial domain features, a GRU structure is introduced to model the sequential dependencies within the fused features. As an improved variant of the RNN, GRU effectively captures temporal patterns in vibration signals while avoiding the vanishing gradient problem. By reshaping the fused CNN features into a time-series format and feeding them into the GRU, joint modeling of spatial and temporal features is achieved.

To further enhance the model's ability to focus on critical temporal segments, a multi-head self-attention mechanism is introduced after the GRU to optimize the weighting of output features. Specifically, the hidden state sequence output by the GRU is linearly transformed into query (Q), key (K), and value (V) vectors. These transformations are divided into H attention heads, where for the k-th head, the linear transformations of the query, key, and value are as follows:

$$Q_t^{(k)} = W_Q^{(k)} h_t + b_Q^{(k)} \tag{6}$$

$$K_t^{(k)} = W_K^{(k)} h_t + b_K^{(k)} \tag{7}$$

$$V_t^{(k)} = W_V^{(k)} h_t + b_V^{(k)} \tag{8}$$

where W_Q, W_K, W_V denote the learnable weight matrix and b_Q, b_K, b_V represent the bias term.

Compute the attention output for each head:

$$\alpha_{t,i} = \text{softmax}(Attnetion_{t,i}^{Score}) = \frac{e^{Attnetion_{t,i}^{Score}}}{\sum_{j=1}^{T} e^{Attnetion_{t,i}^{Score}}}, Attnetion_{t,i}^{Score} = \frac{Q_t \cdot K_i}{\sqrt{d_k}} \tag{9}$$

$$Output_t^{(k)} = \sum_{i=1}^{T} \alpha_{t,i}^{(k)} V_i^{(k)} \tag{10}$$

where Q_t represents the query vector at time step t, K_i is the key vector, d_k denotes the dimensionality of the key vector, and $\alpha_{t,i}$ is the attention weight of time step t with respect to position i.

The outputs from all heads are then concatenated and passed through a linear transformation to obtain the final multi-head self-attention output.

$$Output_t^{Multi-Head} = Concat(Output_t^{(1)}, \cdots, Output_t^{(k)}) W_O + b_O \tag{11}$$

The integration of the multi-head self-attention mechanism emphasizes salient features, while a Dropout layer is added to enhance the model's generalization ability, aiming to further improve fault diagnosis accuracy.

5. Simulation and Analysis of the Model

The proposed network is evaluated in a testing environment consisting of Windows 11, an Intel i9-12900H dual-core 2.50 GHz CPU, an NVIDIA RTX 3060 GPU, and 16 GB of RAM. All simulation experiments are conducted using MATLAB 2024b.

To ensure the reliability and practical applicability of the proposed model in engineering scenarios, a dual verification strategy was adopted: (1) the model was trained and tested on two internationally recognized rotating machinery fault datasets, CWRU and UC, to verify its classification performance under various operating conditions and fault types, while multiple mainstream fault diagnosis models were introduced for comparative experiments to comprehensively validate the model's effectiveness; (2) gamma-distributed noise was injected into the UC dataset to simulate signal disturbances encountered in real industrial environments, thereby evaluating the model's robustness and generalization capability under complex background noise. This dual verification process, equivalent to numerical simulation verification at the signal processing level, helps ensure that the proposed method is not only theoretically effective but also highly adaptable in practical engineering applications.

5.1. Description of the Dataset

5.1.1. Case 1

The dataset provided by Case Western Reserve University (CWRU) [25] is used for the first stage of model performance evaluation. As shown in Table 1, the CWRU dataset contains 10 operating conditions: one normal condition, three rolling element faults, three inner race faults, and three outer race faults. Each category consists of 100 samples, resulting in a total of 1000 samples. Each sample contains 2000 data points. All samples are randomly divided into a training set and a testing set at a ratio of 4:1, with 800 samples used for training and 200 for testing. This partitioning ensures the stability and generalization capability of the model during training, parameter tuning, and final evaluation.

Table 1. Bearing failure modes in CWRU dataset.

Label	Fault Type	Fault Location	Fault Diameter (Inch)
1	Norm	-	0
2	B07	Bearing ball	0.007
3	B14	Bearing ball	0.014
4	B21	Bearing ball	0.021
5	IR07	Inner ring	0.007
6	IR14	Inner ring	0.014
7	IR21	Inner ring	0.021
8	OR07	Outer ring	0.007
9	OR14	Outer ring	0.014
10	OR21	Outer ring	0.021

Subsequently, both the training and testing sets are transformed into two-dimensional images using GAF, as illustrated in Figure 6. This transformation first maps the one-dimensional vibration signals from Cartesian coordinates into polar coordinates. Based on a specific inner product formulation, the time correlations are encoded using trigonometric summation and difference operations, and the results are arranged from the top-left to bottom-right to generate the GASF and GADF images. The resulting GAF-encoded images exhibit significant sparsity, which not only mitigates nonlinearity and suppresses noise but also preserves the complete characteristics of the original one-dimensional signals. Additionally, this encoding captures temporal dependencies and latent correlations among the nonstationary and time-varying signal components.

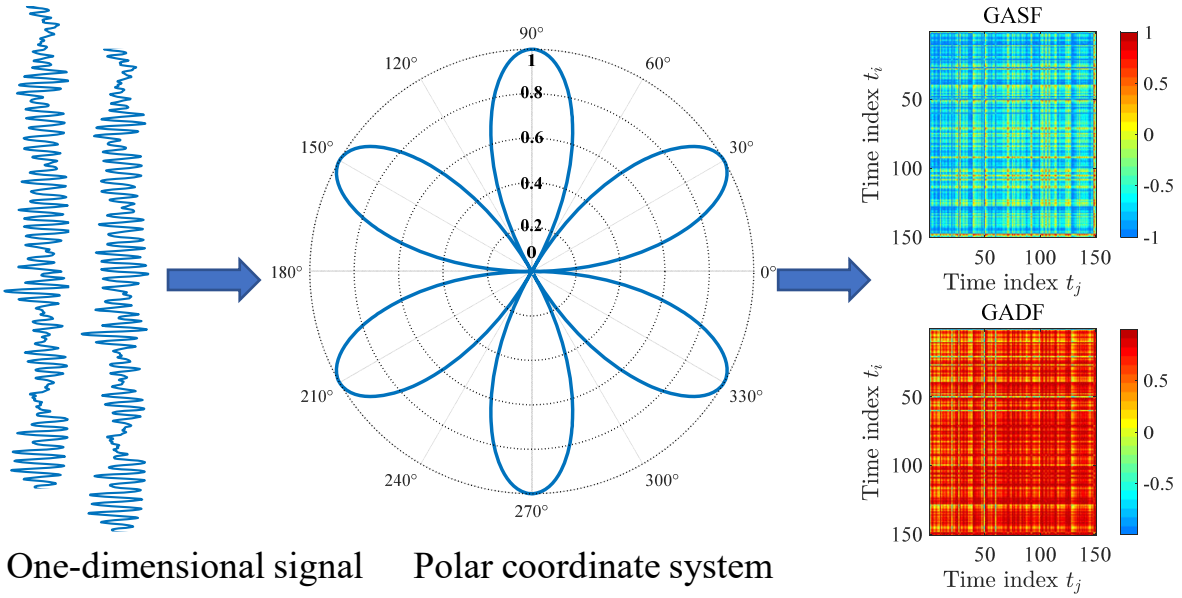


Figure 6. GAF mapping process.

First, the performance of the GAF-PCNN-GRU model is comprehensively evaluated on the CWRU dataset. As shown in Figure 7, the proposed model achieves a diagnostic accuracy of 99.5% on the test set, with only one misclassification among all test samples. The confusion matrix (Figure 8) provides a more detailed analysis of the classification accuracy for each fault category. Except for a single misclassification in label 7, the other nine fault types are all correctly classified with 100% accuracy. These results strongly demonstrate the model's outstanding capability in accurately identifying complex rolling bearing fault patterns in real-world industrial environments.

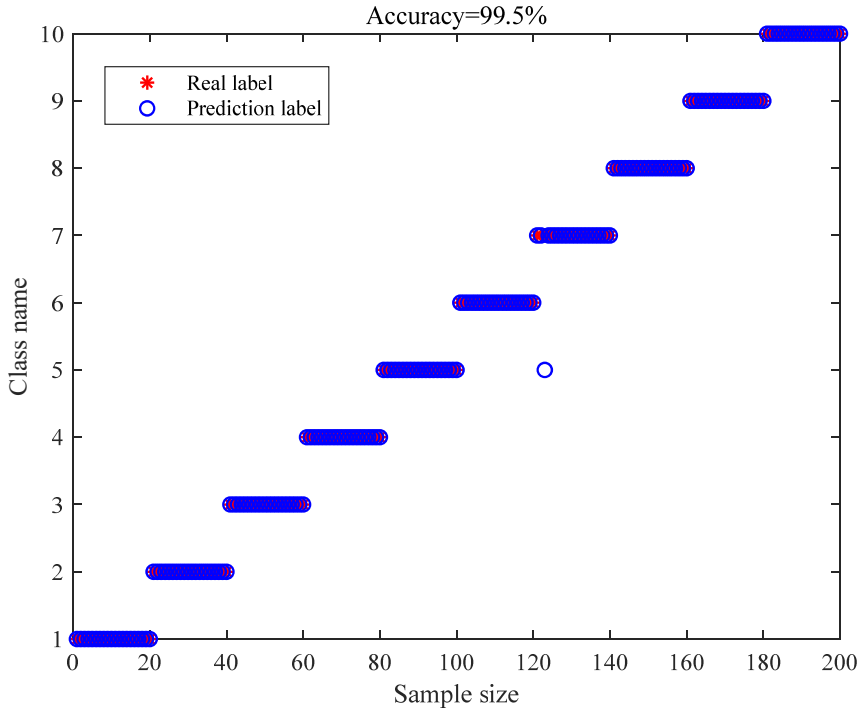


Figure 7. Comparison of predictive results on the CWRU test set.

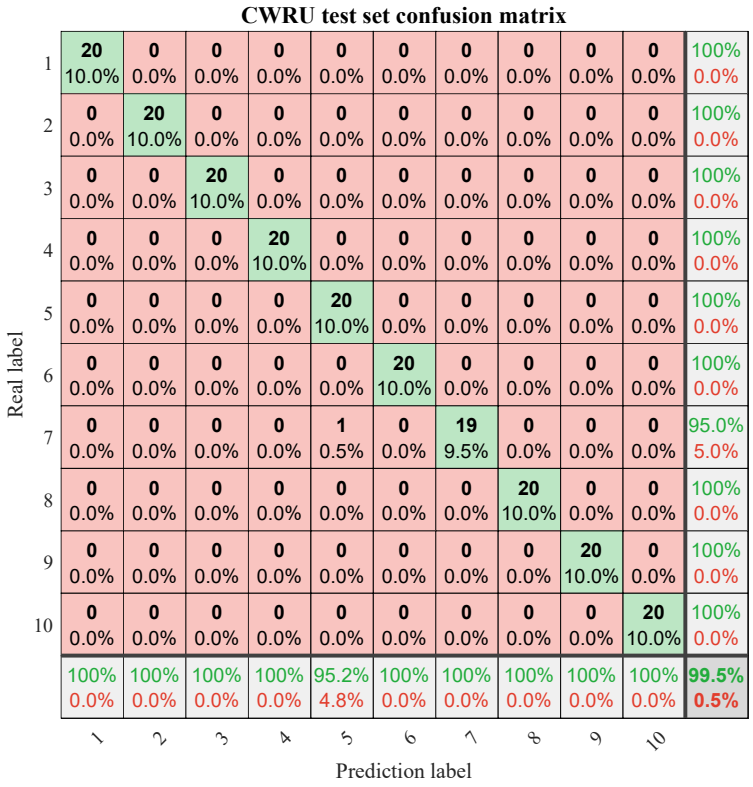


Figure 8. Confusion matrix of the CWRU test set.

5.1.2. Case 2

The gear fault dataset from the University of Connecticut (UC) [26] was collected through experiments conducted on a two-stage gearbox equipped with replaceable gears, with a sampling frequency of 20 kHz. The specific fault types are defined in Table 2. The UC dataset contains a total of 936 samples covering nine distinct fault categories, including healthy (Healthy), missing tooth, root crack, spalling, and tip chipping with five different severity levels. Each category contains 104 samples, and each sample consists of 3600 data points. During the modeling process, the samples are randomly divided into a training set (720 samples) and a test set (216 samples) to ensure a relatively uniform distribution of each fault type across the subsets.

Table 2. Gear failure modes in UC dataset.

Label	Fault Type	Fault Location
1	Undamaged	-
2	Partial Tooth Absence	Tooth
3	Cracked Tooth Foundation	Root
4	Surface Fatigue Spalling	Surface
5	Tip Micro-pitting 5	Gear tip
6	Tip Micro-pitting 4	Gear tip
7	Tip Micro-pitting 3	Gear tip
8	Tip Micro-pitting 2	Gear tip
9	Tip Micro-pitting 1	Gear tip

When utilizing the UC dataset, Figure 9 illustrates the prediction results for the test set, achieving an overall diagnostic accuracy of 99.54%. The proposed model demonstrates superior diagnostic performance on the UC dataset as well. To further analyze the diagnostic outcomes for each fault category, the confusion matrix depicted in Figure 10 reveals that only one sample from label 9 was misclassified into label 6, while the diagnostic accuracy

for all other fault categories reached 100%. These experimental findings collectively demonstrate that the proposed methodology adequately addresses the operational diagnostic requirements for rotating machinery in practical industrial settings.

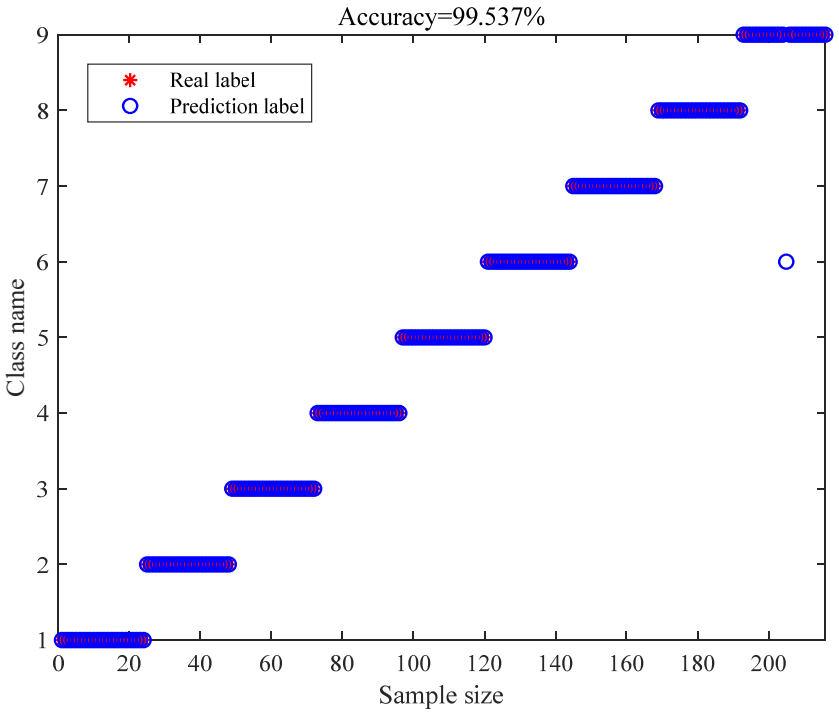


Figure 9. Comparison of predictive results on the UC test set.

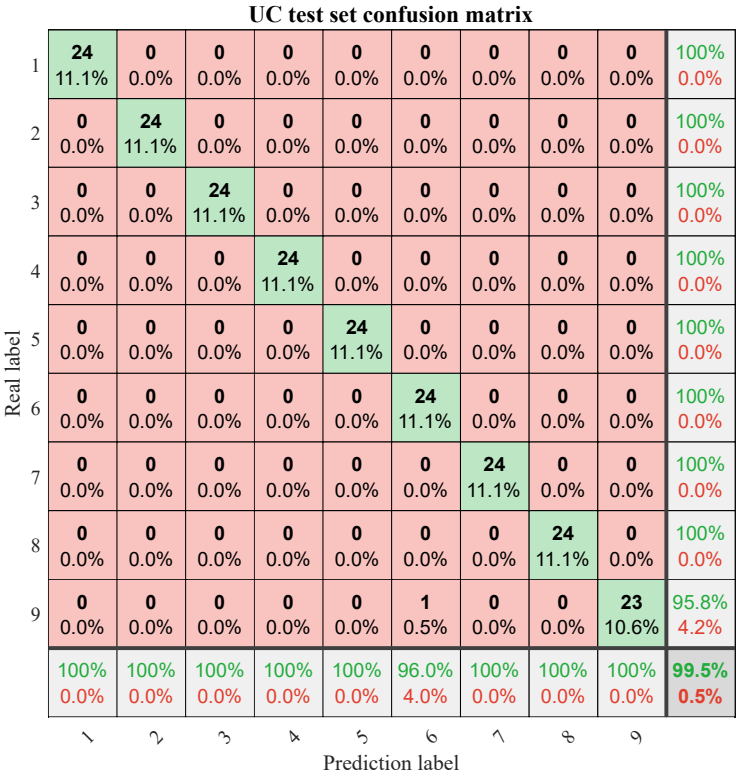


Figure 10. Confusion matrix of the UC test set.

5.2. Comparison Experiments

The proposed model's diagnostic efficacy is benchmarked against GRU, GAF-CNN, GAF-PCNN, and GAF-CNN-GRU architectures. The parameter settings for each comparative model are detailed as follows:

(1) GRU: To maintain uniformity in input data across all comparative networks, images sized $227 \times 227 \times 3$ undergo flatten layer processing before being fed into GRU. The GRU's hidden layer nodes are set to 128.

(2) GAF-CNN: GADF images are selected as input, with 64 filters in the CNN, with filter dimensions of 3×3 and convolutional stride of 2×2.

(3) GAF-PCNN: The first branch of CNN takes GASF images as input, while the second branch takes GADF images. The dual-path architecture employs 64 filters per branch, featuring convolutional kernels of size 3×3 accompanied by a stride length of 2×2.

(4) GAF-CNN-GRU: Similar to GAF-CNN, GADF images are used as input, with 64 filters in the CNN, with filter dimensions of 3×3 and convolutional stride of 2×2.

The CWRU dataset constitutes a standard benchmark in fault diagnosis research due to its distinct categorization of failure modes. Consequently, the UC dataset was adopted for comparative analysis. Model efficacy was quantified through four metrics: accuracy (Acc), precision (P), recall (R), and F1-score ($F1$), computed via Equations (12)–(15).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$P = \frac{TP}{TP + FP} \tag{13}$$

$$R = \frac{TP}{TP + FN} \tag{14}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{15}$$

In the equations, TP represents true positives; TN represents true negatives; FP represents false positives; and FN represents false negatives.

Figure 11 illustrates the performance metrics of each comparative network. From the figure, it is evident that the proposed model outperforms all comparative models across all four-evaluation metrics, thus further validating its superior fault classification capability. Table 3 presents detailed performance metrics of each comparative network. Both the GRU and GAF-CNN models exhibit poorer performance metrics, all below 90%. The GAF-PCNN model achieves metrics around 92%, demonstrating the effectiveness of parallel training of CNN networks. The GAF-CNN-GRU model achieves metrics above 95%, confirming the efficacy of the combined model. The proposed GAF-PCNN-GRU model achieves metrics exceeding 99%, indicating its comprehensive and superior classification capabilities. These findings demonstrate substantive utility in practical rotating machinery diagnostic implementations.

Table 3. Evaluation metrics of various contrastive networks in the UC dataset.

Network	Accuracy	Precision	Recall	F1_Score
GRU	73.15%	75.24%	73.15%	72.04%
GAF-CNN	87.96%	88.78%	87.96%	87.87%
GAF-PCNN	92.59%	93.19%	92.59%	92.57%
GAF-CNN-GRU	95.83%	96.12%	95.83%	95.82%
GAF-PCNN-GRU	99.54%	99.56%	99.54%	99.54%

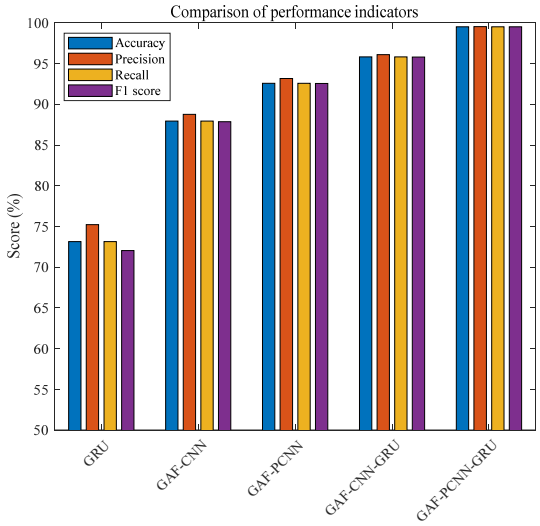


Figure 11. Comparison of network evaluation metrics.

As shown in Table 4, both training and testing times increase progressively with the complexity of the model architecture. The GRU model demonstrates the shortest training time due to its reliance solely on one-dimensional sequential data, which limits its ability to extract deep features. The GAF-CNN series improves spatial feature extraction through image-based representation but lacks the capacity for temporal modeling.

Table 4. Comparison of computational efficiency across models.

Network	CWRU		UC	
	Training Time (s)	Testing Time (s)	Training Time (s)	Testing Time (s)
GRU	35~45	1.0~1.5	40~50	1.2~1.8
GAF-CNN	60~80	1.8~2.2	70~90	1.8~2.3
GAF-PCNN	90~110	2.0~2.6	100~120	2.2~2.8
GAF-CNN-GRU	100~130	2.3~2.8	110~140	2.4~3.0
GAF-PCNN-GRU	130~160	2.5~3.2	140~170	2.6~3.5

In contrast, the proposed GAF-PCNN-GRU model requires slightly more computational time during both training and inference (e.g., 140 s for training and 2.6 s for testing on the UC dataset). However, it integrates dual-modality image modeling, attention mechanisms, and temporal feature learning, resulting in significantly improved diagnostic accuracy, robustness, and generalizability. In practical industrial applications, this level of training cost is acceptable, and the inference time—within a few seconds—is sufficient to support real-time fault diagnosis and rapid response. Therefore, the proposed method achieves a well-balanced trade-off between computational cost and diagnostic performance.

5.3. Noise Adding Experiment

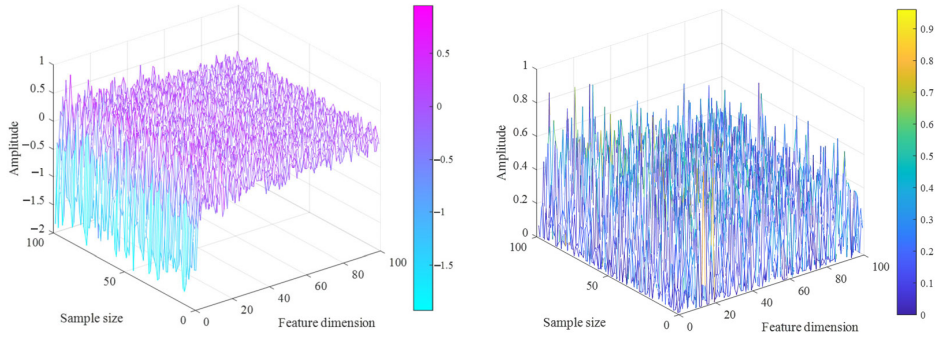
To assess model robustness, gamma-distributed noise was injected into raw data ([936 × 100]) to simulate variations and uncertainties typical of real operational environments, as shown in Equation (16).

$$y = input + gamrnd(1, 2, [936, 100]) \tag{16}$$

where "*input*" denotes the raw data with dimensions [936, 100]; shape = 1, scale = 2.

To intuitively analyze the impact of noise on the distribution characteristics of vibration data, gamma noise was introduced into the original samples of the UC dataset, and three-dimensional visualizations before and after noise injection were plotted, as shown in

Figure 12. Specifically, Figure 12a shows the distribution of the original data without noise, and Figure 12b shows the distribution after gamma noise injection. In the figure, the x-axis represents the feature dimension of the vibration data, the y-axis represents the number of plotted samples, and the z-axis represents the vibration amplitude. It can be observed that after the injection of gamma noise, both the amplitude and spatial distribution of the data changed significantly, reflecting the disturbance effect of noise on the feature state. To enhance the readability of the figure, the first 100 samples were selected for plotting, and the figure was output in vector format to ensure that details remain clear when zooming in or out.



(**a**) (**b**)

Figure 12. Comparative visualization of pristine vs. noise-injected datasets: (**a**) original data; (**b**) original data; noisy data (gamma noise).

To intuitively evaluate the classification performance of the proposed model across different fault types, t-distributed stochastic neighbor embedding (t-SNE) [27] is employed to visualize the high-dimensional features in a two-dimensional space. Specifically, the input to t-SNE is the high-dimensional feature representation obtained from the multi-head self-attention mechanism, and the top three principal dimensions are selected as inputs for dimensionality reduction to preserve representative feature information. As shown in the visualization results in Figure 13, even under the influence of noise, the vast majority of samples are still correctly classified. The nine fault categories form clearly separated clusters in the feature space, with minimal overlap, indicating that the proposed model maintains strong fault discrimination capability under complex operating conditions.

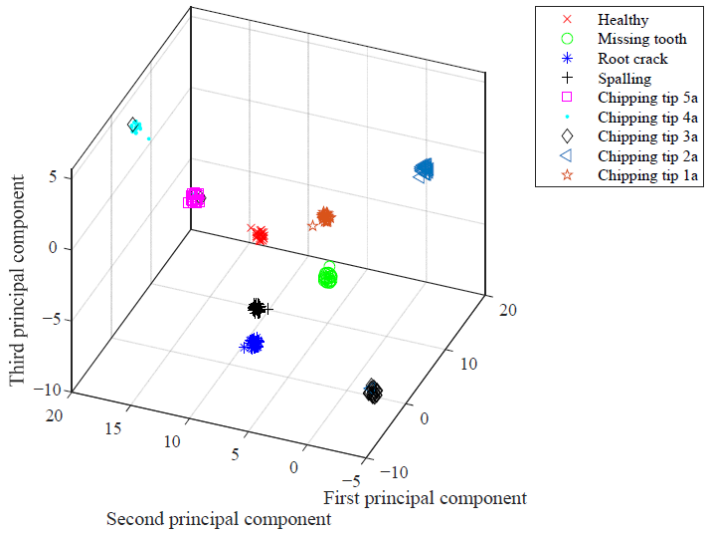


Figure 13. Scatter plot of features in the UC test set.

According to Table 5, the evaluation metrics for each comparative network after noise addition are presented. Following noise introduction, GRU exhibits the poorest robustness, with all metrics declining to below 70%. GAF-PCNN shows a decrease of approximately 10% across metrics, yet maintains higher scores compared to the GAF-CNN model. GAF-CNN-GRU experiences a decrease of around 4% in all metrics. The proposed GAF-PCNN-GRU model demonstrates a marginal decrease of only about 2% across metrics, highlighting its significantly superior robustness compared to other models.

Table 5. Evaluation metrics of contrastive networks on the augmented UC dataset.

Network	Accuracy	Precision	Recall	F1_Score
GRU	68.06%	70.42%	68.06%	68.23%
GAF-CNN	81.48%	82%	81.48%	81.44%
GAF-PCNN	82.87%	85.31%	82.87%	82.99%
GAF-CNN-GRU	91.2%	92.2%	91.2%	91.27%
GAF-PCNN-GRU	97.67%	97.77%	97.67%	97.69%

6. Conclusions

This paper proposes a fault diagnosis method for rotating machinery that integrates GAF, PCNN, and GRU, aiming to address the limitations of one-dimensional vibration signals in feature representation and the difficulty of effectively extracting deep temporal features. Specifically, the GAF method enables the mapping of time-series signals into image space, preserving essential temporal characteristics; the PCNN architecture employs parallel branches to extract multi-modal image features, enhancing spatial representational capability; and the GRU network models temporal dependencies while incorporating a self-attention mechanism to improve the perception of critical fault information.

Extensive experiments conducted on two public datasets, CWRU and UC, demonstrate that the proposed method significantly outperforms several existing models in terms of diagnostic accuracy, robustness, and noise resistance. Notably, the model maintains high classification performance even under gamma noise interference, validating its potential for application in complex industrial environments.

Future work will focus on two main directions: first, compressing the model architecture to improve computational efficiency and reduce deployment costs without compromising diagnostic performance; and second, extending the model to fault diagnosis tasks involving multi-source heterogeneous data, such as multi-sensor fusion and cross-device transfer learning, to enhance its practicality and generalizability.

Author Contributions: Conceptualization, Y.H. and X.D.; methodology, Y.H. and S.C.; software, Y.H.; validation, S.C., Y.H. and X.D.; writing—original draft preparation, X.D. and Y.H.; writing—review and editing, Y.H. and S.C.; visualization, Y.H.; supervision, X.D.; project administration, S.C.; funding acquisition, X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific and Technological Project of Gansu Province, grant number 24CXGA050, the Scientific and Technological Project of Lanzhou City, grant numbers 2024-QN-63 and 2024-3-47, and the Innovation Fund Project of Gansu Provincial Department of Education, grant number 2025A-025. The APC was funded by the Scientific and Technological Project of Lanzhou City, grant number 2024-3-47.

Data Availability Statement: All data used in this study are publicly available: (1) The Case Western Reserve University (CWRU) bearing fault dataset is accessed via a Zenodo mirror: "https://zenodo.org/records/10987113 (accessed on 27 September 2024)". (2) The University of Connecticut (UC) gear fault dataset is hosted on Figure share: "https://figshare.com/articles/dataset/Gear_Fault_Data/6127874/1 (accessed on 27 September 2024)". All analyses and results are based on these open-source datasets, which are fully accessible without restriction.

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

Abbreviations

The following abbreviations are used in this manuscript:

GRU	Gated Recurrent Units
PCNN	Parallel Convolutional Neural Network
GAF	Gramian Angular Field
LSTM	Long Short-Term Memory

References

1. Han, Y.; Zhang, X.; Su, Z.; Huang, Q.; Zhang, Y. An intelligent fault diagnosis of rotating machinery in class-incremental scenarios. *Meas. Sci. Technol.* **2025**, *36*, 046204. [CrossRef]
2. Dalian, Y.; Junjun, Z.; Hui, L. Capsule networks for intelligent fault diagnosis: A roadmap of recent advancements and challenges. *Expert Syst. Appl.* **2025**, *296*, 128814. [CrossRef]
3. Liu, X.; Zhao, Z.; Yang, F.; Liang, F.; Bo, L. Environment adaptive deep reinforcement learning for intelligent fault diagnosis. *Eng. Appl. Artif. Intell.* **2025**, *151*, 110783. [CrossRef]
4. Guan, W.; Wang, S.; Chen, Z.; Wang, G.; Liu, Z.; Cui, D.; Mao, Y. Domain generalization network based on inter-domain multivariate linearization for intelligent fault diagnosis. *Reliab. Eng. Syst. Saf.* **2025**, *261*, 111055. [CrossRef]
5. Wang, S.; Wang, S.; Wang, W.; Zheng, Y.; Ding, B.; Yan, R.; Chen, X. Unrolled sparse coding via lifting wavelet for interpretable intelligent fault diagnosis. *Reliab. Eng. Syst. Saf.* **2025**, *264*, 111371. [CrossRef]
6. Wang, Y.; Zhang, Z.; Xue, C.; Zhu, Q.; Li, X.; Wang, L.; Ding, X. Progressive Transfer Learning: An intelligent fault diagnosis method for unlabeled rotating machinery with small samples. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 1–12. [CrossRef]
7. Zhang, Q.; Xu, C.; Li, J.; Sun, Y.; Bao, J.; Zhang, D. LLM-TSFD: An industrial time series human-in-the-loop fault diagnosis method based on a large language model. *Expert Syst. Appl.* **2025**, *264*, 125861. [CrossRef]
8. Reetz, S.; Neumann, T.; Schrijver, G.; Berg, A.V.D.; Buursma, D. Expert system based fault diagnosis for railway point machines. *Proc. Inst. Mech. Eng. Part F* **2024**, *238*, 214–224. [CrossRef]
9. Lin, K.Y.; Jamrus, T. Industrial data-driven modeling for imbalanced fault diagnosis. *Ind. Manag. Data Syst.* **2024**, *124*, 3108–3137. [CrossRef]
10. Jin, L.; Mao, Y.; Wang, X.; Lu, L.; Wang, Z. Online data-driven fault diagnosis of dual three-phase PMSM drives considering limited labeled samples. *IEEE Trans. Ind. Electron.* **2023**, *71*, 6797–6808. [CrossRef]
11. Zhao, Z.; Jiao, Y. A fault diagnosis method for rotating machinery based on CNN with mixed information. *IEEE Trans. Ind. Inform.* **2022**, *19*, 9091–9101. [CrossRef]
12. Ruan, D.; Wang, J.; Yan, J.; Gühmann, C. CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis. *Adv. Eng. Inform.* **2023**, *55*, 101877. [CrossRef]
13. Qu, J.; Cheng, X.; Liang, P.; Zheng, L.; Ma, X. Fault diagnosis of bearings using wavelet packet energy spectrum and SSA-DBN. *Processes* **2023**, *11*, 1875. [CrossRef]
14. Zhou, H.; Chen, W.; Liu, J.; Cheng, L.; Xia, M. Trustworthy and intelligent fault diagnosis with effective denoising and evidential stacked GRU neural network. *J. Intell. Manuf.* **2023**, *35*, 3523–3542. [CrossRef]
15. Kumar, A.; Parey, A.; Kankar, P.K. A new hybrid LSTM-GRU model for fault diagnosis of polymer gears using vibration signals. *J. Vib. Eng. Technol.* **2024**, *12*, 2729–2741. [CrossRef]
16. Zhu, Y.; Xie, B.; Wang, A.; Qian, Z. Fault diagnosis of wind turbine gearbox under limited labeled data through temporal predictive and similarity contrast learning embedded with self-attention mechanism. *Expert Syst. Appl.* **2024**, *245*, 123080. [CrossRef]
17. Wang, H.; Sun, W.; Sun, W.; Ren, Y.; Zhou, Y.; Qian, Q.; Kumar, A. A novel tool condition monitoring based on Gramian angular field and comparative learning. *Int. J. Hydromechatronics* **2023**, *6*, 93–107. [CrossRef]
18. Li, X.; Zhang, J.; Xiao, B.; Zeng, Y.; Lv, S.; Qian, J.; Du, Z. Fault diagnosis of hydropower units based on Gramian angular summation field and parallel CNN. *Energies* **2024**, *17*, 3084. [CrossRef]
19. Lin, M.; Wu, J.; Meng, J.; Wang, W.; Wu, J. Screening of retired batteries with Gramian angular difference fields and ConvNeXt. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106397. [CrossRef]
20. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
21. Taye, M.M. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation* **2023**, *11*, 52. [CrossRef]

22. Dey, R.; Salem, F.M. Gate-variants of gated recurrent unit (GRU) neural networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017; IEEE: New York, NY, USA, 2017; pp. 1597–1600.
23. Xia, M.; Shao, H.; Ma, X.; de Silva, C.W. A stacked GRU-RNN-based approach for predicting renewable energy and electricity load for smart grid operation. *IEEE Trans. Ind. Inform.* **2021**, *17*, 7050–7059. [CrossRef]
24. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
25. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [CrossRef]
26. Cao, P.; Zhang, S.; Tang, J. Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access* **2018**, *6*, 26241–26253. [CrossRef]
27. Anvari, M.A.; Rahmati, D.; Kumar, S. t-Distributed stochastic neighbor embedding. In *Dimensionality Reduction in Machine Learning*; Morgan Kaufmann: Burlington, MA, USA, 2025; pp. 187–207.

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.