

Article

A Multi-Level Fusion Framework for Bearing Fault Diagnosis Using Multi-Source Information

Xiaojun Deng ¹, Yuanhao Sun ¹ , Lin Li ^{2,3}  and Xia Peng ^{2,3,*}

¹ School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China; little_army@hut.edu.cn (X.D.); m23085400008@stu.hut.edu.cn (Y.S.)

² School of Intelligent Manufacturing, Hunan First Normal University, Changsha 410205, China; lilin19@nudt.edu.cn

³ Key Laboratory of Industrial Equipment Intelligent Perception and Maintenance in College of Hunan Province, Changsha 410205, China

* Correspondence: pandsia@hnfnu.edu.cn

Abstract

Rotating machinery is essential to modern industrial systems, where rolling bearings play a critical role in ensuring mechanical stability and operational efficiency. Failures in bearings can result in serious safety risks and significant financial losses, which highlights the need for accurate and robust methods for diagnosing bearing faults. Traditional diagnostic methods relying on single-source data often fail to fully leverage the rich information provided by multiple sensors and are more prone to performance degradation under noisy conditions. Therefore, this paper proposes a novel bearing fault diagnosis method based on a multi-level fusion framework. First, the Symmetrized Dot Pattern (SDP) method is applied to fuse multi-source signals into unified SDP images, enabling effective fusion at the data level. Then, a combination of RepLKNet and Bidirectional Gated Recurrent Unit (BiGRU) networks extracts multi-modal features, which are then fused through a cross-attention mechanism to enhance feature representation. Finally, information entropy is utilized to assess the reliability of each feature channel, enabling dynamic weighting to further strengthen model robustness. The experiments conducted on public datasets and noise-augmented datasets demonstrate that the proposed method significantly surpasses other single-source and multi-source data fusion models in terms of diagnostic accuracy and robustness to noise.

Keywords: bearing fault diagnosis; multi-source information fusion; Symmetrized Dot Pattern (SDP); cross-attention mechanism



Academic Editor: Maoyin Chen

Received: 6 July 2025

Revised: 20 August 2025

Accepted: 20 August 2025

Published: 21 August 2025

Citation: Deng, X.; Sun, Y.; Li, L.; Peng, X. A Multi-Level Fusion Framework for Bearing Fault Diagnosis Using Multi-Source Information. *Processes* **2025**, *13*, 2657. <https://doi.org/10.3390/pr13082657>

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 serves as the backbone of modern industry and is integral to a wide range of engineering systems. Rolling bearings, as essential components of such machinery [1], often operate in harsh environments. Prolonged operation under such conditions significantly increases the risk of failure. Bearing failures can lead to substantial economic losses, serious safety risks, and even potential threats to human life [2]. These consequences highlight the importance of fault diagnosis for rolling bearings in rotating equipment.

In recent years, deep learning techniques have attracted more and more interest in fault diagnosis because of their powerful capability for automatic feature extraction from raw signals through end-to-end learning architectures. Various deep learning models, including

convolutional neural networks (CNN) [3], long short-term memory networks (LSTM) [4], and graph convolutional networks (GCN) [5], have been widely applied to bearing fault diagnosis and achieved good results. However, most of the current research relies solely on single-source data, making it vulnerable to noise, sensor failures, and external disturbances. These limitations restrict their ability to capture comprehensive equipment information, often resulting in suboptimal diagnostic performance. Meanwhile, advancements in sensor technology have made it easier to acquire multi-source data from multiple sensors. This enables the use of multi-source information to develop more accurate and robust diagnostic models for rotating machinery.

Currently, fault diagnosis methods that utilize multi-source information fusion can be categorized into three levels: data-level fusion, feature-level fusion, and decision-level fusion [6]. The data-level fusion approach uses raw signals from multiple sensors and applies signal processing techniques to generate a unified dataset. This approach aims to exploit the original data fully while minimizing information loss. Generally, multi-source data are either directly concatenated or treated as separate channels to construct fused data. For example, Azamfar et al. [7] proposed a data-level fusion approach by stacking sensor signals in row format to form a two-dimensional input matrix, enabling the integration of multi-source information. Yang et al. [8] proposed a data weighting strategy utilizing enhanced fuzzy support. This strategy employs dynamic bending distance instead of Euclidean distance to compute the support of sensor signals within a given time window, thereby achieving weighted fusion of sensor data. Xie et al. [9] proposed an intelligent diagnostic approach combining multi-sensor fusion with CNN, where principal component analysis (PCA) is utilized to convert multi-source signals into a three-channel RGB image, facilitating data-level fusion. Despite these advancements, most existing data-level fusion strategies rely heavily on direct concatenation or weighted summation of raw signals. These methods often overlook the intrinsic relationships and interdependencies among signals from different sources. As a result, their ability to capture complementary and synergistic information is limited, ultimately restricting the effectiveness of the fusion process.

In the feature-level fusion approach, features extracted from multi-source data through signal analysis or neural networks are integrated using methods such as feature concatenation, weighted fusion, and attention mechanisms. The feature-level fusion method is capable of extracting key features from complex information to enhance the representation of the fused features. Dong et al. [10] proposed an efficient fusion method, MD-1d-DCNN, which combines features from various domains with adaptive features extracted via a one-dimensional dilated CNN. This method enables effective fault diagnosis, even in the presence of noisy data. Xiao et al. [11] proposed a multi-feature fusion strategy based on an attention mechanism, which integrates various feature types extracted by CNN and GCN from multi-source data, improving the quality of the fused feature. Gao et al. [12] proposed a multi-source heterogeneous information fusion method based on a graph convolutional network (MHIF-GCN). The framework uses a convolutional autoencoder (CAE) to extract deep features from different sensor types, which are then used as graph node features to address data heterogeneity. However, most of the current deep learning-based feature-level fusion methods fuse features at only a single stage. This limitation restricts their ability to fully capture equipment information across different stages, leading to partial information loss.

In the decision-level fusion method, classification outcomes from multiple sources are aggregated to achieve information fusion. Common techniques for decision-level fusion include weighted voting, Dempster-Shafer (DS) evidence theory, and fuzzy logic. These methods provide an efficient way to manage conflicting information from diverse sensors and demonstrate robust performance in the presence of interference and noise.

Gao et al. [13] proposed a dual fusion GCN that uses entropy-weighted DS evidence theory to merge diagnostic decisions from multiple parallel branches, enabling effective information fusion. Shao et al. [14] proposed a flexible weighted strategy for decision fusion, which combines the outputs of multiple stacked wavelet autoencoders with Morlet wavelet functions, further enhancing the fusion process. Zhang et al. [15] proposed an improved framework of the Multi-Graph Convolutional Network, termed Trusted Multi-Source Information Fusion (MGCN-TMIF), which integrates multi-channel evidence using the reduced DS evidence theory to achieve reliable information fusion. This approach has high anti-noise performance. However, the effectiveness of decision-level fusion heavily depends on the chosen fusion strategy. When conflicting information arises, it may fail to assign appropriate weights, resulting in inaccurate or unreliable decisions.

To address the aforementioned challenges, this paper proposes a multi-level fusion framework for bearing fault diagnosis using multi-source information. The proposed framework comprises three main components: a Symmetrized Dot Pattern (SDP)-based data-level fusion module, a cross-attention-based multi-view feature fusion strategy, and an information entropy-based feature channel fusion module. The main contributions of this paper are summarized as follows:

(1) A novel data-level fusion strategy is proposed based on the SDP-based data-level fusion module. The raw signals from different information sources are fused into an SDP image to construct a unified data representation, which improves the discrimination of different fault features and achieves efficient fusion of multi-source information.

(2) A cross-attention-based multi-view feature fusion strategy is proposed. By enhancing query features and leveraging a cross-attention mechanism, this method maps multi-modal features into a unified latent space, fully utilizing their complementary information and improving the representation of the fused features.

(3) An information entropy-based feature channel fusion module is proposed. This module quantifies the uncertainty of feature distributions across different channels through entropy analysis, dynamically assigning higher weights to more reliable channels. It effectively leverages reliable features and minimizes the influence of unreliable sensors, improving the robustness of the diagnostic model.

(4) A multi-level fusion framework is proposed, which integrates SDP transformation, RepLKNet, BiGRU, cross-attention mechanisms, and entropy-weighted channel fusion to fully utilize the strengths of each module. This design improves the efficiency of information fusion across multi-source data and enhances diagnostic performance. Additionally, unlike traditional methods that focus on a single fusion level or modality, our framework provides a unified, hierarchical, and end-to-end approach that integrates data, features, and channels from multiple sources. Finally, it shows strong performance even under significant noise interference.

The structure of this paper is as follows: Section 2 introduces the basic principles of SDP and the theoretical foundations of the RepLKNet architecture. Section 3 details the overall architecture of the proposed method and the implementation specifics of each key component. Section 4 describes the relevant experimental setup, including the datasets used and the comparative methods, and discusses the results that validate the effectiveness of the proposed method. Finally, Section 5 summarizes the key points of the paper and suggests directions for future research.

2. Related Theory Introduction

2.1. The Theory of SDP

The Symmetrized Dot Pattern (SDP) [16] is a method for visualizing one-dimensional time series data as a scatter plot in a polar coordinate system. It normalizes the time-

domain signal $X = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ and then converts it into $S(r(i), \theta(i), \varphi(i))$ in the polar coordinate system by generating symmetric points. $r(i)$ is the radial coordinate, and $\theta(i)$ and $\varphi(i)$ are the angular coordinates corresponding to counterclockwise and clockwise rotations relative to the reference axis. This approach represents the time series signal as a simple, intuitive polar coordinate representation, allowing for a clear distinction between different fault types. The principle is illustrated in Figure 1.

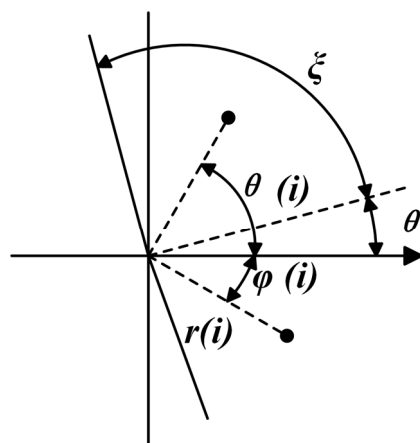


Figure 1. The basic principles of SDP.

For the time series X , x_i denotes the i -th sampling point of X , and x_{i+l} represents the $(i + l)$ -th sampling point after a time interval l . According to the SDP principle, when transforming the time-domain point x_i into the polar coordinate space, its radial coordinate $r(i)$ is defined as:

$$r(i) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where $x_{\max}$ and $x_{\min}$ denote the maximum and minimum values of the signal, respectively. The angular coordinates $\theta(i)$ and $\varphi(i)$ are determined by the adjacent time-domain points x_{i+l} . These can be expressed as follows:

$$\theta(i) = \theta + \frac{x_{i+l} - x_{\min}}{x_{\max} - x_{\min}} \zeta \quad (2)$$

$$\varphi(i) = \theta - \frac{x_{i+l} - x_{\min}}{x_{\max} - x_{\min}} \zeta \quad (3)$$

where l is the time interval parameter ($1 \leq l \leq 10$) [17], and differences in the time series can be reflected through its varying values. The θ represents the rotational angle of the mirror symmetry plane ($\theta = 360i/n, i = 1, 2, \dots, n$) [18], while ζ is the angular gain factor ($\zeta \leq \theta$).

SDP analysis transforms time series into scatter pairs in the polar coordinate system, forming two symmetric petals within the angular range $[\theta - \zeta, \theta + \zeta]$. Here, θ denotes the angle of the axis of symmetry for the two petals, ζ limits the angular distribution range of the petals, and the time interval parameter l affects the shape of the petals. According to Equations (1)–(3), when l and ζ are fixed, the mapping shapes of signals in the polar coordinate system vary across different fault states due to variations in amplitude and frequency. Consequently, the distribution patterns of points in the SDP image reflect differences among various fault states. This characteristic enables the accurate identification and classification of different fault states. When multiple sensors are available, SDP analysis can fuse all sensor signals into a single SDP image by adjusting θ and ζ . This not only achieves data-level fusion of multi-source information but also enhances the discrimination among different fault states, making the classification of different faults more accurate.

2.2. RepLKNet: A Large-Kernel Architecture

Traditional CNNs rely on stacking convolutional layers to expand their receptive fields. However, the increase in their effective receptive field [19] is limited, constraining their ability to capture more global information. In contrast, RepLKNet uses a small number of large convolutional kernels rather than stacking additional convolutional layers. This design enables more efficient capture of global information in images. As a result, RepLKNet exhibits a significantly larger effective receptive field and superior global feature extraction capabilities compared to traditional CNNs of the same depth.

RepLKNet is a convolution-based model that replaces the attention mechanism in the Swin Transformer [20] with large 31×31 convolutional kernels [21]. It uses depth-wise convolution (DW) [22] and re-parameterization [23] to achieve efficient and high-performance convolutions. The overall structure of RepLKNet is illustrated in Figure 2. It consists of three main components: the Stem Layer, which preserves fine-grained details and provides local features; the Stage Module, which combines RepLK Blocks to expand the receptive field and ConvFFN Blocks to enhance cross-channel nonlinearity; and the Transition Block, which adjusts feature resolution and channel dimensions after each stage.

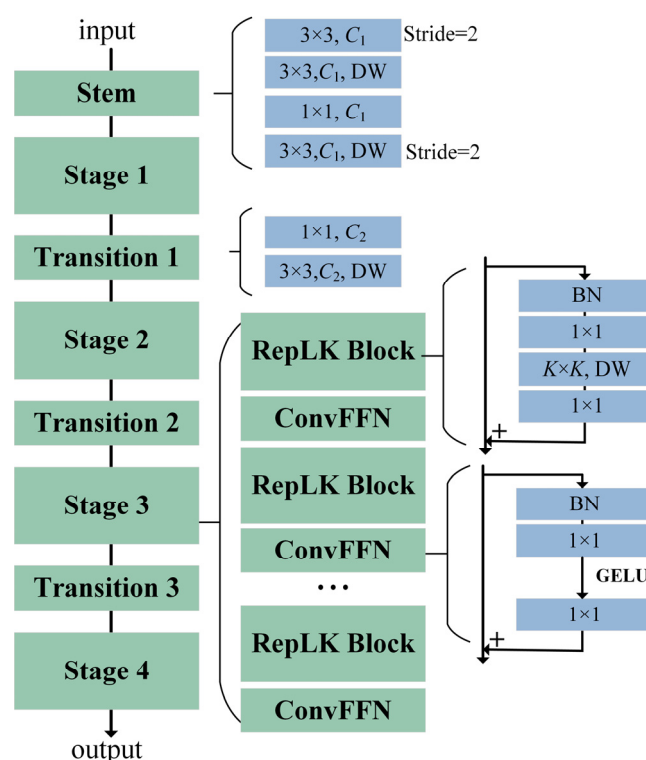


Figure 2. The basic RepLKNet framework.

This design of RepLKNet not only significantly enhances the effective receptive field and shape bias [24] but also preserves the computational efficiency of CNNs. During the recognition of SDP images, extracting correlations between different information sources and distinguishing differences among fault states are crucial. For this task, a larger effective receptive field and stronger shape bias are particularly important. Therefore, RepLKNet not only demonstrates significant advantages in extracting SDP image features but also can perform feature extraction efficiently.

3. Proposed Fault Diagnosis Methods

To address the limitations of fault diagnosis that relies on a single signal source and to utilize the complementary information from multi-source signals, this paper proposes a

multi-level, multi-source information fusion framework. The method comprises three core components: an SDP-based data-level fusion module, a cross-attention-based multi-view feature fusion strategy, and an information entropy-based feature channel fusion module. First, the SDP-based data-level fusion module constructs a unified representation of multi-source information by generating SDP images and utilizes the correlations among multi-source data to achieve data-level fusion. Next, the RepLKNet and Bidirectional Gated Recurrent Unit (BiGRU) models are used to extract multi-modal features from the integrated multi-source information. By combining the attention mechanism, the extracted features are enhanced and fused, effectively compensating for the limitations of individual models in feature extraction. Finally, to further improve the reliability of the fused features, an information entropy-based feature channel fusion module is proposed. This approach assigns different weights to feature channels according to their respective reliability, thereby significantly enhancing the overall robustness and reliability of the model. The overall flowchart is shown in Figure 3.

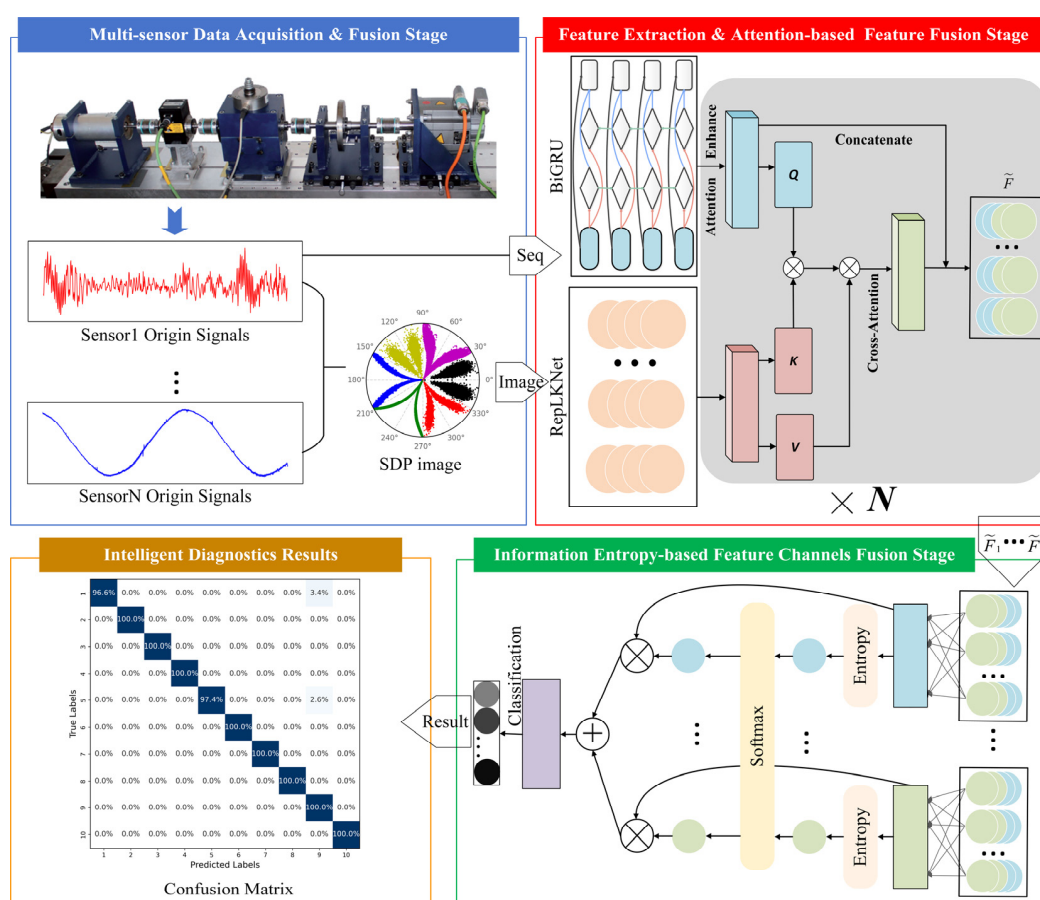


Figure 3. Flowchart of the proposed method.

3.1. SDP-Based Data-Level Fusion Module

This paper proposes a data-level multi-source information fusion method based on SDP analysis. First, it employs Variational Mode Decomposition (VMD) to decompose raw signals acquired from multiple sensors into Intrinsic Mode Functions (IMFs) across various frequency bands adaptively. Then, a modal energy analysis method is used to select more representative IMFs. Finally, both the original signal and the selected IMFs are transformed into the polar coordinate system through SDP analysis, achieving effective data-level fusion of multi-source information. The construction process of the SDP image is illustrated in Figure 4.

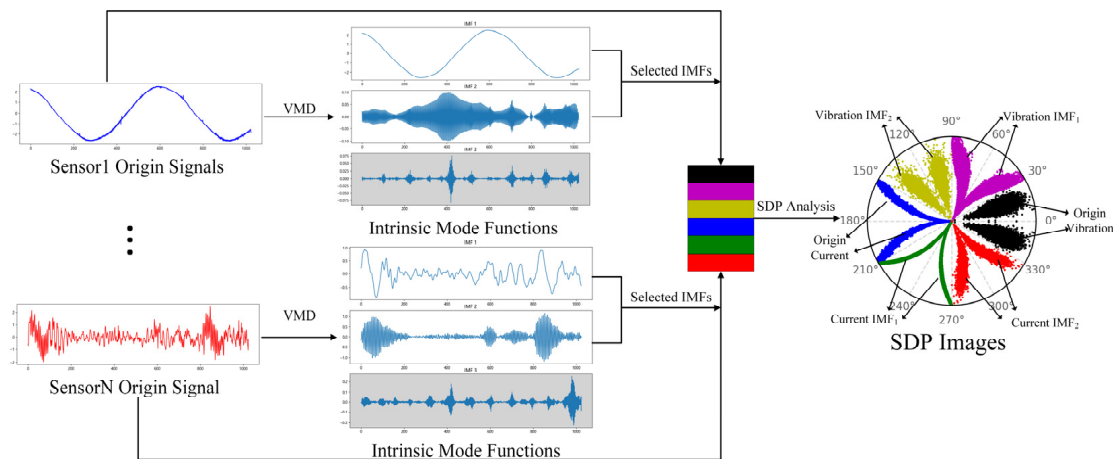


Figure 4. SDP image construction process.

VMD decomposition extracts components from complex signals across different frequency bands, effectively avoiding modal mixing. This makes it suitable for processing multi-source data with various fault types and varying severity levels. During the decomposition process, an appropriate choice of K is crucial for effectively capturing informative components from the original signal. In this study, considering the characteristics of multi-source information and balancing the effectiveness of decomposed information with computational efficiency, the number of modes K is set to 3 [25].

To avoid overlap or insufficient symmetry in the SDP image, which can hinder the extraction of features between sensors, this study sets the number of sectors n to 6 [26]. As a result, not all IMFs derived from VMD can be mapped onto a single SDP image. Therefore, it is essential to identify representative modes in the IMFs. This paper employs the IMF energy analysis method. The IMF components are ranked according to their energy levels, with higher-energy IMFs selected first. The key formulation is as follows:

$$E_{IMF_i} = \sum_{t=1}^T IMF_i(t)^2 \quad (4)$$

where E_{IMF_i} represents the energy of the mode IMF_i , and $IMF_i(t)$ is the value of the mode IMF_i at time t . The energy of each IMF reflects its importance within the signal. Generally, most of the signal's energy is concentrated in a limited number of IMFs, especially the early modes. These high-energy IMFs typically contain the dominant fault-related information [27].

Let N denote the number of sensors and M denote the number of IMFs selected for each sensor. The fused image is composed of n sectors. In this study, given the limited number of sensors ($N = 2$), when the SDP image is divided into $n = 6$ sectors, it follows that $M = (n/N) - 1 = 2$. Consequently, each sensor selects only the two IMFs with the highest energy. The selected IMFs are combined with their respective raw signals and then converted into an SDP image. This approach ensures that multi-source information is fused within the same polar coordinate system, resulting in a consistent representation of data from different sensors.

When $n = 6$, substituting into the formula $\theta = 360i/n$, we obtain $\theta = 60^\circ$. Accordingly, the symmetry angle values of the SDP image are $0^\circ, 60^\circ, 120^\circ, 180^\circ, 240^\circ$, and 300° . To prevent image overlap, ζ is set to 30° , which is half of θ [28]. The parameter l effectively captures signal differences within the range of 1 to 10 [17]. In this paper, under the conditions of $\theta = 60^\circ$ and $\zeta = 30^\circ$, a series of experiments with a step size of 2 were conducted. The results indicated that when $l = 8$, the differences between SDP images under various fault states became more pronounced, particularly in terms of the angles and

curvatures of the petals. The findings are detailed in Table 1. Consequently, the optimal rotational angle of the mirror symmetry plane, angular gain factor, and time interval parameter are determined to be $\theta = 60^\circ$, $\xi = 30^\circ$, and $l = 8$, respectively. Finally, the original signals and their selected IMFs are transformed into the polar coordinate system based on predefined angles to generate the fused SDP image. The process of the SDP-based data-level fusion is shown in Algorithm 1.

Algorithm 1: SDP-based Data-level Fusion

Require: Multi-sensor signals $S = \{S_1(t), \dots, S_N(t)\}$; number of VMD mode K ; number of sectors n ; SDP params θ , ξ , and l ;

Ensure: Fused SDP image I_{fused} ;

```

1: for each  $S_i(t) \in S$  do
2:    $IMF_{i1}, IMF_{i2}, \dots, IMF_{iK} \leftarrow \text{VMD}(S_i(t), K)$ ;
3:    $E_{ik} \leftarrow \sum_{t=1}^T IMF_i(t)^2$ ;
4:    $IMF_{i1}, \dots, IMF_{iM} \leftarrow \text{Select } M \text{ IMFs by highest } E_{ik}; M \leftarrow n/N - 1$ ;
5:    $data_i \leftarrow \text{concat}(S_i(t), IMF_{i1}, \dots, IMF_{iM})$ ;
6: end for
7:  $All\_data \leftarrow \text{concat}(data_1, \dots, data_N)$ ;
8:  $\theta_k \leftarrow k \cdot \theta, k = 0, \dots, n - 1$ ;
9: for each sector  $k = 0$  to  $n - 1$  do
10:  Compute  $r(i), \theta(i), \varphi(i)$ , Plot  $S(r(i), \theta(i), \varphi(i))$  in polar coordinates;
11: end for
12: return  $I_{\text{fused}}$ 

```

Table 1. SDP graph for different states of the bearing with different values of l .

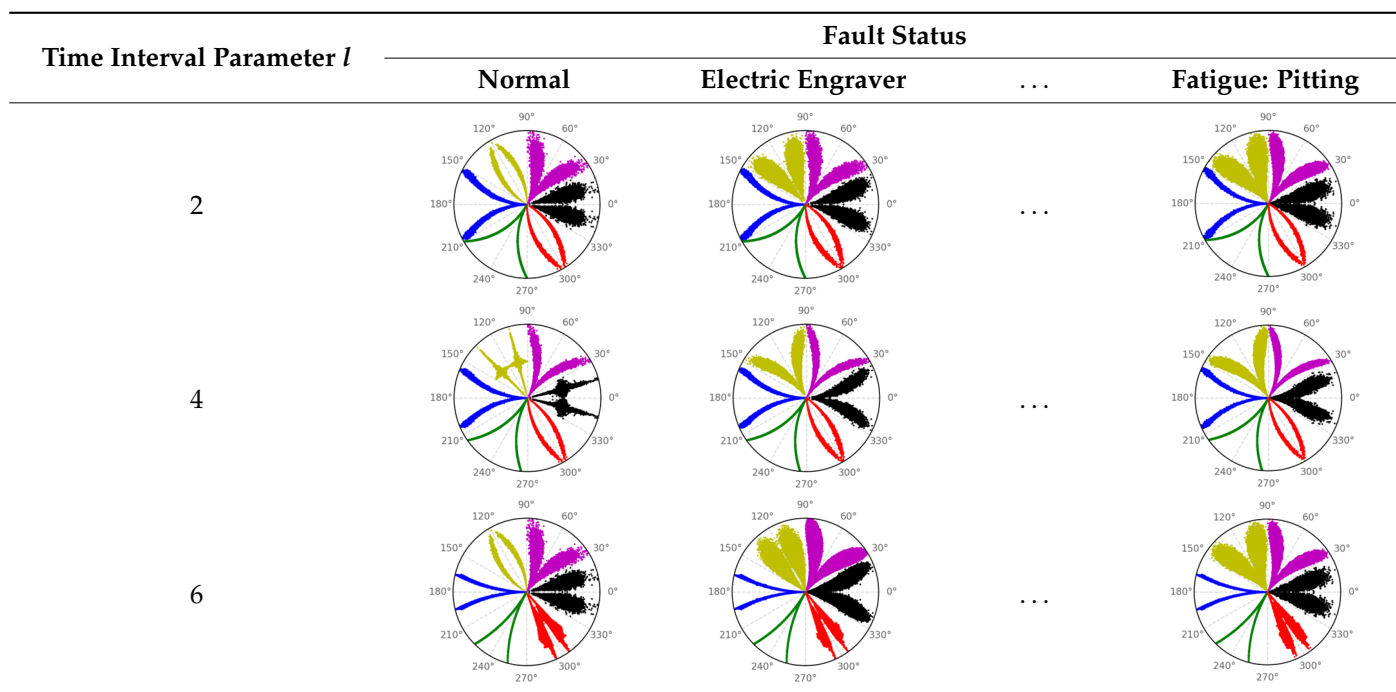


Table 1. Cont.

Time Interval Parameter l	Fault Status			
	Normal	Electric Engraver	...	Fatigue: Pitting
8			...	
10			...	

3.2. Cross-Attention-Based Multi-View Feature Fusion Strategy

In this paper, we employ RepLKNet and BiGRU to process multi-sensor data. Specifically, RepLKNet is used to capture relational features from SDP images, while BiGRU is focused on extracting intrinsic features from time-series signals. However, each feature type has its own limitations. Relying solely on a single feature type for fault diagnosis hinders the model's ability to achieve optimal results. Therefore, we propose a cross-attention-based multi-view feature fusion strategy to address this issue. First, we enhance the temporal features from BiGRU to improve the connection between the output features and global information. Then, the enhanced features are fused with the image features from RepLKNet in a unified latent space, effectively utilizing their complementarity to improve the final diagnostic output. The specific structure of the cross-attention-based multi-view feature fusion strategy is shown in Figure 5.

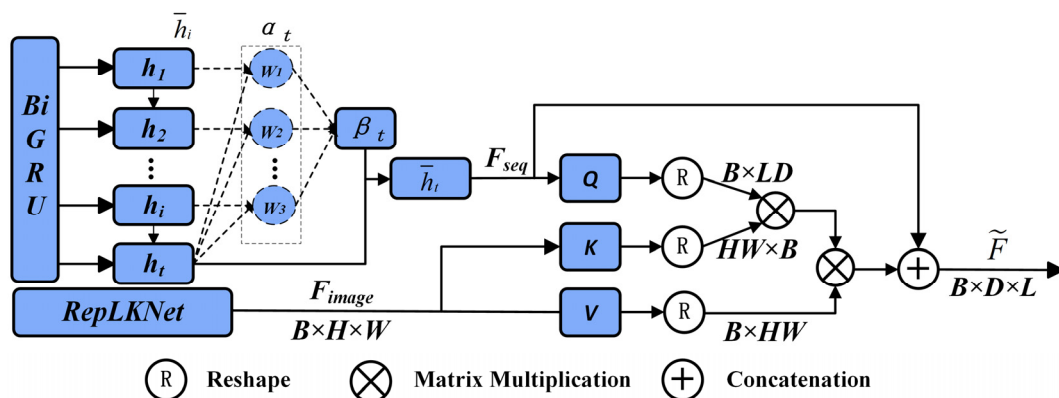


Figure 5. The structure of the attention mechanism.

BiGRU struggles to capture global dependencies in long-range time series due to its inherent limitations. Directly using time-series features for fusion may lead to feature mismatches, particularly when dealing with low-quality queries. To improve query quality, modeling global dependencies becomes necessary. The cross-attention-based multi-view feature fusion strategy captures relationships between the target hidden state and all hidden states in the input time series, facilitating better feature fusion. Specifically, given the hidden states of BiGRU $\bar{h}_i$ ($i = 1, 2, \dots, t$), we compute a globally enhanced representation $\bar{h}_i$ through the following steps:

$$\alpha_t(i) = \frac{\exp(\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_i))}{\sum_j \exp(\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_j))} \quad (5)$$

$$\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_i) = v_\alpha^\top \tanh(W_\alpha[\mathbf{h}_t; \bar{\mathbf{h}}_i]) \quad (6)$$

where the attention weight $\alpha_t(i)$ quantifies the degree of relevance between the hidden state of all historical input information and the target hidden state $\mathbf{h}_t$. The $\text{score}(\mathbf{h}_t, \bar{\mathbf{h}}_i)$ is the relevance score, obtained by comparing the target hidden vector with other hidden vectors $\mathbf{h}_i$. W_α and v_α represent the weight matrices.

$$\beta_t = \sum_i \alpha_t(i) \bar{\mathbf{h}}_i \quad (7)$$

$$\bar{\mathbf{h}}_t = \tanh(W_\beta[\beta_t; \mathbf{h}_t]) \quad (8)$$

where the global vector β_t is calculated based on $\alpha_t(i)$, serving as the weighted value for each hidden state. Subsequently, the weighted vector β_t of the entire hidden layer is combined with the target hidden vector $\mathbf{h}_t$ to compute the final output $\bar{\mathbf{h}}_t$.

Then, the features extracted from images and time-series signals are separately input into the cross-attention mechanism to establish cross-modal associations. In the proposed method, enhanced time-series features act as the query, while SDP image features serve as the key-value pair, facilitating more effective cross-modal fusion. The computational formulation is presented as follows:

$$F_{\text{fusion}} = \text{softmax}\left(\frac{Q_{\text{seq}} K_{\text{image}}^T}{\sqrt{d_k}}\right) \cdot V_{\text{image}} \quad (9)$$

$$\tilde{F} = \text{concat}(F_{\text{fusion}}, F_{\text{seq}}) \quad (10)$$

where d_k denotes the dimension of the key vector, Q_{seq} is derived from temporal features extracted by BiGRU, while K_{image} and V_{image} are provided by SDP image features extracted by RepLKNet. Finally, the fused features F_{fusion} are concatenated with the input sequence features F_{seq} to produce the final representation $\tilde{F}$, preserving the original sensor information while enhancing their interdependencies. The process of the cross-attention-based multi-view feature fusion is shown in Algorithm 2.

Traditional feature fusion methods, such as direct concatenation or weighted averaging, rely on simplistic operations that project different features into a joint space. These methods often fail to capture the underlying dependencies between feature modalities. In contrast, our proposed cross-attention-based multi-view feature fusion strategy addresses this limitation. It explicitly captures the correlations between temporal and image-based features, which can effectively model cross-modal dependencies. This strategy enables more effective fusion of spatiotemporal information across modalities, thereby enhancing both the efficiency and accuracy of multi-modal feature integration.

Algorithm 2: Cross-attention-based Multi-view Feature Fusion

Require: Multi-sensor signals $S = \{S_1(t), \dots, S_N(t)\}$; SDP image I_{fused} ;
Ensure: fused feature representation $\tilde{F}_1, \dots, \tilde{F}_N$;
1: Extract image feature: $F_{\text{img}} \leftarrow \text{RepLKNet}(I_{\text{fused}})$;
2: **for** each signal $S = \{S_1(t), \dots, S_N(t)\}$ **do**
3: Hidden states $\bar{h}_i \leftarrow \text{BiGRU}(S)$;
4: **for** each $h_t \in \bar{h}_i$ **do**
5: Compute attention weights: $\text{score}(h_t, \bar{h}_i) \leftarrow v_\alpha^\top \tanh(W_\alpha[h_t; \bar{h}_i])$, $\alpha_t(i) \leftarrow \text{softmax}(\text{score}(h_t, \bar{h}_i))$;
6: Global vector $\beta_t \leftarrow \sum_i \alpha_t(i) \bar{h}_i$;
7: Enhanced state $F_{\text{seq}} \leftarrow \bar{h}_t = \tanh(W_\beta[\beta_t; h_t])$;
8: **end for**
9: Cross-attention: $Q \leftarrow F_{\text{seq}} W_Q$, $K \leftarrow F_{\text{img}} W_K$, $V \leftarrow F_{\text{img}} W_V$, $A \leftarrow \text{softmax}(QK^\top / \sqrt{d}) V$;
10: Fused feature $\tilde{F} \leftarrow \text{concat}(A, F_{\text{seq}})$;
11: **end for**
12: **return** $\tilde{F}_1, \dots, \tilde{F}_N$

3.3. Information Entropy-Based Feature Channel Fusion Module

This study aims to build a robust diagnostic model based on multi-source information, where multiple sensors contribute diverse feature channels. To enhance the discriminative power of the fused features, it is necessary to consider the varying reliability of each channel. Traditional average-weight fusion methods ignore these differences, which can lead to underutilization of informative features or amplification of noise, ultimately reducing diagnostic accuracy. To address this issue, we propose an information entropy-based fusion module that adaptively assigns weights to each channel based on its feature distribution. Entropy is used to quantify the uncertainty within each channel. As is widely recognized, higher entropy indicates more dispersed features and lower classification confidence, thus resulting in lower assigned weights. This dynamic weighting mechanism not only improves the reliability of feature fusion but also boosts diagnostic performance.

Given the feature information from multiple sensor channels, let the feature vector of the i -th sensor channel be $\tilde{F}_i = \{f_{i1}, f_{i2}, \dots, f_{in}\}$. A lower entropy value indicates that the feature information exhibits a higher degree of discrimination. Therefore, a larger weight should be assigned to it. To achieve this, we need to calculate the weight w_i and then obtain the final fused feature F_{fused} .

$$F_{\text{fused}} = \sum_n^{i=1} w_i \cdot \tilde{F}_i \quad (11)$$

$$w_i = \frac{\exp(1/(H_i + \epsilon))}{\sum_{j=1}^n \exp(1/(H_j + \epsilon))} \quad (12)$$

where ϵ denotes a small constant (e.g., 1×10^{-8}) introduced to ensure numerical stability and prevent division by zero errors. The weights from the calculations are used for the weighted fusion of the features from each channel, resulting in the fused feature F_{fused} . The H_i represents the information entropy of the feature vector from the i -th sensor channel. Each output feature $\tilde{F}_i$ is normalized, followed by computation of its information entropy H_i according to the following formulation:

$$P_i^{(k)} = \frac{\exp(\tilde{F}_i^{(k)})}{\sum_{j=1}^n \exp(\tilde{F}_i^{(j)})} \quad (13)$$

$$H_i = - \sum_{k=1}^n P_i^{(k)} \log P_i^{(k)} \quad (14)$$

where $P_i^{(k)}$ represents the probability value at the k -th dimension of the i -th sensor ($k = 1, 2, \dots, n$). The process of the information entropy-based feature channel fusion is shown in Algorithm 3.

Algorithm 3: Information Entropy-based Feature Channel Fusion

Require: Feature from N channels $\tilde{F} = \{\tilde{F}_1, \dots, \tilde{F}_N\}$, Each $\tilde{F}_i = \{f_{i1}, f_{i2}, \dots, f_{in}\}$, small constant ϵ ;

Ensure: Final fused feature F_{fused} ;

1: **for** each channel $\tilde{F}_i \in \tilde{F}$ **do**

2: **for** each dimension $k = 1$ to n **do**

3: probability value $P_i^{(k)} \leftarrow \text{softmax}(\tilde{F}_i^{(k)})$;

4: **end for**

5: Compute entropy $H_i \leftarrow - \sum_{k=1}^n P_i^{(k)} \log P_i^{(k)}$;

6: **end for**

7: weight $w_i \leftarrow \text{softmax}(1/H_i + \epsilon)$;

8: final fused feature $F_{\text{fused}} \leftarrow \sum_{i=1}^n w_i \cdot \tilde{F}_i$

9: **return** F_{fused}

4. Discussion

This study used an AMD Ryzen 5 CPU and an NVIDIA GeForce RTX 4060 GPU as the computational platform. The research environment was implemented using PyTorch 2.0, and the model code was written in Python 3.9. Experiments were conducted using the publicly available Paderborn University dataset and a noisy signal dataset to validate the efficacy and superiority of the proposed method.

4.1. Dataset

This experiment utilized the Paderborn University dataset provided by Lessmeier et al. [29]. The experimental platform comprises a test motor, torque measurement shaft, bearing test module, flywheel, and load motor. This test rig provides multi-source data, including motor current signals and bearing vibration signals, under varying operational conditions. The dataset classifies bearing conditions into four classes: healthy condition, inner ring fault (IR), outer ring fault (OR), and compound fault (IR&OR). Each class includes varying degrees of fault severity and symptoms. For this study, 10 classes of data were selected for the experiment. The operational conditions were set as follows: radial force of 1000 N, load torque of 0.7 Nm, and bearing speed of 1500 rpm. Since both the vibration and current signals in the dataset share the same sampling frequency of 64 kHz, the sample length was set to 1024 during dataset preparation. Further details regarding the dataset are summarized in Table 2.

Table 2. The detailed description of the Paderborn datasets.

Label	Bearing Code	Bearing Running Status	Damage Symptom	Damage Level
1	K001	Normal	/	/
2	KA01	Electrical discharge machining	OR	1
3	KA03	Electric engraver	OR	2
4	KA05	Electric engraver	OR	1

Table 2. Cont.

Label	Bearing Code	Bearing Running Status	Damage Symptom	Damage Level
5	KI01	Electrical discharge machining	IR	1
6	KI03	Electric engraver	IR	1
7	KI07	Electric engraver	IR	2
8	KB23	Fatigue: pitting	IR&OR	2
9	KB24	Fatigue: pitting	IR&OR	3
10	KB27	Fatigue: pitting	IR&OR	1

4.2. Experimental Detail Configuration

The proposed model integrates a RepLKNet and multiple BiGRUs, where each BiGRU is assigned to a specific sensor. The RepLKNet employs a multi-stage architecture consisting of four core modules. Each module is characterized by three hyperparameters: the number of RepLK Blocks (B), channel dimension (C), and kernel size (K). In this paper, the BiGRU employs a two-layer architecture, containing 32 and 64 hidden units, respectively. All attention mechanisms are implemented as single-head scaled dot-product types, with a dropout rate of 0.3 uniformly applied across the model. During training, the model uses cross-entropy as the loss function and the Adam optimizer for optimization. After multiple iterations, the optimal settings are achieved. The detailed parameter configurations are listed in Table 3.

The size of the SDP images is set to 224×224 . In the Paderborn dataset, the experimental data are randomly divided into three subsets: 70% for training, 20% for validation, and 10% for testing. Ten fault states are defined, and their corresponding SDP images are presented in Figure 6.

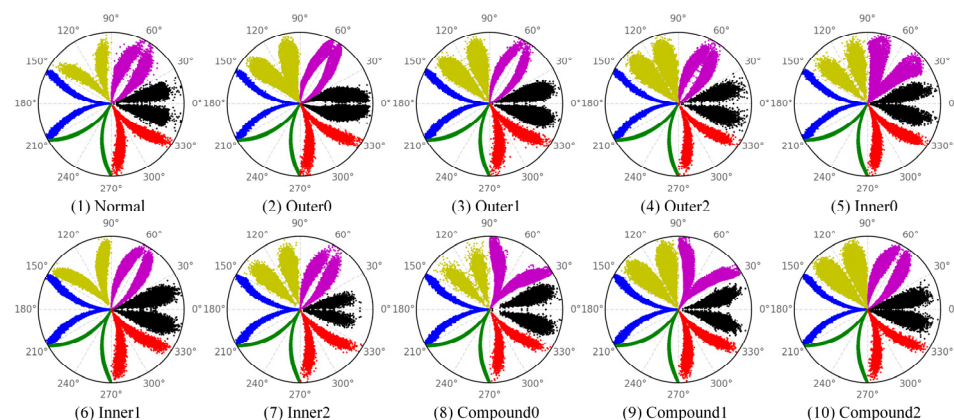


Figure 6. SDP images of different fault states.

Table 3. Hyperparameters of the model.

Description	Parameters	Value
RepLKNet Stage 1	B, C, K	(2, 128, 31)
RepLKNet Stage 2	B, C, K	(2, 256, 29)
RepLKNet Stage 3	B, C, K	(18, 512, 27)
RepLKNet Stage 4	B, C, K	(2, 1024, 13)
BiGRU layer	number of layers	2
BiGRU units	hidden layer units	(32, 64)
Learning rate	lr	0.0003
Attention	head	1
Dropout	dropout	0.3

Table 3. Cont.

Description	Parameters	Value
Min-batch learning	batch size	32
Maximum iteration times	epoch	100
Number of sample classes	class	10
Optimization	optimizer	Adam
Loss function	loss	cross-entropy

4.3. Experimental Results and Analysis

To validate the effectiveness of the proposed method, four experiments are conducted on the Paderborn dataset, including a multi-source information fusion experiment, a comparative experiment, an ablation experiment, and a robustness against noise experiment.

4.3.1. Multi-Source Information Fusion Experiment

To evaluate the effectiveness of different input sources, we conduct comparative experiments using vibration signals, current signals, and fused data. To further verify the roles of the RepLKNet and BiGRU architectures, each signal type is tested across three model configurations: the full proposed method, the standalone RepLKNet architecture, and the standalone BiGRU architecture. This results in a total of nine combinations. Notably, when using vibration signals or current signals as inputs for the proposed method, two identical copies of the same signal are used to construct the fused input. For fused data, the single RepLKNet employs only the SDP-based data-level fusion module, while the single BiGRU utilizes only the information entropy-based feature channel fusion module. The fault diagnosis results for all nine combinations are summarized in Figure 7, with their corresponding confusion matrices presented in Figure 8.

As shown in Figure 7, the comparison shows that the fault diagnosis accuracy using current signals is significantly lower than that using vibration signals. Fused data significantly outperforms both individual signal types. For example, in the proposed model, vibration signals achieve an accuracy of 96.88%, while current signals achieve 78.54%. When the fused data is used, the accuracy increases to 99.38%. This indicates that fused data effectively integrates complementary information from different sources, contains richer fault-related information, and greatly improves diagnostic accuracy.

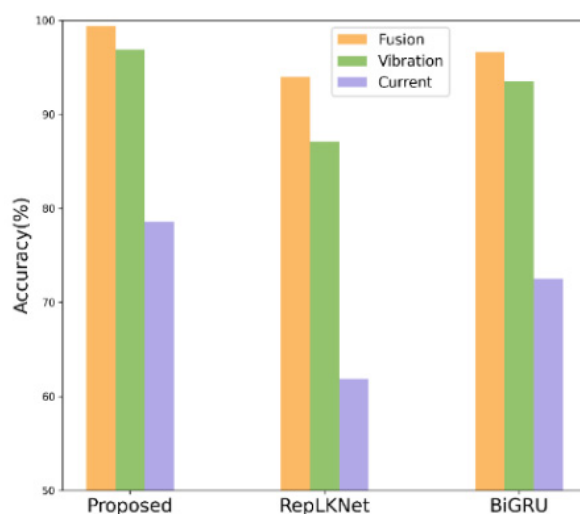


Figure 7. Diagnostic results of different models under different data.

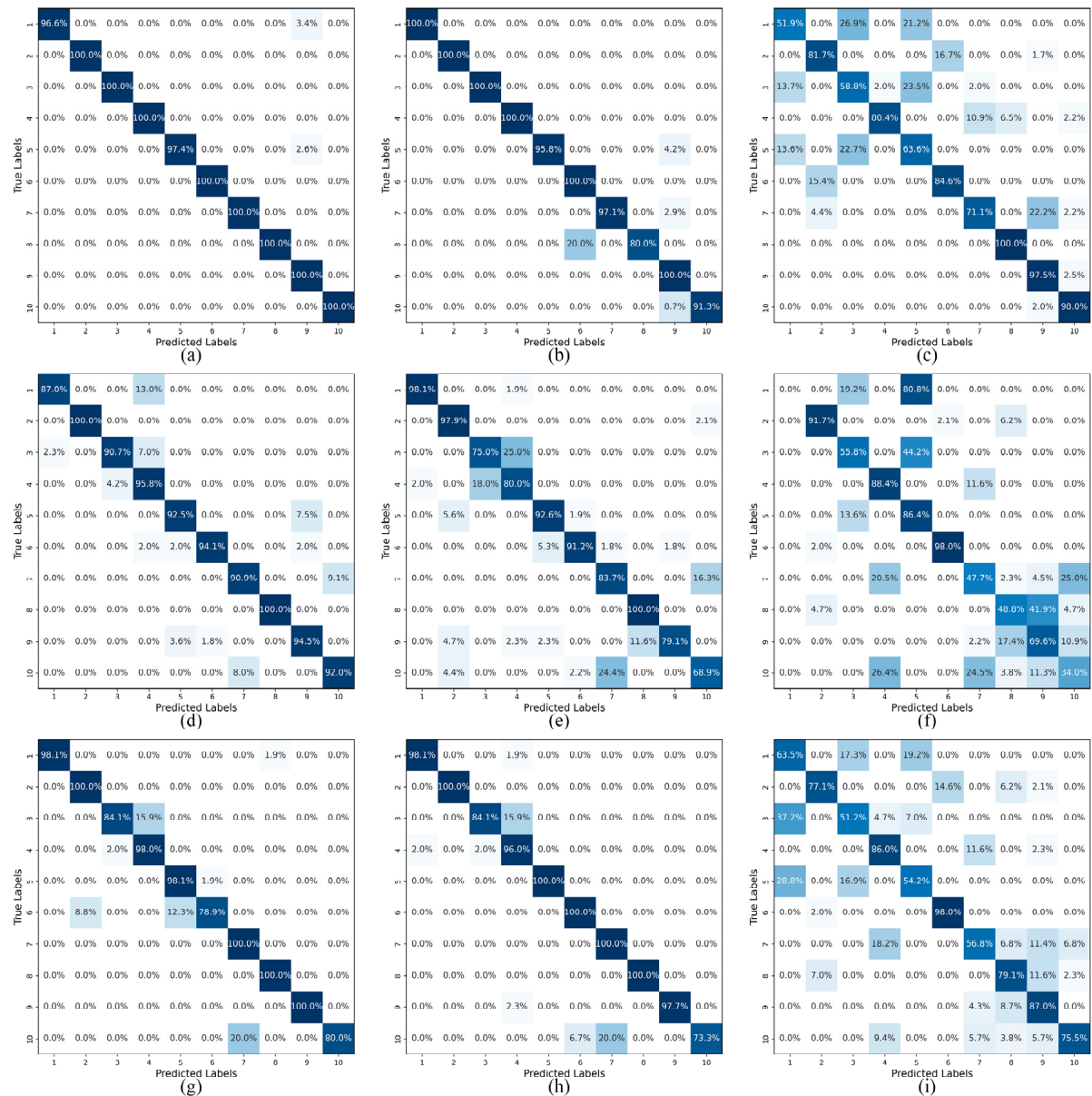


Figure 8. Confusion matrices of various models under different data; (a) Proposed model with fusion data; (b) Proposed model with vibration signal; (c) Proposed model with current data; (d) RepLKNet with fusion data; (e) RepLKNet with vibration signal; (f) RepLKNet with current signal; (g) BiGRU with fusion data; (h) BiGRU with vibration signal; (i) BiGRU with current signal.

As illustrated in the confusion matrix presented in Figure 8, under varying data type conditions, the performance of either the RepLKNet or BiGRU model used independently is suboptimal compared to the proposed method. This suggests that individual models have limitations in feature extraction. In contrast, the proposed model combines RepLKNet and BiGRU to leverage RepLKNet’s sensitivity to SDP image shapes and BiGRU’s ability to capture temporal correlations. This integration provides complementary multi-modal features for downstream modules. Compared to the individual components, the combined approach significantly improves diagnostic accuracy. Therefore, the model proposed in this paper effectively integrates the advantages of RepLKNet and BiGRU and comprehensively utilizes multi-source data and the correlations between features of different models, thereby maximizing diagnostic accuracy.

4.3.2. Comparative Experiment

To empirically validate the effectiveness of our method, we compared it with state-of-the-art diagnosis methods, including AMDC-CNN [30], MRSFN [31], Matrix-CNN [7], MCFCNN [32], CWT-RepLKNet [33], and SDP-CNN [17]. Among these models, AMDC-CNN and MRSFN employ feature-level fusion for multi-source information integration. Matrix-CNN and MCFCNN incorporate both data-level and feature-level fusion strategies. CWT-RepLKNet and SDP-CNN rely solely on single-source data. To evaluate diagnostic performance, accuracy, precision, recall, and F1-score were selected as evaluation metrics for assessing the effectiveness of fault diagnosis. To ensure reproducibility, we re-implemented all comparative models and followed the experimental settings and hyperparameters listed in Table 3. All methods were tested under the standardized conditions of the Paderborn dataset, using consistent preprocessing and model configurations. The comparative performance metrics for all methods are summarized in Table 4.

Table 4. Fault diagnosis metrics for different models.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
AMDC-CNN	87.08	86.93	86.56	86.56
MRSFN	91.78	92.08	92.13	92.02
Matrix-CNN	96.49	96.69	96.60	96.60
MCFCNN	95.21	95.16	95.03	95.07
CWT-RepLKNet	98.80	97.95	97.86	97.90
SDP-CNN	97.71	98.07	98.10	97.99
Proposed	99.38	99.39	99.37	99.37

As shown in Table 4, AMDC-CNN and MRSFN exhibit inferior performance due to their use of single fusion strategies. AMDC-CNN applies 1D convolutions to extract features from multi-sensor data, followed by simple concatenation, which leads to unreliable feature selection and lower diagnostic accuracy. MRSFN employs CNN and LSTM architectures for richer feature extraction, yielding better performance than AMDC-CNN but still falling short of the proposed method. By adopting a multi-level fusion strategy that incorporates both data-level and feature-level fusion, the proposed model overcomes the limitations of single fusion approaches and significantly improves diagnostic performance.

Matrix-CNN and MCFCNN perform data-level fusion by stacking the original signals into two-dimensional matrices or treating them as separate channels. Despite demonstrating superior performance compared to AMDC-CNN and MRSFN, which rely solely on feature fusion methods, they primarily capture local relationships between adjacent data sources and fail to fully exploit deeper cross-source correlations. This limitation restricts their classification performance. CWT-RepLKNet and SDP-CNN extract features by converting single-source data into two-dimensional images. While these methods achieve relatively high diagnostic accuracy, their reliance on single-source data makes them more vulnerable to noise interference. The proposed model outperforms all comparison methods in fault classification across four evaluation metrics. By integrating complementary information from multiple sources, it enhances both feature representation and diagnostic accuracy.

In summary, these limitations of the comparison methods prevent them from capturing interactions between information sources and often cause them to overlook cross-level feature correlations, resulting in constrained expressiveness and suboptimal performance. In this study, we propose a method that constructs a multi-level information fusion framework. By systematically considering the interrelationships among different information sources, the proposed framework effectively addresses these limitations. Through the implementation of a multi-level fusion strategy, it not only overcomes the constraints of

single-stage fusion methods but also significantly enhances the representational capacity of the final fused result.

4.3.3. Ablation Experiment

In this section, ablation experiments were conducted to verify the effectiveness of the three key modules proposed in this work. Three experimental scenarios were defined as (a) removing the SDP-based data-level fusion module during the data fusion stage; (b) disabling the cross-attention-based multi-view feature fusion strategy in the feature fusion stage; and (c) generating the final output features without the information entropy-based feature fusion module. Each of the three scenarios was tested, and the results were compared to the proposed method. The comparison results are presented in Table 5.

Table 5. Fault diagnosis metrics for different experimental scenarios.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Proposed	99.38	99.39	99.37	99.37
Strategy A	97.29	97.32	97.26	97.27
Strategy B	95.79	95.73	96.05	95.82
Strategy C	98.54	98.43	98.53	98.46

The results of the ablation experiments show that, compared to the proposed method, the performance evaluation metrics in all three experimental scenarios have decreased. Specifically, the removal of the SDP-based data-level fusion module results in a 2.18% decrease in the model's accuracy. Without multi-source data fusion, the network relies solely on single-level features, resulting in incomplete feature representations and lower diagnostic accuracy. The method proposed in this paper effectively enhances the discrimination of different fault states by fusing multi-source information at the data level, thereby significantly improving diagnostic accuracy.

Removing the cross-attention-based feature fusion method decreased diagnostic accuracy by 3.59% when simple concatenation was used. This approach failed to capture feature relationships and underutilized complementary information across modalities. We propose a cross-attention-based multi-view feature fusion strategy that maps multi-modal features into a unified latent space, leveraging their complementarity to enhance representation capability. Similarly, mean-weighted channel feature fusion reduced accuracy by 0.84%, amplifying the influence of unreliable channels on classification. Our proposed channel fusion module dynamically adjusts weights based on channel reliability, improving accuracy and robustness.

Additionally, to better demonstrate the effectiveness of the three key modules proposed in this paper, t-SNE (t-Distributed Stochastic Neighbor Embedding) is used to visualize the results for each channel. The channels are defined as follows: Channel 1 represents raw vibration data; Channel 2 represents raw current data; Channel 3 represents features extracted from SDP images using RepLKNet; Channel 4 represents the fusion of vibration signal features and image features via a cross-attention-based multi-view feature fusion strategy; Channel 5 represents the fusion of current signal features and image features via a cross-attention-based multi-view feature fusion strategy, and the output of the information entropy-based feature channel fusion module is denoted as Final Fusion.

As illustrated in Figure 9, the raw vibration and current signals exhibit random distributions with no clear class separation. Channel 3 shows some preliminary classification among different categories, but the effect is not very distinct. Channels 4 and 5, which adopt the feature fusion strategy, fuse multi-modal features. The observation results indicate that Channel 4, primarily utilizing vibration signal features, demonstrates excellent perfor-

mance in the classification task. Channel 5, primarily relying on current signal features, exhibits unclear classification boundaries among categories, although some sample types still show certain clustering tendencies. By fusing the temporal features of each channel with the SDP image features, both Channels 4 and 5 exhibit relatively excellent classification performance, which is particularly evident when compared with Channels 1 and 2. This highlights the effectiveness of the SDP-based multi-source data and cross-attention-based multi-view feature fusion strategy. The final fused result, shown in Figure 9f, achieves near-optimal clustering, with distinct class boundaries and only a few scattered samples. This clearly demonstrates the effectiveness of the information entropy-based feature channel fusion module.

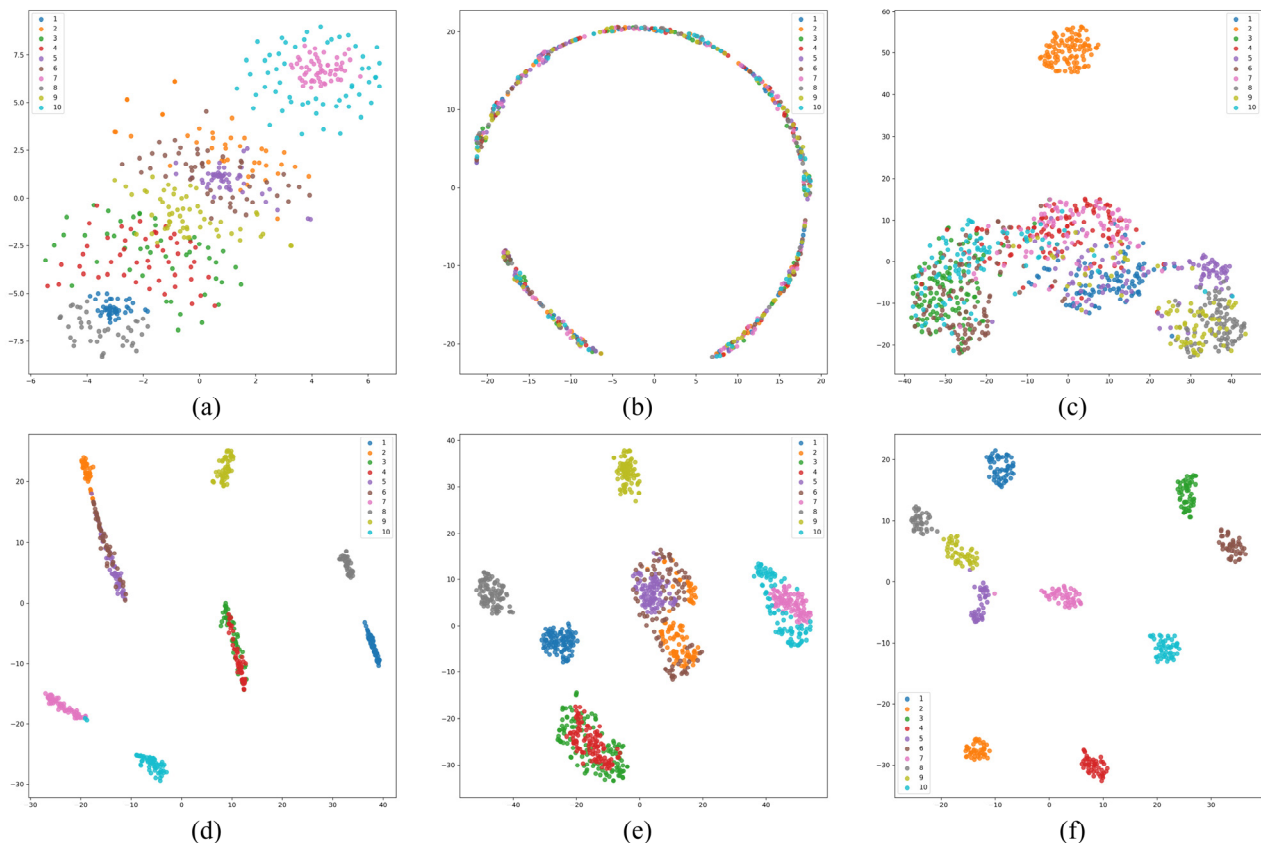


Figure 9. Visualization of the results of different channels on the test set. (a) Channel 1 (b) Channel 2 (c) Channel 3 (d) Channel 4 (e) Channel 5 (f) final fusion.

In summary, the ablation experiments confirmed the necessity of each component in the construction of a multi-level fusion framework. The multi-level fusion strategy proposed in this paper overcomes the limitations of relying on a single component. Through the coordinated operation of modules, the strengths of each individual module are fully utilized. This design enhances both the efficiency of multi-source information fusion and the accuracy of fault diagnosis. Compared with methods that only rely on a single fusion level, this framework has the advantages of unity, hierarchy, and being end-to-end. It can effectively integrate data, features, and channel information from multiple sources, which cannot be achieved by using only a single fusion method.

4.3.4. Robustness Against Noise Experiment

In this section, we evaluate the noise robustness of the proposed method by designing noise scenarios with varying signal-to-noise ratios (SNR). Gaussian white noise is commonly found in practical production environments and effectively simulates the impact of

noise under real-world operational conditions. To generate noisy signals, Gaussian white noise was added to the raw signals at specific SNR levels. The SNR is the logarithmic relationship between the original signal's power and the noise's power:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \quad (15)$$

where P_{signal} represents the power of the signal, and P_{noise} represents the power of the noise.

In Figure 10, Gaussian white noise is added to the raw compound fault signals. At an SNR of 0 dB (where the noise power equals the signal power), the raw signal is severely corrupted by noise, as clearly shown in Figure 10. To evaluate the robustness of the model against noise, we added noise signals with SNRs ranging from -5 dB to 10 dB in 5 dB increments to the vibration signals from the Paderborn dataset, generating four distinct noisy signal datasets. The performance comparison of various models under different noise conditions is presented in Table 6.

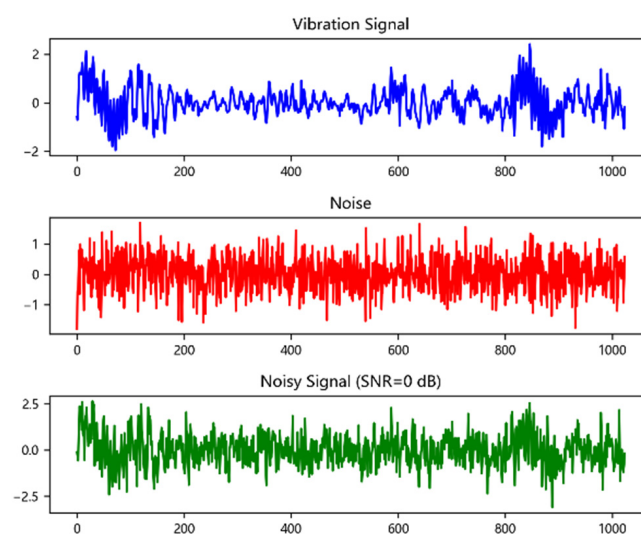


Figure 10. The original signal, Gaussian noise and noise signal diagram at SNR = 0.

Table 6. Diagnostic results on the Paderborn dataset under different SNRs (%).

Methods	SNR			
	−5	0	5	10
AMDC-CNN	61.39	75.67	82.16	85.69
MRSFN	68.42	76.15	87.75	90.77
Matrix-CNN	77.72	83.08	89.68	92.66
MCFCNN	80.10	86.18	91.28	94.81
CWT-RepLKNet	43.15	48.85	81.34	89.31
SDP-CNN	61.15	72.29	88.69	94.93
Proposed	87.23	95.79	98.43	99.11

The performance results of different models on noisy signal datasets with varying SNR levels clearly show that noise significantly affects diagnostic accuracy. As the noise level increases, the diagnostic accuracy of all models declines. The CWT-RepLKNet and SDP-CNN methods exhibit a substantial decrease in accuracy compared to their performance on the Paderborn dataset. Specifically, when the SNR is above 0 dB, their accuracy decreases by less than 10% . However, when the SNR is 0 dB or lower, the accuracy drops by over 30% . This vulnerability arises from their reliance on a single information source, which significantly reduces their robustness to noise interference. Among the comparative

methods, AMDC-CNN and MRSFN perform worse than Matrix-CNN and MCFCNN under noisy conditions. This discrepancy occurs because AMDC-CNN and MRSFN rely on single-level fusion, causing their outputs to be excessively dependent on the quality of individual features, which can be easily corrupted by noise. In contrast, the proposed model shows superior noise robustness, achieving accuracies of 99.11%, 98.43%, 95.79%, and 87.23% across four noise scenarios. It leverages multi-source signals and feature interactions to compensate for feature loss caused by noise, enhancing fusion performance and noise robustness.

5. Conclusions

This study proposes a novel bearing fault diagnosis method based on a multi-level fusion framework for multi-source information. To overcome the limitations of traditional single-source approaches, we first introduce an SDP-based data-level fusion module. This module integrates multi-source data into unified SDP images, which improves the discrimination of different fault features and enables efficient fusion of multi-source information. Next, we design a cross-attention-based multi-view feature fusion strategy to fuse feature information from different modalities and information sources. It makes full use of the complementarity of different features and enhances the representational capability of the fused features. Finally, we propose an information entropy-based channel fusion module. It assigns higher weights to more informative channels, enhancing feature utilization and improving model robustness. The proposed framework offers a unified, hierarchical, and end-to-end solution that effectively integrates multi-source data, features, and channels, thereby enhancing information fusion efficiency and maximizing the performance of individual modules. Through comprehensive comparisons with existing methods on both public datasets and noisy signal datasets, the proposed method clearly demonstrates its superiority and robustness under noise interference.

Although the proposed method demonstrates strong performance in bearing fault diagnosis under specific operating conditions, it has certain limitations. While the current SDP representation effectively captures key features from multiple sources, information loss or redundancy may occur as the number of sensors increases. Such losses can weaken the discrimination ability of downstream modules, reducing the model's robustness and accuracy. To address this issue, we propose a potential solution based on the SDP-based data-level fusion module. The solution builds a multi-layer SDP graph data, where each layer represents a signal source or its IMF. It leverages sensor spatial correlations to connect layers, achieving effective multi-source data fusion.

The model trained under specific conditions shows limited generalizability when applied to unseen or varying operating environments. In future research, we intend to improve the model's cross-domain adaptability by integrating transfer learning techniques and validating its performance using real-world industrial multi-source datasets. This approach will enable the model to generalize more effectively across different types of machinery and operational environments, decrease its dependence on large labeled datasets, and improve its adaptability and efficiency.

Author Contributions: Conceptualization, X.D. and X.P.; methodology, X.D. and Y.S.; software, Y.S.; validation, X.D., Y.S. and L.L.; formal analysis, Y.S. and L.L.; investigation, X.D. and Y.S.; resources, X.P.; data curation, X.D.; writing—original draft preparation, X.D. and Y.S.; writing—review and editing, X.D., Y.S., L.L. and X.P.; visualization, Y.S.; supervision, X.P.; project administration, X.P.; funding acquisition, X.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Key Technologies for Intelligent Monitoring and Analysis of Equipment Health Status of High of Science and Technology Innovation Team in College of Hunan Province, grant numbers 2023-233 (Xiang Jiao Tong), in part by Natural Science Foundation of Hunan Province, grant numbers 2024JJ7091, 2024JJ7092, 2024JJ7093 and 2024JJ7148, and in part by Key Project of Hunan Provincial Department of Education, grant numbers 24A0671.

Data Availability Statement: This article uses publicly available datasets.

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

References

1. Zhang, P.; Chen, R.; Xu, X.; Yang, L.; Ran, M. Recent Progress and Prospective Evaluation of Fault Diagnosis Strategies for Electrified Drive Powertrains: A Comprehensive Review. *Measurement* **2023**, *222*, 113711. [\[CrossRef\]](#)
2. Huang, R.; Xia, J.; Zhang, B.; Chen, Z.; Li, W. Compound Fault Diagnosis for Rotating Machinery: State-of-the-Art, Challenges, and Opportunities. *J. Dyn. Monit. Diagn.* **2023**, *2*, 13–29. [\[CrossRef\]](#)
3. Choudhary, A.; Mishra, R.K.; Fatima, S.; Panigrahi, B.K. Multi-Input CNN Based Vibro-Acoustic Fusion for Accurate Fault Diagnosis of Induction Motor. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105872. [\[CrossRef\]](#)
4. Dao, F.; Zeng, Y.; Qian, J. Fault Diagnosis of Hydro-Turbine via the Incorporation of Bayesian Algorithm Optimized CNN-LSTM Neural Network. *Energy* **2024**, *290*, 130326. [\[CrossRef\]](#)
5. Yu, Z.; Zhang, C.; Deng, C. An Improved GNN Using Dynamic Graph Embedding Mechanism: A Novel End-to-End Framework for Rolling Bearing Fault Diagnosis Under Variable Working Conditions. *Mech. Syst. Signal Process.* **2023**, *200*, 110534. [\[CrossRef\]](#)
6. Kibrete, F.; Engida Woldemichael, D.; Shimels Gebremedhen, H. Multi-Sensor Data Fusion in Intelligent Fault Diagnosis of Rotating Machines: A Comprehensive Review. *Measurement* **2024**, *232*, 114658. [\[CrossRef\]](#)
7. Azamfar, M.; Singh, J.; Bravo-Imaz, I.; Lee, J. Multisensor Data Fusion for Gearbox Fault Diagnosis Using 2-D Convolutional Neural Network and Motor Current Signature Analysis. *Mech. Syst. Signal Process.* **2020**, *144*, 106861. [\[CrossRef\]](#)
8. Yang, J.; Gao, T.; Zhang, H.; Li, Y. A Multi-Sensor Fault Diagnosis Method for Rotating Machinery Based on Improved Fuzzy Support Fusion and Self-Normalized Spatio-Temporal Network. *Meas. Sci. Technol.* **2023**, *34*, 125112. [\[CrossRef\]](#)
9. Xie, T.; Huang, X.; Choi, S.-K. Intelligent Mechanical Fault Diagnosis Using Multisensor Fusion and Convolution Neural Network. *IEEE Trans. Ind. Inform.* **2022**, *18*, 3213–3223. [\[CrossRef\]](#)
10. Dong, K.; Lotfipoor, A. Intelligent Bearing Fault Diagnosis Based on Feature Fusion of One-Dimensional Dilated CNN and Multi-Domain Signal Processing. *Sensors* **2023**, *23*, 5607. [\[CrossRef\]](#)
11. Xiao, X.; Li, C.; He, H.; Huang, J.; Yu, T. Rotating Machinery Fault Diagnosis Method Based on Multi-Level Fusion Framework of Multi-Sensor Information. *Inf. Fusion* **2025**, *113*, 102621. [\[CrossRef\]](#)
12. Gao, S.; Noman, K.; Mao, G.; Deng, Z.; Li, Y.; Ge, W. Multi-Source Heterogeneous Information Fusion Based on Graph Convolutional Network for Gearbox Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3547515. [\[CrossRef\]](#)
13. Gao, L.; Liu, Z.; Gao, Q.; Li, Y.; Wang, D.; Lei, H. Dual Data Fusion Fault Diagnosis of Transmission System Based on Entropy Weighted Multi-Representation DS Evidence Theory and GCN. *Measurement* **2025**, *243*, 116308. [\[CrossRef\]](#)
14. Shao, H.; Lin, J.; Zhang, L.; Galar, D.; Kumar, U. A Novel Approach of Multisensory Fusion to Collaborative Fault Diagnosis in Maintenance. *Inf. Fusion* **2021**, *74*, 65–76. [\[CrossRef\]](#)
15. Zhang, K.; Li, H.; Cao, S.; Lv, S.; Yang, C.; Xiang, W. Trusted multi-source information fusion for fault diagnosis of electromechanical system with modified graph convolution network. *Adv. Eng. Inform.* **2023**, *57*, 102088. [\[CrossRef\]](#)
16. Tang, Y.; Zhang, X.; Huang, S.; Qin, G.; He, Y.; Qu, Y.; Xie, J.; Zhou, J.; Long, Z. Multisensor-Driven Motor Fault Diagnosis Method Based on Visual Features. *IEEE Trans. Ind. Inform.* **2023**, *19*, 5902–5914. [\[CrossRef\]](#)
17. Huang, F.; Zhang, X.; Qin, G.; Xie, J.; Peng, J.; Huang, S.; Long, Z.; Tang, Y. Demagnetization Fault Diagnosis of Permanent Magnet Synchronous Motors Using Magnetic Leakage Signals. *IEEE Trans. Ind. Inform.* **2023**, *19*, 6105–6116. [\[CrossRef\]](#)
18. Yang, F.; Tian, X.; Ma, L.; Shi, X. An Optimized Variational Mode Decomposition and Symmetrized Dot Pattern Image Characteristic Information Fusion-Based Enhanced CNN Ball Screw Vibration Intelligent Fault Diagnosis Approach. *Measurement* **2024**, *229*, 114382. [\[CrossRef\]](#)
19. Chen, Q.; Li, C.; Ning, J.; Lin, S.; He, K. GMConv: Modulating Effective Receptive Fields for Convolutional Kernels. *IEEE Trans. Neural Netw. Learn. Syst.* **2025**, *36*, 6669–6678. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021. [\[CrossRef\]](#)
21. Ding, X.; Zhang, X.; Zhou, Y.; Han, J.; Ding, G.; Sun, J. Scaling up your kernels to 31×31 : Revisiting large kernel design in CNNs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022. [\[CrossRef\]](#)

22. Khan, Z.Y.; Niu, Z. CNN with Depthwise Separable Convolutions and Combined Kernels for Rating Prediction. *Expert Syst. Appl.* **2021**, *170*, 114528. [[CrossRef](#)]
23. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. RepVGG: Making VGG-Style ConvNets Great Again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021. [[CrossRef](#)]
24. Tuli, S.; Dasgupta, I.; Grant, E.; Griffiths, T.L. Are Convolutional Neural Networks or Transformers more like human vision? *arXiv* **2021**, arXiv:2105.07197. [[CrossRef](#)]
25. Li, H.; Wu, X.; Liu, T.; Li, S.; Zhang, B.; Zhou, G.; Huang, T. Composite Fault Diagnosis for Rolling Bearing Based on Parameter-Optimized VMD. *Measurement* **2022**, *201*, 111637. [[CrossRef](#)]
26. Fu, Y.; Chen, X.; Liu, Y.; Son, C.; Yang, Y. Multi-source information fusion fault diagnosis for gearboxes based on SDP and VGG. *Appl. Sci.* **2022**, *12*, 6323. [[CrossRef](#)]
27. Han, H.; Cho, S.; Kwon, S.; Cho, S.-B. Fault diagnosis using improved complete ensemble empirical mode decomposition with adaptive noise and power-based intrinsic mode function selection algorithm. *Electronics* **2018**, *7*, 16. [[CrossRef](#)]
28. Xu, X.; Liu, H.; Zhu, H.; Wang, S. Fan fault diagnosis based on symmetrized dot pattern analysis and image matching. *J. Sound Vib.* **2016**, *374*, 297–311. [[CrossRef](#)]
29. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification. *PHM Soc. Eur. Conf.* **2016**, *3*, 1–17. [[CrossRef](#)]
30. Wang, D.; Li, Y.; Jia, L.; Song, Y.; Liu, Y. Novel Three-Stage Feature Fusion Method of Multimodal Data for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3514710. [[CrossRef](#)]
31. 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](#)]
32. Li, H.; Huang, J.; Gao, M.; Yang, L.; Bao, Y. Multi-View Information Fusion Fault Diagnosis Method Based on Attention Mechanism and Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 11410. [[CrossRef](#)]
33. Zhu, H.; Fang, T. Compound Fault Diagnosis of Rolling Bearings with Few-Shot Based on DCGAN-RepLKNet. *Meas. Sci. Technol.* **2024**, *35*, 066105. [[CrossRef](#)]

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.