

Article

Dynamic Condition-Matching Network for End-to-End Machinery Health Monitoring Under Time-Varying Operating Conditions

Jiangling Wang ¹, Juntao Wang ¹, Caiming Zhong ¹  and Yu Tian ^{1,2,*} ¹ College of Science & Technology, Ningbo University, Ningbo 315300, China² School of Mechanical Engineering, Tongji University, Shanghai 201814, China

* Correspondence: tianyu@nbu.edu.cn

Abstract

Dynamic monitoring of critical machinery under time-varying operating conditions remains a challenging problem. Existing methods often fail to efficiently and conveniently generate reliable health indicators. To address this issue, a Dynamic Condition-Matching Network (DCMN) is proposed to end-to-end generate machinery health indicators under time-varying operating conditions. First, a self-supervised matching-pair-based sample generation method is developed. Real and virtual vibration–speed pairs are constructed to produce pseudo-regression labels for exploring latent cross-modal relationships. Second, cross-attention is introduced into DCMN to model the vibration–speed cross-modal representation association. Fused features that reflect deviation from the reference baseline are thereby generated. Finally, a cross-modal matching-rate prediction head is established in DCMN. The health indicators under given vibration and speed inputs can be computed in an end-to-end manner. The effectiveness of DCMN is validated through a case study under time-varying operating conditions. Experimental results demonstrate that DCMN can effectively generate condition-invariant health indicators. It outperforms the compared methods and provides an elegant approach for machinery health monitoring under time-varying conditions.

Keywords: health monitoring; time-varying condition; variant speeds; cross attention; self-supervised learning

1. Introduction

Dynamic health monitoring of machinery remains a persistent challenge in the prognostics and health management field. In practical industrial scenarios, the operating conditions of critical equipment are not fixed. Frequent and large fluctuations in operating conditions often lead to the failure of conventional models. In addition, frequent false alarms and missed detections may cause substantial economic losses, which is unacceptable in engineering practice and remains an urgent problem [1].

Machinery health monitoring is widely applied in engineering practice. Real-time operating states are obtained based on various collected signals, and the operating conditions of critical components are tracked to ensure continuous and safe production [2]. Traditional methods are typically established based on statistical models. Various simple and efficient indicators are calculated from the measured signals to provide a preliminary multi-dimensional assessment. However, different statistical measures carry different



Academic Editor: Sheng Du

Received: 30 June 2026

Revised: 20 September 2026

Accepted: 23 September 2026

Published: 24 September 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

physical meanings and vary in sensitivity to specific phenomena. Therefore, joint interpretation with expert knowledge is usually required [3]. Over the past decade, deep learning has attracted significant attention in the field of health management [4]. Duan et al. [5] proposed a spatio-temporal fusion method (SALICAE) integrating self-attention and LSTM to construct unsupervised health indicators (HIs). Li et al. [6] developed a spatiotemporal enhanced convolutional autoencoder (STECAE) to mine local spatial and global temporal features directly from raw data. Koutroulis et al. [7] introduced an anticausal learning framework to ensure robust remaining useful life (RUL) prediction under time-varying conditions. However, these approaches may lack robustness when machinery operates under highly dynamic conditions. To address this, various schemes have been developed for time-varying monitoring. Xu et al. [8] developed an enhanced stacked autoencoder (ESAE) using exponential weight moving average to eliminate noise from unlabeled signals. For frequent speed fluctuations, Liu et al. [9] designed an EPSDNet model. This model utilizes a nonlinear regression network, Spyder, to identify the mapping between vibration signals and speeds for adaptive feature alignment. Within the transfer learning framework, Zhang et al. [10] developed a latent dependency relationship-aware network (LDRN) to preserve cross-domain spectral consistency. Furthermore, Liang et al. [11] proposed a semisupervised subdomain adaptation graph convolutional network (SSAGCN) to reduce distribution differences between source and target domains. To mitigate feature coupling, Chen et al. [12] presented a sliding mixing graph contrastive domain adaptation (Mix-GraphCDA) method. This approach employs contrastive learning to maximize similarity between domains and their mixed views, ensuring high diagnostic accuracy in dynamic environments. However, existing methods often struggle with non-stationary conditions or rely on overly complex architectures, which increases the computational cost and reduces the flexibility in industrial deployment. A more lightweight and adaptive framework is therefore required.

The cross-attention mechanism has attracted wide attention in Prognostics and Health Management (PHM) owing to its capability of multi-source information alignment and feature fusion. Yang et al. [13] proposed a model combining temporal convolutional network (TCN)-Transformer with Kolmogorov-Arnold networks (KCA) to capture degradation features for accurate RUL prediction. To handle strong noise, Li et al. [14] developed the MSIF-Convformer framework to achieve deep fusion of vibration and acoustic signals. Yu et al. [15] presented the MSIFT framework, which utilizes cross-attention to guide feature extraction and mitigate label bias under imbalanced data. Furthermore, Zhu et al. [16] integrated synergy graphs with Transformers to mine intrinsic correlations between different feature modalities. In the field of transfer learning, Liu et al. [17] proposed an explainable cross-attention-driven method to achieve feature alignment between source and target domains. Yu et al. [18] utilized cross-attention for spatio-temporal fusion to optimize fault identification in rotating machinery. Additionally, Han et al. [19] constructed a simulation-to-real transfer learning framework to synergize physical modeling and data-driven techniques. However, these cross-attention-based approaches still face certain limitations. The high computational complexity of cross-attention modules poses challenges for real-time monitoring with large-scale sensor data. Moreover, existing studies often focus on specific scenarios and lack a universal modeling mechanism for cross-modal interactions under highly variable conditions.

Self-supervised learning (SSL), especially contrastive learning, has emerged as a powerful paradigm for feature representation by leveraging massive unlabeled data. In the computer vision domain, Doersch et al. [20] introduced a context prediction task to learn visual representations by predicting the relative positions of image patches. He et al. [21] developed Masked Autoencoders (MAE) that reconstruct original signals from a high

proportion of masked patches. These foundational works have paved the way for intelligent fault diagnosis. Chen et al. [22] proposed the SimCLR framework, which demonstrates that the composition of data augmentations and nonlinear transformations significantly improves the quality of learned features. To further enhance scalability, Peng et al. [23] presented a supervised contrastive learning method with negative out-of-distribution data augmentation to address open-set fault diagnosis. Similarly, Wan et al. [24] proposed a self-supervised simple Siamese framework (SSF) to extract distinguishable fault features using only unlabeled vibration samples. However, most contrastive learning methods in PHM rely on manual augmentation strategies, often require large batch sizes or elaborate negative-sample mining to converge, and lack explicit physical constraints, so the learned representations may be sensitive to operational noise rather than health states.

In addition, deep adversarial and domain-adaptation learning have been explored to extract speed-invariant features, where adversarial discriminators or subdomain alignment reduce the distribution discrepancy across operating conditions. These methods, however, often rely on labeled target-domain data or complex adversarial training, which limits their flexibility in unsupervised health monitoring. A representative work is the quantitative estimation neural network (QENN) [25], which jointly trains a feature extractor, an adversarial speed discriminator, and a severity classifier to estimate the fault severity of rolling bearings under variable working conditions.

Although the above-mentioned approaches have achieved significant progress, there still exist several gaps in dynamic condition monitoring under time-varying conditions: (1) Modality fusion limitations: Simple feature concatenation fails to capture the complex bidirectional interactions between vibration patterns and fluctuating operating physics. (2) Difficulties in end-to-end health indicator generation: Indirect proxies like reconstruction error are highly sensitive to speed variations, leading to false alarms in healthy states. (3) Unbounded health indicators: Unbounded metrics lack a unified scale and physical interpretability, making universal threshold setting extremely difficult across different industrial scenarios.

In order to address these problems, a novel Dynamic Condition-Matching Network (DCMN) is proposed to implement end-to-end machinery health monitoring. DCMN performs deep feature interaction between vibration and speed. The main contributions are summarized as follows: (1) Self-supervised matching sample pair strategy: A novel paradigm uses artificially generated positive and negative pairs to learn speed-invariant representations using only healthy data, eliminating manual labeling. (2) Bidirectional cross-attention mechanism: A dedicated fusion module enables vibration and speed to query each other, capturing intricate cross-modal dependencies that simple concatenation misses. (3) End-to-end bounded health indicator prediction: An integrated prediction head directly outputs a matching rate in the $[0, 1]$ range, providing a unified, speed-invariant metric for elegant threshold management.

Experimental results demonstrate that DCMN significantly outperforms representative methods, and demonstrates competitive ball-fault detection and overall stability under variable speed conditions. The rest of this paper is organized as follows. The model construction of DCMN is presented in Section 2. Section 3 provides the test results on a time-varying speed case. Section 4 discusses the mechanism behind the results and the limitations, and Section 5 concludes this article.

2. Methodology

In this section, a novel dynamic condition-matching network (DCMN) is proposed for vibration–speed matching prediction in bearing fault diagnosis under varying operating conditions. The core innovation lies in the introduction of a bidirectional cross-attention

fusion mechanism that explicitly models the intrinsic correlation between vibration signals and rotational speed information, thereby enabling accurate matching rate prediction in a self-supervised manner.

DCMN comprises four primary components: a vibration encoder, a speed encoder, a cross-attention fusion module, and a prediction head of matching rate, as shown in Figure 1. The vibration encoder is constructed based on a 1-dimensional DenseNet [26] architecture, through which raw vibration signals of length 8192 are progressively encoded into a common representation space. This DenseNet architecture facilitates efficient feature reuse and enhances deep feature extraction for non-stationary vibration patterns. Meanwhile, speed information encoding is achieved via a multilayer perceptron (MLP), mapping scalar rotational speed values into the same feature space. Subsequently, a bidirectional cross-attention fusion module with multi-head attention (four heads) is employed to perform feature fusion. Specifically, vibration features attend to speed features and vice versa, and residual connections together with layer normalization are incorporated to stabilize training. Through this bidirectional cross-attention mechanism, complementary information across modalities is adaptively integrated, and the fused representation is forwarded to an MLP-based prediction head for matching rate estimation within the range of $[0, 1]$. A self-supervised learning strategy is adopted to train the DCMN without manual fault labels. Positive samples are constructed by pairing real vibration signals with their corresponding true speeds, while negative samples are generated by pairing the same vibration signals with randomly sampled speeds, where the pseudo matching rates are computed based on the true speed and corresponding sampled speeds. This forms a regression-based contrastive learning objective. For each sample, one positive and multiple negative pairs are generated, and the model is optimized using mean squared error between predicted and target matching rates. Through this process, normal vibration–speed matching patterns are learned exclusively from healthy data.

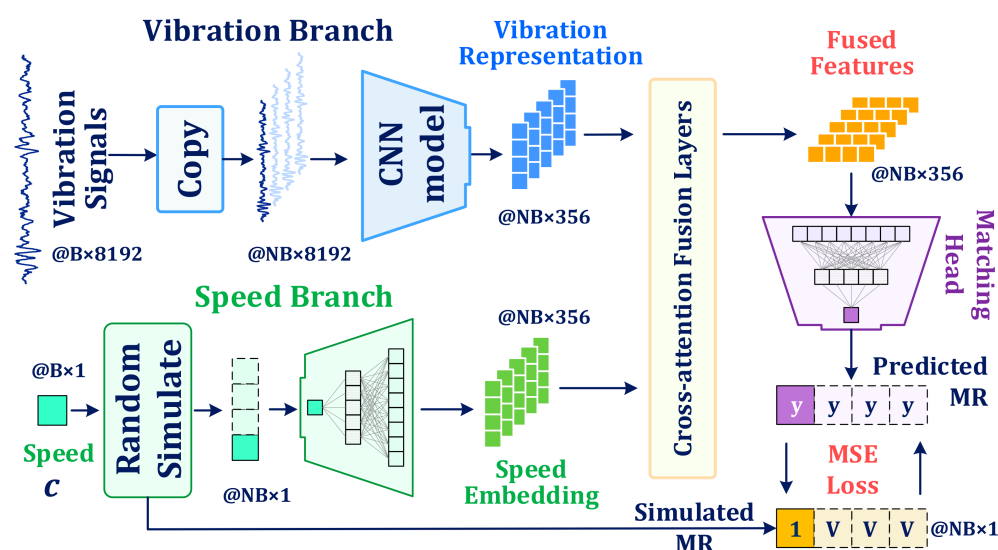


Figure 1. Diagram of DCMN.

During testing phase, the predicted matching rate can be directly used as the health indicator for machinery health monitoring under time-varying operating conditions. Lower matching rates correspond to higher abnormality levels. Fault detection is achieved by comparing the health indicator against a threshold computed from training data. The well-trained DCMN thus provides a unified framework that couples cross-modal fusion with self-supervised matching-rate prediction, offering a practical solution for dynamic fault detection of critical machinery.

2.1. Self-Supervised Matching Rates Generation

To construct training pairs without manual annotation, the DCMN dataset adopts a self-supervised learning paradigm with physics-guided matching rate labels, enabling contrastive learning between vibration signals and rotational speeds. As shown in Figure 2, instead of relying on human-labeled fault categories, supervision is automatically derived from the intrinsic physical relationship between vibration responses and operating speeds, thereby forming a continuous similarity metric for matching rate prediction.

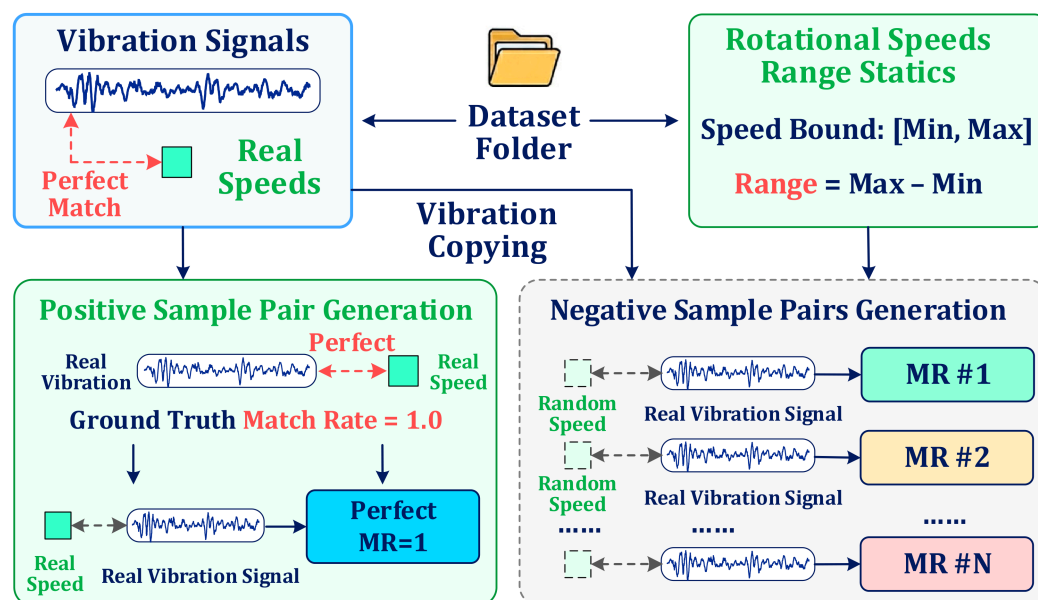


Figure 2. Self-supervised sample pairs generation for matching rates.

$Speed_{min} Speed_{max} Speed_{range} = Speed_{max} - Speed_{min} Speed_{true}(v, Speed_{true})$ First, the speed boundaries are computed by scanning the rotational speeds parsed from the file-names of all samples, yielding, and the corresponding normalization factor. The true speed associated with each vibration segment serves as the positive anchor, forming a positive pair that is assigned a perfect ground-truth matching rate of 1.0. This positive sample establishes the reference point for subsequent contrastive learning.

For each positive pair, N negative samples are generated via uniform random sampling, i.e., $Speed_{gen} \sim U(Speed_{min}, Speed_{max})$. These generated speeds are paired with the same vibration signal to form negative pairs, which introduce diverse contrastive signals across the entire speed range. Such uniform random sampling ensures broad coverage of operating conditions and enhances the robustness of representation learning.

The matching rate for each negative pair is calculated as:

$$MR = 1 - \frac{|Speed_{true} - Speed_{gen}|}{Speed_{range}} \quad (1)$$

which represents a normalized distance-based similarity. This physics-based formulation reflects the continuity of rotational speed: closer speeds imply higher similarity, whereas larger deviations result in lower matching rates. The resulting value is clamped via $\max(0.0, \min(1.0, MR))$, to guarantee a valid range of $[0, 1]$. Unlike binary contrastive labels, this continuous similarity metric enables regression-based training and preserves fine-grained relational information. Physically, within the operating-speed range of a given machine, the kinematic components of the vibration response vary smoothly and, to first order, linearly with the rotational speed, so a normalized linear distance is the simplest prior

consistent with this smoothness. A faulty vibration breaks the correspondence expected at its speed, which lowers the matching rate and thereby exposes the fault.

Finally, each batch is organized as $[B, 1 + N, L, 1]$ for vibration signals and $[B, 1 + N]$ for both speed and matching rate tensors, where each sample contains one positive and negative pairs. These real and generated matching rates serve as regression targets to train DCMN end to end. Through this self-supervised contrastive mechanism, the model learns the intrinsic vibration–speed matching relationship in a scalable and annotation-free manner.

2.2. Vibration and Speed Encoding Branches

The DCMN framework adopts a dual-branch encoding architecture to extract representative features from heterogeneous modalities, namely vibration signals and rotational speed values. Specifically, a 1D DenseNet-based vibration encoder [26] and a multilayer perceptron (MLP)-based speed encoder are designed to perform modality-specific feature extraction while projecting both inputs into a shared embedding space for subsequent cross-modal fusion.

The vibration encoder is built upon a 1D DenseNet with bottleneck layers and a predefined growth rate, as shown in Figure 3. Each bottleneck layer follows a normalization–activation–convolution scheme and is densely connected to all preceding layers. Through feature concatenation and progressive temporal down-sampling, the encoder captures multi-scale temporal dependencies of the vibration signal while reusing features and strengthening gradient propagation. The final features are aggregated into a compact representation of the dynamic characteristics of the mechanical system.

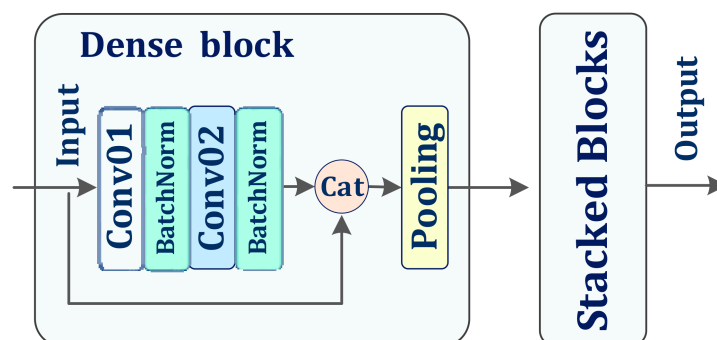


Figure 3. Network structure of vibration branch.

In contrast, the speed encoder employs a lightweight MLP structure tailored for scalar operating-condition inputs. It is illustrated in Figure 1. The network adopts a progressive expansion strategy to project low-dimensional speed values into a high-dimensional latent space. Nonlinear activation functions are incorporated to enhance representation capacity, while dropout regularization is applied to mitigate overfitting and improve generalization. This design ensures that speed information, although inherently simple, is sufficiently transformed to encode operating-condition characteristics relevant to vibration behavior.

The two encoders differ in complexity according to their inputs: the vibration branch uses deep convolutional modeling to extract temporal patterns, whereas the speed branch uses a compact fully connected structure to embed scalar information. Both branches output aligned features in a shared space, which enables the subsequent bidirectional cross-attention fusion.

2.3. Bidirectional Cross-Attention Fusion

The cross-attention fusion mechanism [27] of DCMN adopts a bidirectional design in which vibration and speed representations mutually query each other through independent cross-attention operations, followed by symmetric fusion. This formulation explicitly models the interdependence between signal dynamics and operating conditions, enabling refined cross-modal feature interaction, as shown in Figure 4.

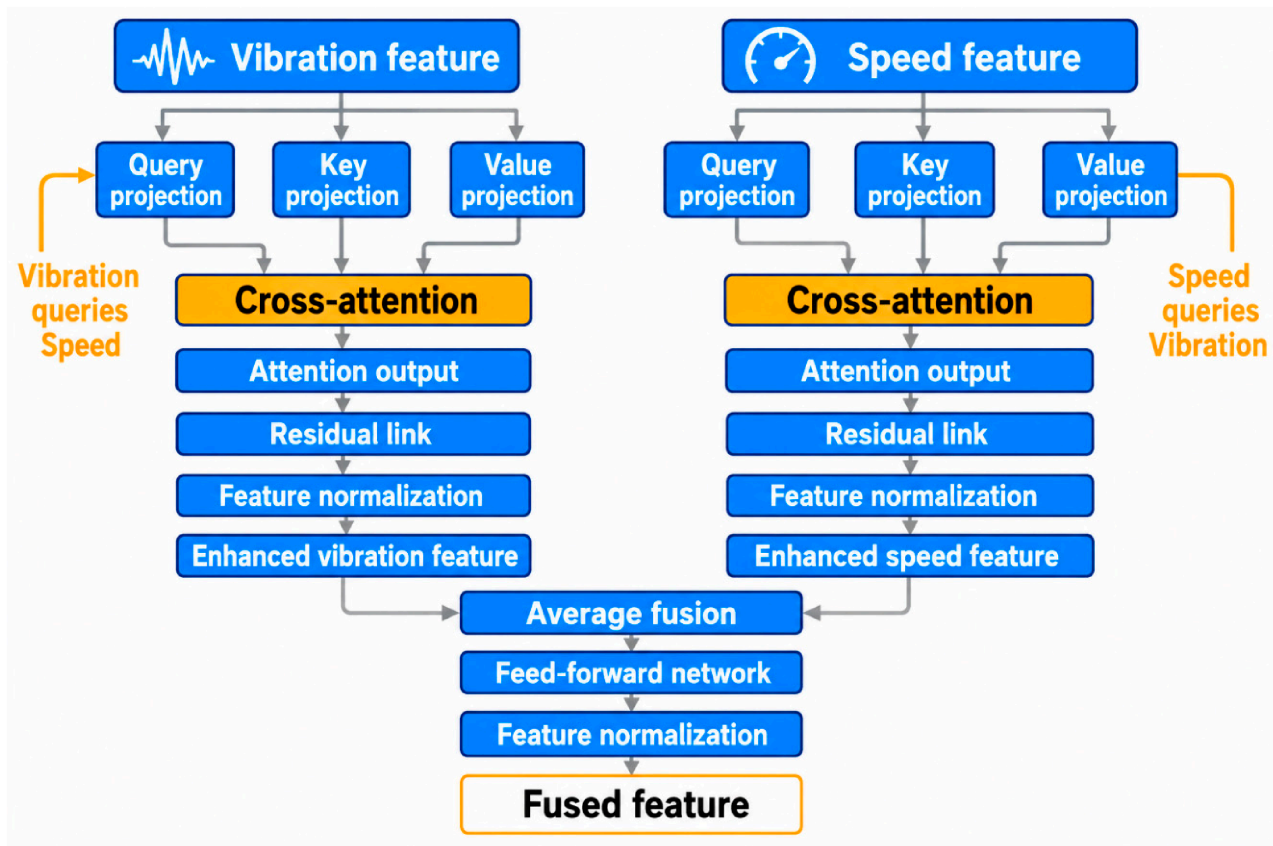


Figure 4. Cross Attention in DCMN for dual-branch feature fusion.

(1) Vibration querying speed: let $v \in \mathbb{R}^{d_{model}}$ denote the speed feature. In the first direction, vibration features serve as the query, while speed features provide the key and value. For each attention head $i \in \{1, \dots, h\}$, linear projections are performed as:

$$Q_v^{(i)} = vW_{q,v}^{(i)}, K_s^{(i)} = sW_{k,s}^{(i)}, V_s^{(i)} = sW_{v,s}^{(i)} \quad (2)$$

where $W_{q,v}^{(i)}, W_{k,s}^{(i)}, W_{v,s}^{(i)} \in \mathbb{R}^{d_{model} \times d_k}$, the scaled dot-product attention for head i is computed as:

$$\text{head}_{v \rightarrow s}^{(i)} = \text{softmax} \left(\frac{Q_v^{(i)} K_s^{(i)\top}}{\sqrt{d_k}} \right) V_s^{(i)} \quad (3)$$

The multi-head output is then obtained by concatenation and linear projection:

$$\text{CrossAttn}_{v \rightarrow s}(\mathbf{v}, \mathbf{s}) = \text{Concat}(\text{head}_{v \rightarrow s}^{(1)}, \dots, \text{head}_{v \rightarrow s}^{(h)}) \mathbf{W}_o \quad (4)$$

Finally, residual connection and layer normalization are applied:

$$\mathbf{v}_{enh} = \text{LayerNorm}(\mathbf{v} + \text{CrossAttn}_{v \rightarrow s}(\mathbf{v}, \mathbf{s})) \quad (5)$$

(2) Speed querying vibration: Symmetrically, speed features act as the query while vibration features provide the key and value. The head-wise projections are defined as:

$$Q_s^{(i)} = sW_{q,s}^{(i)}, K_v^{(i)} = vW_{k,v}^{(i)}, V_v^{(i)} = vW_{v,v}^{(i)} \quad (6)$$

The attention computation becomes

$$\text{head}_{s \rightarrow v}^{(i)} = \text{softmax} \left(\frac{Q_s^{(i)} K_v^{(i)\top}}{\sqrt{d_k}} \right) V_v^{(i)} \quad (7)$$

After multi-head concatenation and output projection, the enhanced speed representation is obtained as

$$\mathbf{s}_{enh} = \text{LayerNorm}(\mathbf{s} + \text{CrossAttn}_{s \rightarrow v}(\mathbf{s}, \mathbf{v})) \quad (8)$$

(3) Cross feature fusion: Upon obtaining the bidirectionally enhanced representations $\mathbf{v}_{enh}$ and $\mathbf{s}_{enh}$, symmetric fusion is performed via averaging:

$$fea_0 = \frac{\mathbf{v}_{enh} + \mathbf{s}_{enh}}{2} \quad (9)$$

The fused representation is further refined through a feed-forward network with residual normalization:

$$fea = \text{LayerNorm}(fea_0 + \text{FFN}(fea_0)) \quad (10)$$

Through this explicitly separated bidirectional formulation, vibration and speed modalities mutually enhance each other before integration, ensuring thorough cross-modal feature alignment and stable gradient propagation for subsequent matching rate prediction.

2.4. Matching Rates Prediction Using Fused Features

The matching rate prediction head maps the fused features into a continuous correspondence measure between vibration and speed. As shown in Figure 1, it is an MLP that progressively reduces the feature dimensionality. Each hidden layer applies a linear transformation, ReLU activation, and dropout. The final layer projects the hidden representation into a scalar, and a sigmoid activation constrains the output to $[0, 1]$ as the interpretable matching rate. Training minimizes the mean squared error between the prediction and the physics-guided self-supervised label. Values approaching one indicate strong vibration–speed correspondence, i.e., a healthy state, while lower values indicate weaker correspondence, i.e., a faulty state. This design translates high-dimensional cross-modal features into a bounded and interpretable metric, supporting the end-to-end optimization of DCMN.

2.5. Implementation Details

The vibration encoder is a 1-D DenseNet with 11 bottleneck layers and a growth rate of 8, mapping an 8192-point signal to a 356-d feature. The speed encoder is a three-layer MLP (1-128-256-356) with LeakyReLU and dropout. Speed values measured online by an optical encoder are min-max normalized before entering the speed encoder. The bidirectional cross-attention uses four heads with a hidden dimension of 356, followed by a feed-forward network (356-1424-356). The prediction head is an MLP (356-128-64-1) with a sigmoid output. The model contains 2,266,679 trainable parameters. Detailed configurations are summarized in Table 1. The mean squared error loss is adopted because the matching rate is a continuous regression target that preserves fine-grained vibration-speed similarity; binary contrastive or ranking losses would discard this information. The encoder depth

and the growth rate are kept moderate so that the receptive field covers the 8192-point signal without inflating the parameter count, and the shared embedding dimension of 356 allows the two branches to be fused without an additional projection. The model is trained for 100 epochs with the Adam optimizer, a batch size of 32, a dropout of 0.2, and a ReduceLROnPlateau scheduler, under a fixed random seed of 42 to ensure reproducibility.

Table 1. Hyperparameter and implementation configuration of DCMN.

Component	Configuration
Vibration encoder	1-D DenseNet, 11 bottlenecks, growth rate 8; 8192 -> 356-d
Speed encoder	MLP 1-128-256-356, LeakyReLU, dropout 0.2; input min-max normalized
Cross-attention	4 heads, hidden dim 356, head dim 89
Fusion FFN	356-1424-356, GELU
Prediction head	MLP 356-128-64-1, Sigmoid
Trainable parameters	2,266,679
Loss/Optimizer	MSE/Adam (lr 1×10^{-3} , weight decay 1×10^{-5})
LR scheduler	ReduceLROnPlateau (factor 0.5, patience 10)
Batch size/Epochs	32/100
Dropout	0.2
Negative samples N	2
Speed normalization	min-max over training speed range
Hardware	NVIDIA RTX 3090, PyTorch 1.11

3. Experiment and Result Analysis

This section presents two rotating machinery cases to validate DCMN's effectiveness in end-to-end health indicator generation for health monitoring under time-varying conditions. Case 1 provides a self-built rolling bearing test rig, while Case 2 presents the testing results on the benchmark dataset with varying speeds.

3.1. Case 1: Rolling Bearing Test Rig

Figure 5 shows a rolling bearing test rig. The experimental investigation was conducted on a rolling bearing test rig designed to emulate realistic operational conditions, featuring an induction motor coupled with a controllable braking system and precision reduction gearbox to regulate shaft speed. High-fidelity vibration signals were captured using piezoelectric accelerometers mounted on the bearing housings, while rotational speed data were simultaneously recorded via an optical encoder, ensuring synchronized acquisition of both modalities. The test apparatus was configured to examine six distinct bearing conditions, including healthy (HC), inner race fault (IRF), outer race fault (ORF), ball fault (BF), and two composite faults involving combined inner-outer race defects (IOF) and outer race-ball defects (OBF), allowing systematic evaluation of fault detection performance across diverse degradation modes, as shown in Figure 6. To simulate dynamic operating environments, the motor controller-imposed time-varying speed profiles, comprising sequences of increasing, decreasing, and steady-speed segments, each maintained for predefined durations, thereby reflecting realistic transient and steady-state conditions. Data collection for each condition entailed repeated acquisition sessions, with vibration signals segmented into fixed-length samples and speed values averaged or smoothed to minimize measurement noise, resulting in a high-quality, temporally aligned dataset. Detailed information on the bearing dataset is presented in Table 2. The entire dataset was systematically organized, with files named according to bearing location, fault type, rotational speed, and load, facilitating straightforward identification and retrieval. For subsequent analysis, the dataset was partitioned into training and testing subsets, maintaining an 80:20 ratio to ensure adequate representation of all fault types while preventing data

leakage. The training phase employed the fused vibration-speed data to optimize model parameters, while the testing phase assessed the model's capability to generalize across unseen operational scenarios, thereby providing a rigorous framework to evaluate the effectiveness of DCMN in capturing the complex relationships between bearing vibrations and rotational speed under realistic and time-varying conditions.

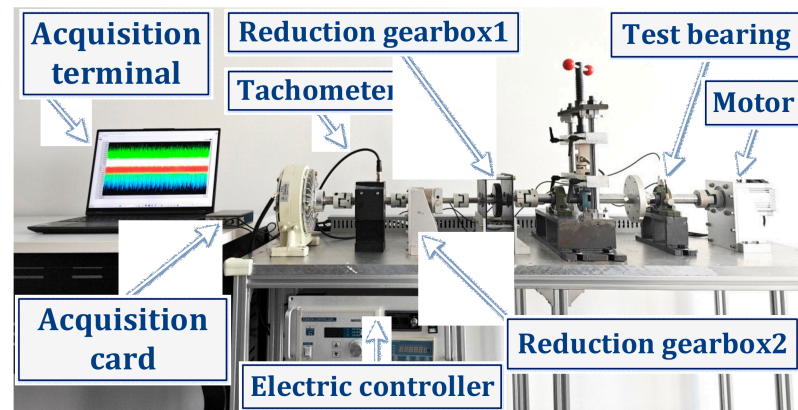


Figure 5. Rolling bearing test rig under time-varying speed conditions.

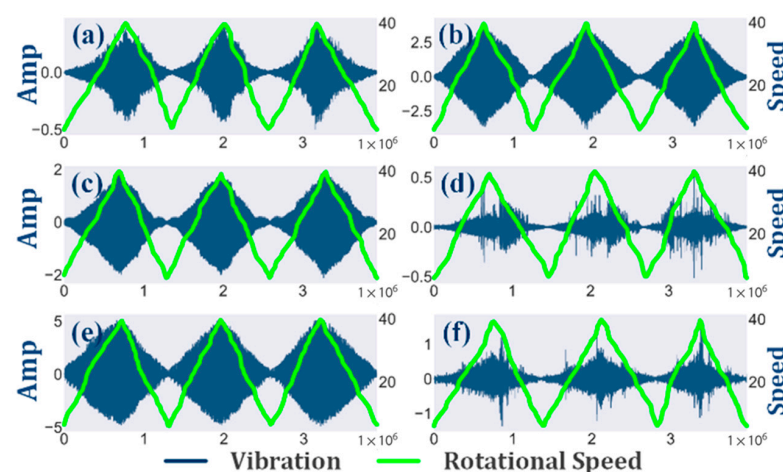


Figure 6. Vibration signals and rotational speeds. (a) HC; (b) IRF; (c) ORF; (d) BF; (e) IOF; (f) OBF.

Table 2. Detailed Information of Rolling Bearing Dataset.

Fault Type	Condition	Max Speed (r/min)	Min Speed (r/min)	Frequency (HZ)	Sample Length	Sample Interval	Train	Test
HC	(INC20→DEC20) × 3	2400 (40 HZ)	350 (5.83 HZ)	25,600	8192	8192	300	75
IRF	(INC20→DEC20) × 3	2400 (40 HZ)	350 (5.83 HZ)	25,600	8192	8192	/	75
ORF	(INC20→DEC20) × 3	2400 (40 HZ)	350 (5.83 HZ)	25,600	8192	8192	/	75
BF	(INC20→DEC20) × 3	2400 (40 HZ)	350 (5.83 HZ)	25,600	8192	8192	/	75
IOF	(INC20→DEC20) × 3	2400 (40 HZ)	350 (5.83 HZ)	25,600	8192	8192	/	75
OBF	(INC20→DEC20) × 3	2400 (40 HZ)	350 (5.83 HZ)	25,600	8192	8192	/	75

3.1.1. Training Visualization Results

Figures 7 and 8 show the dimensionality-reduction visualization results of vibration representations and speed embedding, respectively. The t-distributed stochastic neighbor embedding (t-SNE) visualization depicts the 2D projection of vibration feature embeddings, color-coded by rotational speed, illustrating how the high-dimensional representations organize according to speed variations. Embeddings corresponding to similar speeds form

coherent clusters, while distinct speed ranges are clearly separated, showing smooth transitions across the spectrum and minimal overlap between dissimilar groups. This continuous organization highlights the encoder's ability to capture speed-dependent characteristics, ensuring that features from comparable operating conditions are closely aligned. The self-supervised negative sampling strategy further reinforces this effect by providing diverse speed-conditioned signals during training, enabling the model to simulate a wide range of operational scenarios and learn a continuous mapping between vibration patterns and speed values. Moreover, the clustering demonstrates successful cross-modal alignment, as vibration embeddings implicitly encode rotational speed information, confirming the effective coupling between the two modalities. The distinct and discriminative feature patterns across the speed spectrum validate that DCMN preserves fine-grained speed variations within its learned representations. Collectively, these results demonstrate DCMN's capability in learning robust vibration-speed matching relationships, providing embeddings that are both discriminative and consistent with operational dynamics, thereby supporting accurate matching rate prediction and enhancing the potential for fault diagnosis under variable-speed conditions.

Vibration Features Visualization via t-SNE (DenseNet)

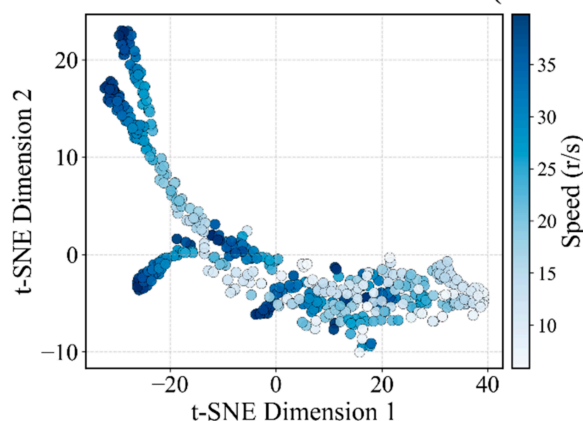


Figure 7. Visualization results of healthy vibration representations.

Speed Embedding Features Visualization via t-SNE

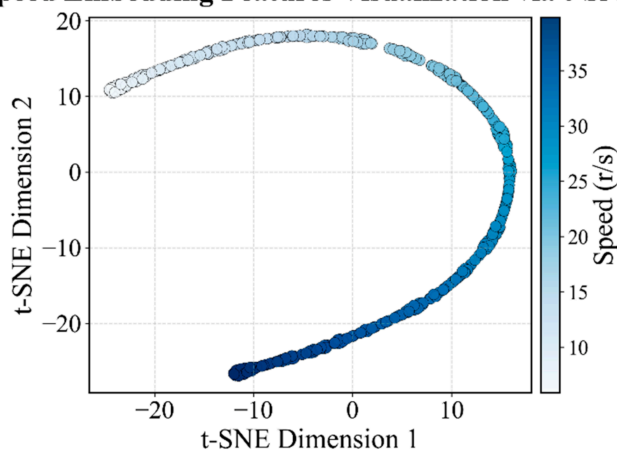


Figure 8. Visualization results of speed embeddings.

Figure 9 shows the dimensionality-reduction visualization results of fused features from different fault types. The t-SNE visualization of fused features from the cross-attention fusion module reveals clear clustering according to fault types, with color coding distinguishing healthy, IRF, ORF, BF, IOF, and OBF conditions, demonstrating remarkable gener-

alization from healthy-only training. Remarkably, DCMN learns exclusively from healthy samples using a self-supervised match-pair strategy, without any exposure to fault labels or faulty data during training, yet the fused features exhibit distinct separation between different fault categories. The visualization highlights that the model naturally discovers fault patterns in an unsupervised manner, forming well-defined clusters where healthy and various faulty conditions are clearly delineated. This self-supervised vibration-speed matching framework trains the model to predict matching rates between vibration signals and their corresponding rotational speeds, where healthy vibrations achieve high matching rates with true speeds, while faulty vibrations produce mismatched predictions. Such a contrast-based learning mechanism effectively amplifies dynamic discriminability, as the model differentiates matched from mismatched pairs, creating separable feature representations for different fault types despite zero-shot exposure to faults. Consequently, the fused feature embeddings preserve both cross-modal information and discriminative characteristics, enabling robust inter-class separability. Importantly, the resulting matching rate predictions directly support health indicator generation, providing an unsupervised metric for fault detection and operational monitoring. Collectively, these results demonstrate that DCMN can transform exclusively healthy training data into a powerful foundation for zero-shot fault classification, revealing both the effectiveness of the self-supervised match-pair strategy and the practical value of fused vibration-speed representations for unsupervised health monitoring.

Fused Features Visualization via t-SNE (by Fault Type)

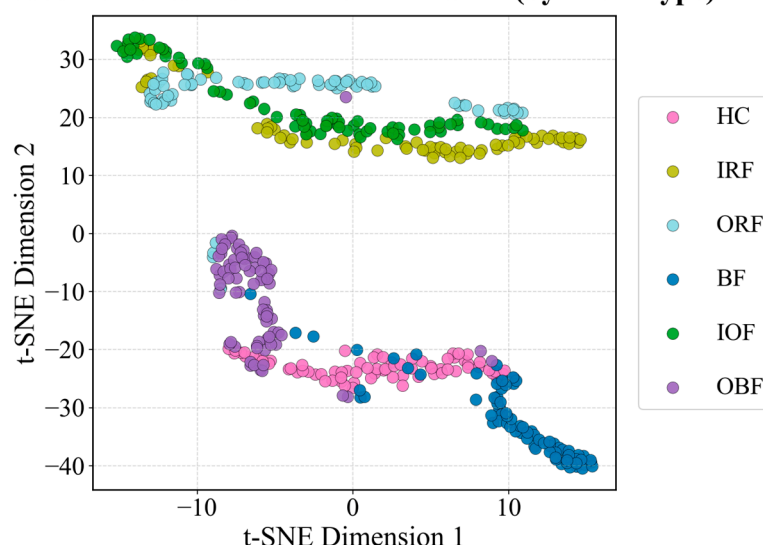


Figure 9. Visualization results of different fault types.

3.1.2. Health Indicator Analysis and Machinery Monitoring

Figure 10 presents the statistical summary of matching rates across different health conditions, while Figure 11 shows the corresponding full distributions using violin plots, offering a complete view of matching rate behavior for healthy and faulty vibrations. Healthy samples exhibit tightly clustered distributions near high matching rates, reflecting consistent recognition of normal vibration-speed correspondence by DCMN. In contrast, all fault types—including IRF, ORF, BF, IOF, and OBF—show broader distributions shifted toward lower matching rates, with increased variance and occasional multi-modal shapes, reflecting diverse fault manifestations and severity levels. Figure 10 further quantifies these trends, showing minimal overlap between healthy and faulty groups and large effect sizes, confirming statistically significant separation. This clear distinction demonstrates

the feasibility of matching rate-based health monitoring, as simple thresholds can reliably discriminate healthy from faulty conditions.

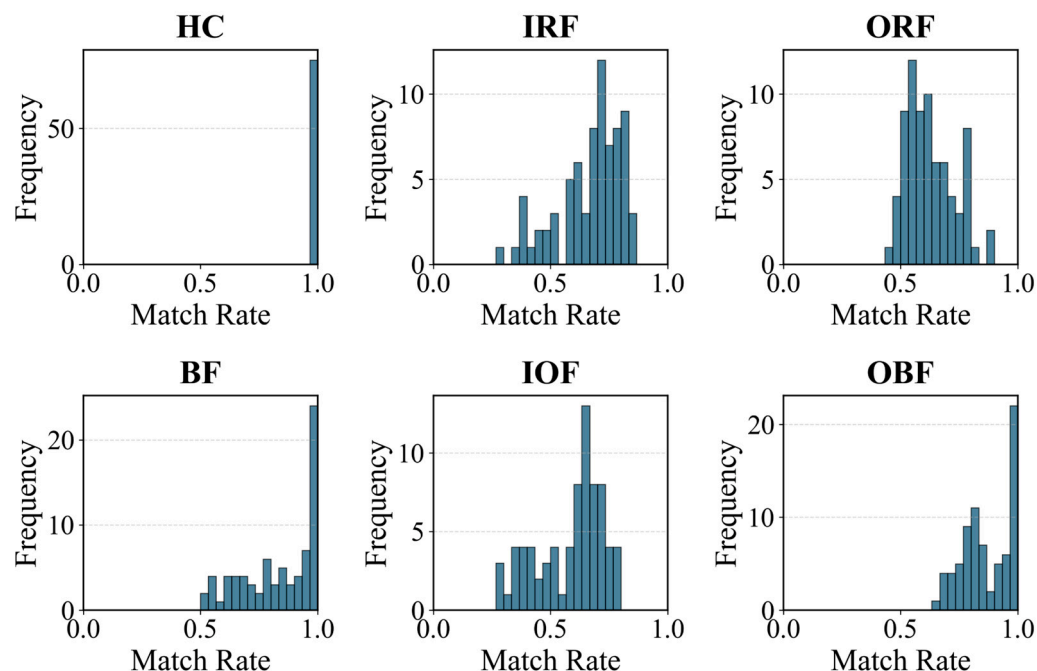


Figure 10. Statistical summary of matching rates across different health conditions.

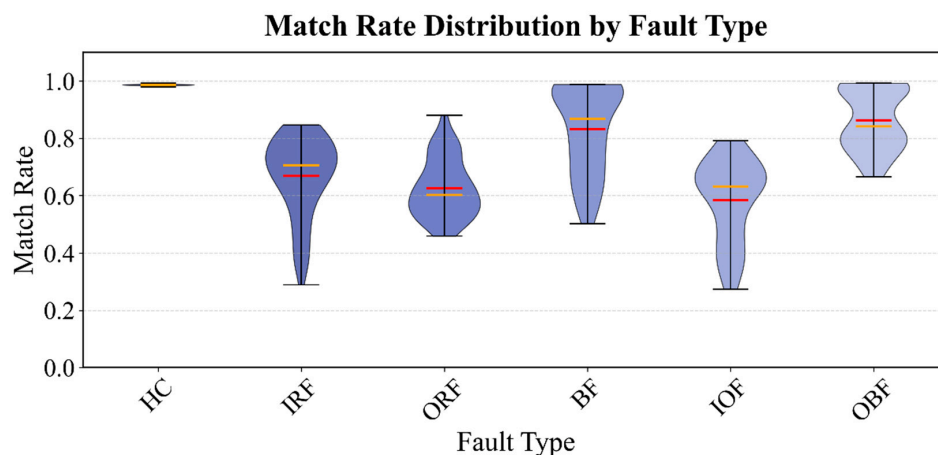


Figure 11. Violin plots of matching-rate distributions for healthy and faulty vibrations.

These patterns arise from DCMN's healthy-only training, where the model learns intrinsic vibration-speed correspondence via self-supervised learning on normal samples. Healthy vibrations form a narrow recognition region, while faulty vibrations deviate from this pattern, producing low matching rates. The cross-attention fusion of vibration and speed embeddings further amplifies this effect, ensuring that disruptions caused by faults manifest as significant drops in predicted matching rates. As a result, the matching rate serves as a continuous health indicator: high values indicate normal operation, while lower values correspond to increasing fault severity. Minimal distribution overlap allows reliable threshold-based detection, and the continuous scale enables assessment of different fault types. Importantly, this method operates without labeled fault data, highlighting its unsupervised capability and generalization to unseen faults. Collectively, the summary statistics and violin plots confirm that matching rate distributions provide a robust, interpretable,

and statistically sound metric for real-world health monitoring, capturing both consistent healthy behavior and discriminative deviations caused by faults.

Figure 12 illustrates the predicted matching rates of DCMN for all six fault modes on the Case 1 test set, where each bar corresponds to one sample. The health indicator is defined as the model output matching rate, and a single threshold is determined by requiring that every healthy test sample stays on the normal side of the threshold (i.e., zero false alarms on healthy samples); the fault states are then evaluated at this threshold. Healthy samples all remain above the threshold, whereas the faulty samples fall below it, confirming that DCMN provides a bounded, interpretable health indicator for online monitoring. Because the rotational speed is explicitly fused into the prediction, the matching rate does not drift with the operating condition, so a single threshold remains valid across the whole speed range. The predicted matching rates under each fault type are detailed in Table 3. The bounded $[0, 1]$ output further makes the indicator directly comparable across different machines and operating points.

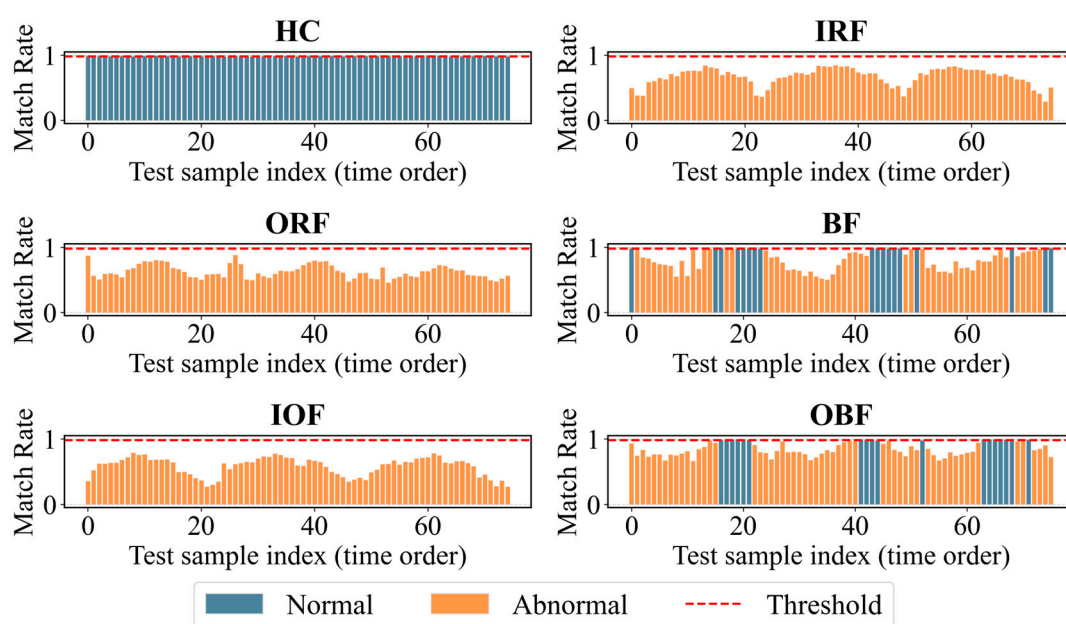


Figure 12. Match-rate bars of DCMN on Case 1.

Table 3. Health Monitoring Performance of DCMN.

Fault Type	Total	Right Alarm	Accuracy (%)
Healthy	75.00	75.00	100.00
IRF	75.00	75.00	100.00
ORF	75.00	75.00	100.00
BF	76.00	58.00	76.32
IOF	75.00	75.00	100.00
OBF	76.00	58.00	76.32

3.1.3. Comparison with Other Representative Approaches

In this section, one conventional deep neural network (DNN), i.e., Convolutional autoencoder (CAE) [28], four state-of-the-art DNNs: UNet [29], one-dimensional residual convolutional auto-encoder (1DRCAE) [30], densely-connected UNet (DenseUNet) [31], long and short-term memory-based autoencoder (LSTMAE) [32], and the quantitative estimation neural network (QENN) [25] are considered to perform comparison to illustrate the superiority of DCMN. The health indicator generated by DCMN can be found in

Figure 12. Even under large speed fluctuations, the health indicators produced by DCMN remain consistently stable in the healthy state. In contrast, the health indicators of the compared methods shift markedly with the speed, which makes the threshold difficult to set and causes frequent false alarms and missed detections. In addition, a ball fault produces abnormal impulses only when the damaged region of the rolling element contacts the inner or outer race during rotation; otherwise, its vibration remains close to that of the healthy state, so the compared methods can hardly detect ball faults. In contrast, DCMN achieves a detection rate of 76.32% for ball faults. DCMN is capable of accurately distinguishing the healthy state and abnormal patterns such as IRF, ORF, and IOF. It also demonstrates satisfactory performance for BF and OBF. An average abnormal detection rate of 92.11% is achieved, which is superior to that of the compared methods. These results demonstrate the effectiveness and superiority of DCMN.

Since the original QENN is trained with severity and speed labels, an unsupervised adaptation is implemented for a fair comparison under the healthy-only protocol. The feature extractor and the adversarial speed discriminator behind a gradient reversal layer are retained, and the severity classifier is replaced by a reconstruction decoder, which is the only label-free substitute that keeps the features informative. The health indicator of QENN is computed as the relative similarity between the feature of each sample and the healthy training features, and the threshold is set as the maximum training indicator, identical to the other methods. All compared methods share the same training data, epochs, and optimizer settings as DCMN.

To further validate the superiority of DCMN, a recent Transformer-based autoencoder (TransformerAE) is additionally compared under the identical healthy-only training protocol. As shown in Table 4 and Figure 13, DCMN outperforms all compared methods on the average detection rate (92.11%) and, notably, is the only method capable of detecting ball faults (76.32% vs. 0% for all baselines). The reconstruction-based models fail to capture the intermittent impulses of ball faults, whereas the cross-modal matching mechanism of DCMN preserves this discriminability. The adapted QENN achieves an average detection rate of 55.04%, on par with the unsupervised reconstruction baselines, and its IRF and IOF detection rates (76.00% and 72.00%) exceed those of most of them, confirming the benefit of the adversarial speed-invariant design. This level is expected rather than accidental: the original QENN is a fault-supervised method whose severity classifier provides direct discriminative supervision, whereas the healthy-only protocol of this comparison deliberately removes that supervision. If fault-labeled data were available for training, the performance of QENN would be expected to improve significantly. Under the identical unsupervised protocol, DCMN still outperforms it by 37.07 percentage points on average. If fault-labeled data were available for training, the performance of QENN would be expected to improve significantly. Under the identical unsupervised protocol, DCMN still outperforms it by 37.07 percentage points on average.

Table 4. Comparison results with Representative methods.

Models	Healthy	IRF	ORF	BF	IOF	OBF	AVG
CAE	98.67	89.33	61.33	0.00	100.00	38.16	64.58
UNet	98.67	82.67	58.67	0.00	80.00	42.11	60.35
1DRCAE	98.67	82.67	58.67	0.00	80.00	43.42	60.57
DenseUNet	97.33	82.67	57.33	0.00	73.33	47.37	59.67
LSTMAE	100.00	78.67	57.33	0.00	73.33	34.21	57.26
TransformerAE	98.67	82.67	60.00	0.00	81.33	48.68	61.89
QENN	98.67	76.00	52.00	0.00	72.00	31.58	55.04
DCMN	100.00	100.00	100.00	76.32	100.00	76.32	92.11

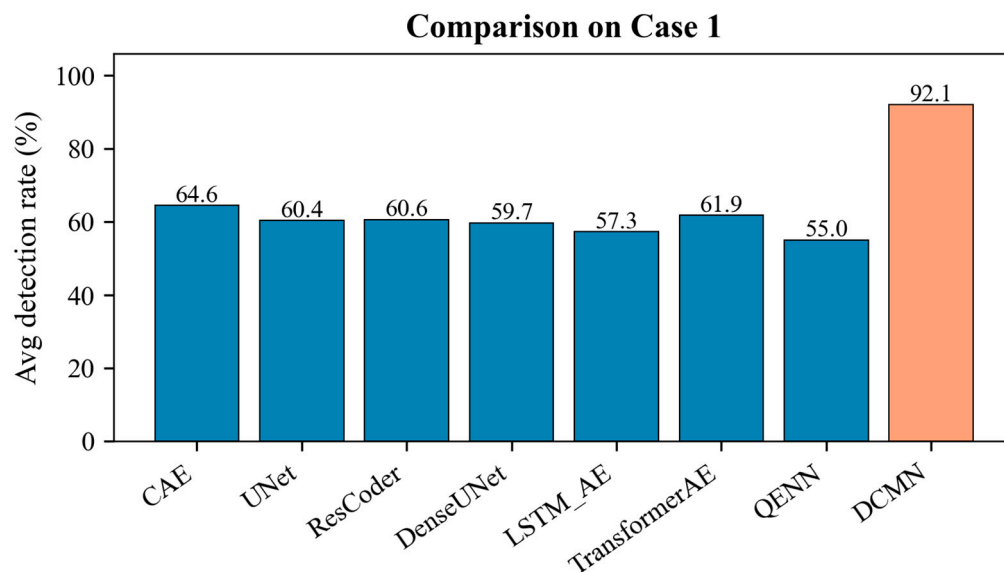


Figure 13. Detection-rate comparison on Case 1 (DCMN vs. compared methods).

3.1.4. Ablation Study

To quantify the contribution of each component, an ablation study is conducted on Case 1, as summarized in Table 5 and Figure 14. Five variants are compared: V1 (concatenation fusion), V2 (single-direction attention), V3 (proposed bidirectional), V4 (without speed input), and V5 (binary labels). Removing the speed input (V4) causes the model to lose fault-detection capability (its average detection rate drops to 22.88%), confirming that speed conditioning is essential; adopting binary labels (V5) collapses the health indicator into a non-discriminative output that alarms on almost every sample. The proposed V3 achieves the lowest test loss and the best balanced performance, i.e., zero false alarms on healthy samples together with a high average detection rate (92.11%). V2 (single-direction attention) yields a marginally higher average rate but at the cost of a non-zero false alarm rate. These results confirm that each component contributes to the overall performance.

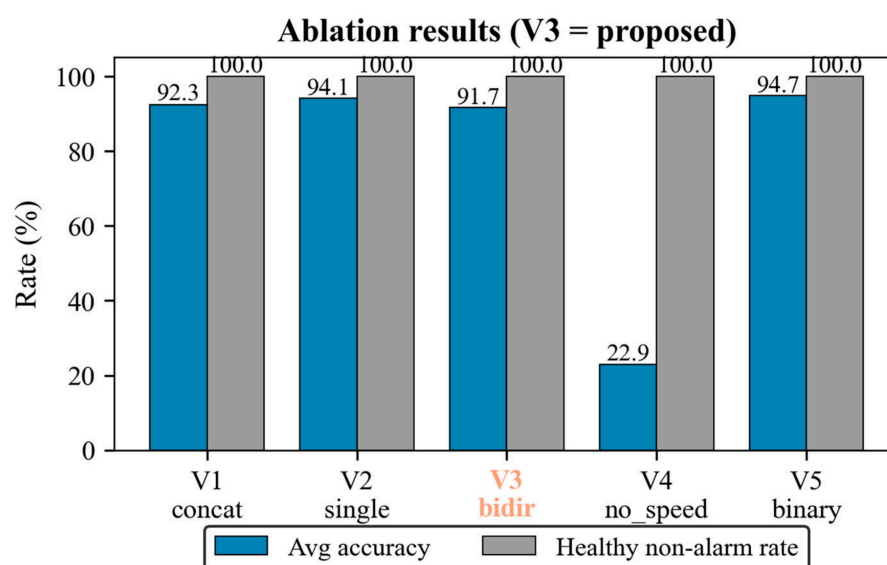


Figure 14. Ablation results: average accuracy and ball-fault detection rate per variant.

Table 5. Ablation results on Case 1.

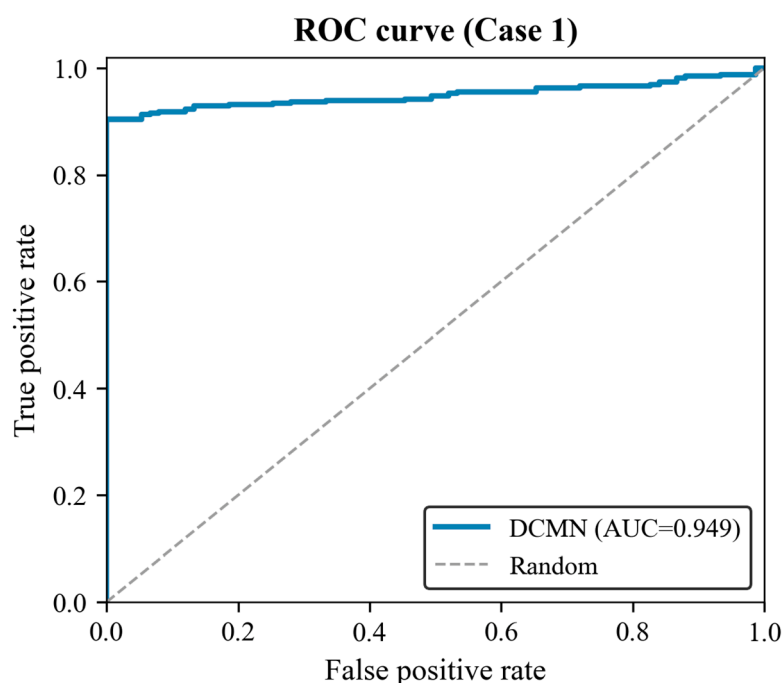
Variant	Params	Healthy	IRF	ORF	BF	IOF	OBF	AVG
V1	1.50 M	100.00	100.00	100.00	68.42	100.00	75.00	90.57
V2	1.76 M	93.33	100.00	98.67	76.32	100.00	93.42	93.62
V3	2.27 M	100.00	100.00	100.00	76.32	100.00	76.32	92.11
V4	1.12 M	100.00	0.00	32.00	5.26	0.00	0.00	22.88
V5	2.27 M	0.00	100.00	100.00	100.00	100.00	100.00	83.33

3.1.5. Evaluation Metrics

Beyond the alarm-rate evaluation, a binary anomaly-detection assessment (healthy vs. fault), including the receiver operating characteristic (ROC) curve and the area under the curve (AUC), is performed, as shown in Table 6 and Figure 15. The AUC reaches 0.949, indicating strong separability. At the operating threshold, DCMN achieves a precision of 1.000, a recall of 0.9045, an F1-score of 0.9499, a false alarm rate of 0.00%, and a missed detection rate of 9.55%. The combination of a high AUC and a zero false alarm rate is particularly meaningful for safety-critical monitoring: a precision of 1.000 means that every alarm raised by DCMN corresponds to a true fault, while the residual missed detections are concentrated on the intermittent ball fault and outer-race-ball fault, whose vibrations partly resemble the healthy state.

Table 6. Binary anomaly-detection metrics of DCMN on Case 1.

Metric	Value
AUC	0.949
Precision	1.000
Recall	0.9045
F1-score	0.9499
False alarm rate	0.00%
Missed detection rate	9.55%

**Figure 15.** ROC curve of DCMN on Case 1.

3.1.6. Computational Complexity

The computational complexity is reported in Table 7 and Figure 16. DCMN requires 2.27 M parameters and 31.6 M multiply-accumulate operations (MACs), with an inference time of 4.8 ms and a peak graphics-processing-unit (GPU) memory of 64 MB. The UNet baseline consumes 22.4 G MACs, over 700 times more expensive than DCMN, and the Transformer-based autoencoder requires 139.5 M MACs, about 4.4 times more than DCMN, while both deliver lower detection performance. DCMN thus strikes a favorable balance between accuracy and efficiency.

Table 7. Computational complexity of DCMN and compared methods.

Model	Params	MACs	Infer. bs1 (ms)	GPU (MB)
QENN	1.23 M	159.3 M	1.32	10
DCMN	2.27 M	31.6 M	4.78	64
CAE	0.43 M	22.2 M	2.18	65
UNet	10.82 M	22.4 G	8.13	236
1DRCAE	1.62 M	55.7 M	2.82	70
DenseUNet	0.02 M	419.0 M	9.99	71
LSTMAE	0.44 M	10.1 M	2.41	75
TransformerAE	0.84 M	139.5 M	1.81	16

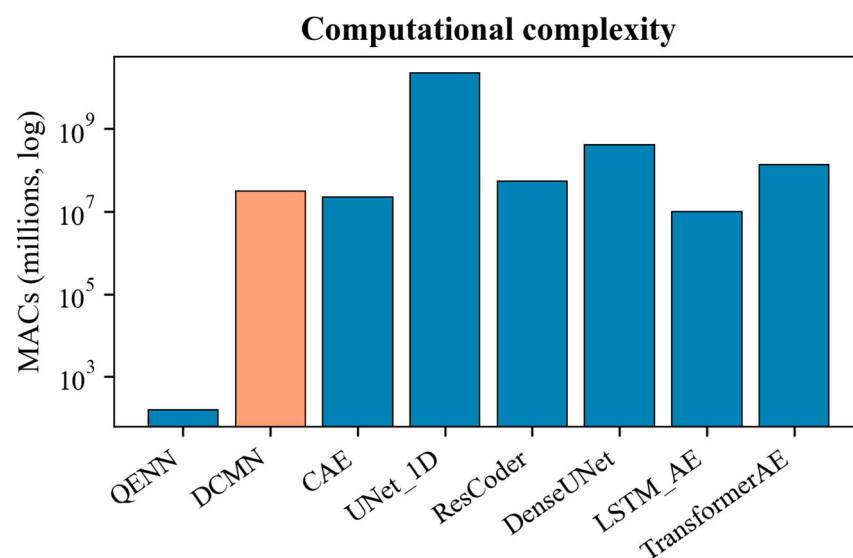


Figure 16. Model computation comparison (log scale).

3.1.7. Sensitivity Analysis

The sensitivity to the number of negative samples N is examined in Figure 17. The average accuracy remains stable across $N = 1.5$ (89.25–93.42%), which confirms that DCMN is robust to this hyperparameter. The default $N = 2$ is retained as a favorable trade-off between detection performance and training cost. An additional study on the matching-rate formulation (linear, Gaussian, cosine) is presented in Figure 18. The linear formulation achieves the highest average rate (92.11%), which is the proposed configuration, followed by cosine (90.34%) and Gaussian (90.13%); the linear formulation is therefore retained.

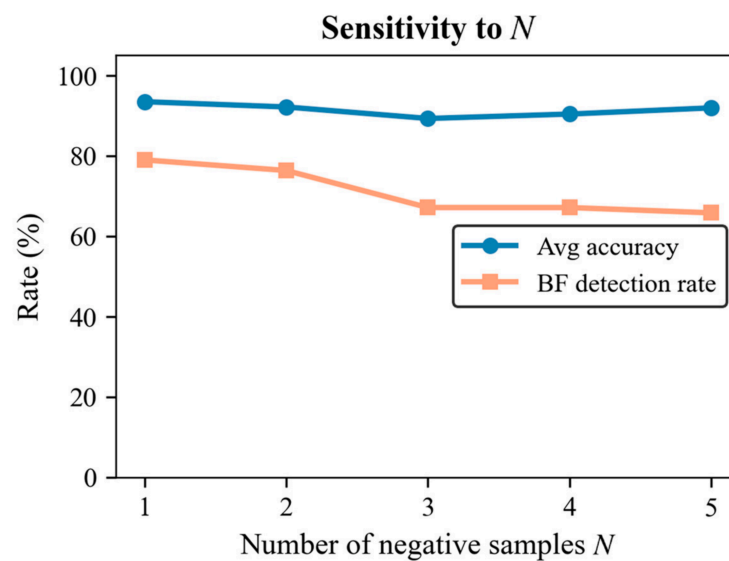


Figure 17. Sensitivity to the number of negative samples N .

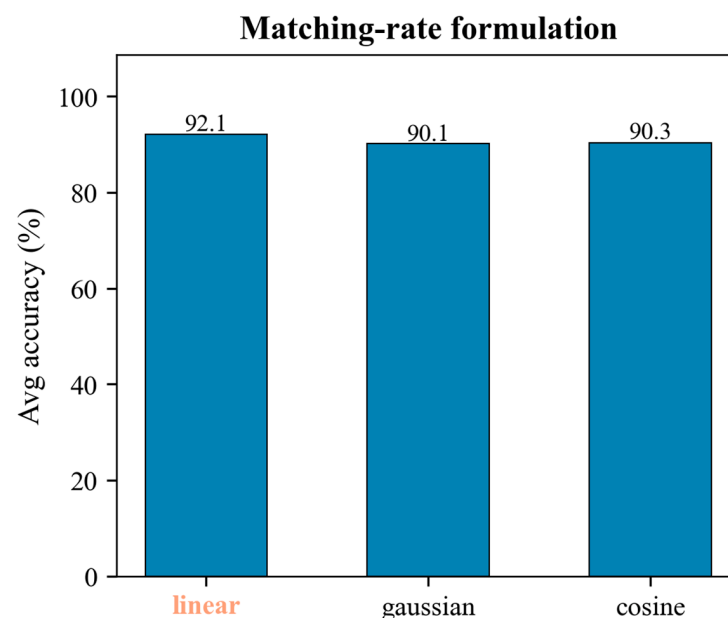


Figure 18. Sensitivity to the matching-rate formulation.

3.2. Case 2: Bearing WTH Dataset

To validate the generalization of DCMN, a second case study is conducted on the Bearing WTH dataset [33] under varying-speed conditions. The dataset contains four fault types (Healthy, IRF, ORF, BF) with 586 test samples per class and 2342 healthy training samples, as summarized in Table 8. DCMN is trained with the same protocol as in Case 1. The per-class detection rates and a comparison with the representative methods are reported in Table 9. Corresponding bar chart visualizations are presented in Figure 19. DCMN achieves an average detection rate of 86.35%, with ball-fault detection reaching 98.81%, confirming that DCMN generalizes across different bearing test rigs. On this dataset the outer-race fault is the most difficult class, which is consistent with the intermittent-impulse mechanism discussed above; nevertheless, DCMN is again the only method that detects it at a usable rate, and its ball-fault detection rate reaches 98.81%. This result indicates that the matching-rate mechanism generalizes across datasets rather than overfitting to a particular test rig. The adapted QENN obtains an average detection rate of 38.44% on

this dataset, again on par with the unsupervised baselines and below DCMN (86.35%). This is consistent with the Case 1 observation: the removal of severity supervision in the adaptation weakens the discriminative power of QENN, and its advantage is expected to return when fault-labeled training data are available. This is consistent with the Case 1 observation: the removal of severity supervision in the adaptation weakens the discriminative power of QENN, and its advantage is expected to return when fault-labeled training data are available.

Table 8. Detailed information of the Bearing WTH dataset.

Fault Type	Test Samples	Speed Range (Hz)
Healthy	586	10.0–30.6
Inner race fault (IRF)	586	10.0–30.6
Outer race fault (ORF)	586	10.0–30.6
Ball fault (BF)	586	10.0–30.6

Table 9. Detection results on Case 2.

Model	Healthy	IRF	ORF	BF	AVG
QENN	100.00	51.19	2.56	0.00	38.44
DCMN	100.00	98.81	46.59	100.00	86.35
CAE	99.83	98.63	0.00	15.70	53.54
UNet	100.00	92.49	0.00	4.27	49.19
1DRCAE	100.00	99.15	0.85	28.50	57.12
DenseUNet	100.00	99.32	1.37	24.91	56.40
LSTMAE	100.00	96.42	0.00	8.87	51.32
TransformerAE	100.00	99.32	1.88	53.41	63.65

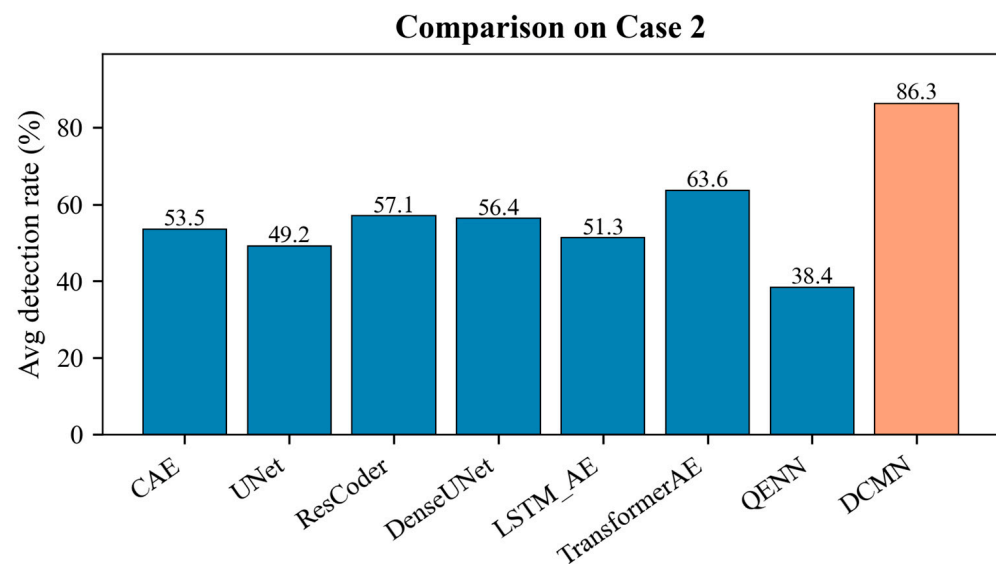


Figure 19. Detection-rate comparison on Case 2.

4. Discussion

The results in Section 3 show that the proposed matching-rate health indicator outperforms reconstruction-based indicators under time-varying speeds, and the mechanism is twofold. First, a reconstruction error conflates fault-related deviations with speed-induced amplitude variations, so a healthy sample acquired at an unseen speed may be reconstructed poorly and trigger a false alarm. In contrast, the matching rate directly measures the correspondence between the vibration pattern and the instantaneous operating con-

dition, so it remains high for healthy samples regardless of the speed value. Second, the matching rate is bounded in $[0, 1]$ and its threshold is the maximum training indicator, which provides a unified and interpretable decision rule across all classes.

The comparison with QENN further clarifies the contribution of DCMN. QENN suppresses condition-related variations through adversarial speed discrimination, which is effective for speed invariance. However, in the unsupervised setting where only healthy data are available, removing speed information alone does not guarantee fault sensitivity, because nothing forces faulty samples to deviate from the healthy baseline in feature space. DCMN resolves this dilemma with self-supervised matching pairs: the negative pairs spanning the whole speed range enforce speed invariance, while the positive pairs at the true speed preserve the vibration-condition correspondence that faults destroy. This explains why DCMN detects intermittent ball faults (76.32% on Case 1) whereas the adapted QENN and all reconstruction-based baselines fail (0%). It should also be noted that this comparison is conservative for QENN: the original QENN trains with fault severity labels, which are unavailable under the healthy-only protocol, and QENN is expected to perform substantially better when labeled fault data exist. From an engineering perspective, DCMN requires 2.27 M parameters and 31.6 M MACs with a 4.78 ms inference latency at batch size 1, which is comparable to the lightest baselines and is deployable on edge controllers. The decision threshold is derived from training data only, so no manual calibration is needed after deployment. The main practical requirement is a speed measurement channel, which is already available in most modern rotating machinery through key-phaser or encoder signals.

5. Conclusions

In this article, a novel Dynamic Condition-Matching Network (DCMN) was proposed for end-to-end health indicator generation under time-varying operating conditions. The model addresses the dual challenges of multi-modal feature alignment and the lack of bounded, condition-invariant metrics. Three technical innovations were developed: First, a self-supervised matching-pair strategy enables the network to learn speed-invariant representations using only healthy data, eliminating the need for manual fault labeling. Second, a bidirectional cross-attention mechanism was introduced to model the complex dependencies between vibration and speed signals, capturing subtle deviations from the reference baseline that conventional concatenation methods fail to detect. Finally, an end-to-end prediction head provides a matching rate bounded in the $[0, 1]$ range, offering a unified metric for elegant threshold management across diverse industrial scenarios. Experimental validation on bearing datasets demonstrated the superiority of DCMN over traditional autoencoder-based baselines. Notably, DCMN achieves a detection rate of 76.32% for ball faults, where the compared methods fail to detect any ball faults (0%), and raised the overall average detection rate to 92.11%. Beyond fault diagnosis, DCMN reveals the potential of cross-modal alignment, providing a robust and promising paradigm for intelligent machinery monitoring in dynamic industrial environments. Despite the encouraging results, this study has several limitations that deserve further investigation. First, the ball fault and the outer-race-ball fault remain the most difficult cases, because their abnormal impacts are intermittent and appear only at certain rotation phases, so part of their vibrations still resembles the healthy state. Second, the speed encoder currently takes a single scalar per sampling window, which assumes a quasi-steady operating condition and does not explicitly model the acceleration-versus-deceleration dynamics. Third, the validation is limited to bearing datasets. Accordingly, future work will focus on strengthening the sensitivity to intermittent impulses through impulse-aware augmentation, extending the speed input to a short sequence that captures the within-window dynamics together with a quantitative

window-length sensitivity analysis, and validating the framework on additional machine types such as gearboxes.

Author Contributions: Conceptualization, J.W. (Jiangling Wang), C.Z. and Y.T.; Methodology, J.W. (Jiangling Wang), J.W. (Juntao Wang) and Y.T.; Software, J.W. (Juntao Wang); Validation, J.W. (Juntao Wang), C.Z. and Y.T.; Formal analysis, J.W. (Jiangling Wang), C.Z. and Y.T.; Writing—original draft, J.W. (Juntao Wang) and Y.T.; Writing—review & editing, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 62172242), the Ningbo Science and Technology Fund (Grant Nos. 2023Z213, 2023Z228, 2024Z202, 2024Z296, 2025Z036), the Ningbo Key Laboratory of Intelligent Home Appliances, and the Ningbo “Innovation Yongjiang 2035” Key R&D Programme (Grant No. 2026Z005).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Gao, Z.; Cecati, C.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques-part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3768–3774. [\[CrossRef\]](#)
2. Zhang, G.; Wang, T.; Chu, F. Mechanism-informed degradation assessment for machinery incorporating monitoring and diagnosis under nonstationary operating conditions. *IEEE/ASME Trans. Mechatron.* **2026**, *31*, 861–872. [\[CrossRef\]](#)
3. Zhou, J.; Yang, J.; Xiang, S.; Qin, Y. Remaining useful life prediction methodologies with health indicator dependence for rotating machinery: A comprehensive review. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3528519. [\[CrossRef\]](#)
4. Liu, W.; Chang, J.; Wang, L.; Yu, J. Speed-normalized expansion-and-shrinkage network for health indicator generating under time-varying conditions. *Adv. Eng. Inform.* **2025**, *67*, 103571. [\[CrossRef\]](#)
5. Duan, Y.; Cao, X.; Zhao, J.; Xu, X. Health indicator construction and status assessment of rotating machinery by spatio-temporal fusion of multi-domain mixed features. *Measurement* **2022**, *205*, 112170. [\[CrossRef\]](#)
6. Li, S.; Zhang, C.; Zhang, X. A novel spatiotemporal enhanced convolutional autoencoder network for unsupervised health indicator construction. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2514510. [\[CrossRef\]](#)
7. Koutroulis, G.; Mutlu, B.; Kern, R. Constructing robust health indicators from complex engineered systems via anticausal learning. *Eng. Appl. Artif. Intell.* **2022**, *113*, 104926. [\[CrossRef\]](#)
8. Xu, F.; Wang, L. Constructing a health indicator for bearing degradation assessment via an unsupervised and enhanced stacked autoencoder. *Adv. Eng. Inform.* **2022**, *53*, 101708. [\[CrossRef\]](#)
9. Liu, W.; Xiao, C.; Yu, J. Incorporate rotational speeds into deep neural network for machinery health monitoring. *IEEE Trans. Ind. Inform.* **2024**, *20*, 13330–13340. [\[CrossRef\]](#)
10. Zhang, Y.; Yuan, X.; Yao, X.; Liu, J.; Zhou, Y.; Zhai, N.; Niu, X. Preserving cross-domain spectral consistency: A latent dependency relationship-aware network for transfer fault diagnosis under time-varying speed conditions. *Expert Syst. Appl.* **2026**, *314*, 131554. [\[CrossRef\]](#)
11. Liang, P.; Xu, L.; Shuai, H.; Yuan, X.; Wang, B.; Zhang, L. Semisupervised subdomain adaptation graph convolutional network for fault transfer diagnosis of rotating machinery under time-varying speeds. *IEEE/ASME Trans. Mechatron.* **2023**, *29*, 730–741. [\[CrossRef\]](#)
12. Chen, K.; Meng, Z.; Sun, D.; Cao, W.; Ren, Y.; Sun, W.; Wu, J. A novel sliding mixing graph contrastive domain adaptation method for fault diagnosis under time-varying speeds. *Expert Syst. Appl.* **2025**, *270*, 126576. [\[CrossRef\]](#)
13. Yang, J.; Liu, S.; Wang, Y.; Zhang, X.; Yang, X. Remaining useful life prediction study of rolling bearings based on TCN-transformer and KCA. *IEEE Sens. J.* **2025**, *25*, 28716–28726. [\[CrossRef\]](#)
14. Li, M.; Huang, J.; Zhang, F.; Yu, Y.; Gu, F.; Chu, F. MSIF-Convformer: A novel end-to-end fault diagnosis framework with multi-source sensors under strong noise. *Inf. Fusion* **2026**, *129*, 104000. [\[CrossRef\]](#)
15. Yu, Y.; Karimi, H.R.; Gelman, L.; Cetin, A.E. MSIFT: A novel end-to-end mechanical fault diagnosis framework under limited & imbalanced data using multi-source information fusion. *Expert Syst. Appl.* **2025**, *274*, 126947. [\[CrossRef\]](#)
16. Zhu, L.; Wang, J.; Chen, M.; Liu, L. Fusion-driven fault diagnosis based on adaptive tuning feature mode decomposition and synergy graph enhanced transformer for bearings under noisy conditions. *Expert Syst. Appl.* **2025**, *260*, 125441. [\[CrossRef\]](#)
17. Liu, D.; Cheng, X.; Wang, G.; Cui, L. Explainable cross-attention-driven data fusion transfer learning for machinery fault diagnosis. *Measurement* **2026**, *267*, 120588. [\[CrossRef\]](#)

18. Yu, M.; Li, Y.; Meng, G.; Wang, Y. An intelligent fault diagnosis for rotating machine under strong noise based on cross-attention-driven spatial-temporal feature fusion and duplexing time sequence convolution optimization. *Eng. Appl. Artif. Intell.* **2025**, *153*, 110897. [\[CrossRef\]](#)
19. Han, Z.; Xia, W.; Shen, W.; Zhu, Q.; Liu, H.; Zhang, C. Simulation-to-real transfer learning for bearing fault diagnosis across working conditions: A hybrid approach combining physical modeling and data-driven techniques. *Adv. Eng. Inform.* **2026**, *69*, 103998. [\[CrossRef\]](#)
20. Doersch, C.; Gupta, A.; Efros, A.A. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2015; pp. 1422–1430. [\[CrossRef\]](#)
21. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 15979–15988. [\[CrossRef\]](#)
22. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning (ICML)*; JMLR Inc.: Norfolk, MA, USA, 2020; pp. 1597–1607. [\[CrossRef\]](#)
23. Peng, P.; Lu, J.; Xie, T.; Tao, S.; Wang, H.; Zhang, H. Open-set fault diagnosis via supervised contrastive learning with negative out-of-distribution data augmentation. *IEEE Trans. Ind. Inform.* **2023**, *19*, 2463–2473. [\[CrossRef\]](#)
24. Wan, W.; Chen, J.; Zhou, Z.; Shi, Z. Self-supervised simple Siamese framework for fault diagnosis of rotating machinery with unlabeled samples. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 6380–6392. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Yu, Y.; Gu, X.; Ma, W.; Guo, L.; Gao, H.; Zhang, G. A new method for quantitative estimation of rolling bearings under variable working conditions. *IEEE/ASME Trans. Mechatron.* **2024**, *29*, 41–51. [\[CrossRef\]](#)
26. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2017; pp. 4700–4708. [\[CrossRef\]](#)
27. Hao, Y.; Zhang, Y.; Liu, K.; He, S.; Liu, Z.; Wu, H.; Zhao, J. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 1, pp. 221–231. [\[CrossRef\]](#)
28. Li, H.; Meng, L.; Zhang, J.; Tan, Y.; Ren, Y.; Zhang, H. Multiple description coding based on convolutional auto-encoder. *IEEE Access* **2019**, *7*, 26013–26021. [\[CrossRef\]](#)
29. Guo, Z.; Xu, L.; Zheng, Y.; Xie, J.; Wang, T. Bearing fault diagnostic framework under unknown working conditions based on condition-guided diffusion model. *Measurement* **2025**, *242*, 115951. [\[CrossRef\]](#)
30. Yu, J.; Zhou, X. One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6347–6358. [\[CrossRef\]](#)
31. Yang, Z.; Xu, P.; Yang, Y.; Bao, B.K. A densely connected network based on U-Net for medical image segmentation. *ACM Trans. Multimed. Comput. Commun. Appl.* **2021**, *17*, 89. [\[CrossRef\]](#)
32. Chen, H.; Liu, H.; Chu, X.; Liu, Q.; Xue, D. Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network. *Renew. Energy* **2021**, *172*, 829–840. [\[CrossRef\]](#)
33. Huang, H.; Baddour, N.; Liang, M. A method for tachometer-free and resampling-free bearing fault diagnostics under time-varying speed conditions. *Measurement* **2019**, *134*, 101–117. [\[CrossRef\]](#)

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.