10.37965/jdmd.2026.1583

Intelligent Bearing Fault Diagnosis via Feature Fusion of Multi-Source Heterogeneous Data

Hongzhou Li[1*], Xinkun Yang[1], Meiyu Wu[1]

1. Hebei Huaxi Special Steel Co., Ltd., Tangshan Haigang Development Zone, Tangshan 063611, Hebei, China

*Corresponding author: 18830086161@163.com

Received 28 June 2026; Revised 21 July 2026; Accepted 28 July 2026; Published online 30 July 2026

Slewing bearings in low-speed, heavy-load equipment generate weak and heterogeneous fault signatures that are difficult to characterize using a single sensor. This study develops a compact dual-branch feature-fusion framework that jointly exploits six-channel vibration and one-channel acoustic-emission (AE) signals. To prevent source-record leakage, complete raw recording groups are assigned to training, validation, and test subsets before segmentation; non-overlapping 1024-point windows are then generated, and channel normalization is fitted using training groups only. Five matched random-seed runs are performed, with the best checkpoint selected exclusively by validation Macro-F1. The fusion model achieves mean test accuracy of 99.10% and Macro-F1 of 0.9910, compared with 97.38%/0.9738 for vibration-only and 98.03%/0.9802 for AE-only. The improvement over vibration-only is statistically significant ($p = 0.022$ for both Accuracy and Macro-F1), whereas the improvement over AE-only is numerical but does not reach the 0.05 significance level ($p = 0.063$). The fusion model also obtains the lowest mean Davies-Bouldin index (0.963). These results support compact vibration-AE fusion as an effective diagnostic baseline while also defining its statistical and deployment limitations.

Keywords: Bearing, Fault Diagnosis, Deep Learning, Multi-source Heterogeneous Data, Feature Fusion

1. Introduction

Rolling bearings, as among the most widely adopted critical fundamental components in industrial rotating machinery, directly govern the reliability and operational safety of wind turbines [1], rail transit systems [2], aerospace equipment [3], and high-end CNC machine tools. Subjected to prolonged operation under severe conditions characterized by high speed, heavy load, variable working conditions, and intense noise, bearings are highly susceptible to faults such as fatigue spalling, wear, and fracture. Statistics indicate that approximately 30%–40% of rotating machinery failures can be attributed to bearing degradation [4]. An unscheduled shutdown triggered by bearing failure not only incurs substantial economic losses but may also lead to catastrophic safety incidents. Therefore, conducting research on efficient and accurate intelligent fault diagnosis of bearings carries profound engineering significance for ensuring the safe and reliable operation of industrial equipment and facilitating the transition from reactive maintenance to predictive maintenance.

Early bearing fault diagnosis relied heavily on the empirical judgment of field technicians. With the advancement of sensor technologies and signal processing theories, signal processing-based diagnostic methods have gradually become the mainstream. These methods leverage multi-source health information—such as vibration [5, 6, 7], acoustic emission (AE) [8, 9, 10], and temperature [11, 12]—to extract physical quantities indicative of fault characteristics (e.g., peak value, energy, kurtosis, characteristic frequencies) through time-domain, frequency-domain, or time-frequency domain analysis techniques (e.g., Fourier transform [13], wavelet transform [14]). Fault identification is subsequently achieved via threshold discrimination or traditional machine learning algorithms. However, large-scale mechanical bearings exhibit distinctive characteristics including substantial size, low rotational speed, and complex structure, which result in significant discrepancies in fault mechanisms and signal propagation behavior compared with conventional small- and medium-sized bearings. Their incipient fault signals are inherently weak and frequently submerged in intense background noise, posing considerable challenges to traditional feature extraction methods grounded in physical principles, thereby impeding accurate capture of fault signatures.

In recent years, fueled by the rapid development of big data and computing power, deep learning—as a representative of the new generation of artificial intelligence—has achieved groundbreaking progress in fields such as image recognition and natural language processing [15–17], offering novel avenues for complex mechanical fault diagnosis. Deep learning approaches, particularly convolutional neural networks (CNN) [18, 19, 20] and recurrent neural networks (RNN) [21, 22], possess exceptional nonlinear mapping and automatic feature extraction capabilities. They can directly mine more discriminative deep semantic features from raw, high-dimensional industrial sensing data, effectively overcoming the heavy reliance of traditional methods on expert experience and manual feature engineering.

Nevertheless, most bearing-diagnosis studies still rely on a single sensing modality, usually vibration. Vibration is effective for capturing macroscopic periodic impacts, whereas AE is sensitive to high-frequency elastic waves generated by friction, crack initiation, and local material damage [24, 25]. Recent studies have explored heterogeneous sensing and multimodal learning. Correlation feature distribution matching aligns source and target feature distributions for cross-domain machinery diagnosis [26], while standardized collaborative multiscale Manhattan entropy characterizes nonlinear train-bearing time series at multiple scales [27]. For multimodal diagnosis, audio-vibration fusion with a quadratic CNN [28], acoustic-vibration decomposition followed by CNN-BiGRU-attention learning [29], and current-vibration attention fusion [30] have shown the value of complementary sensors.

10.37965/jdmd.2026.1583

However, these studies primarily address domain adaptation, handcrafted complexity measures, conventional rolling bearings, or other sensor combinations. They do not directly establish a compact raw-signal fusion protocol for low-speed, heavy-load slewing bearings using synchronized multi-channel vibration and AE data.

Accordingly, this study does not claim a new universal fusion operator. Its contribution is application-oriented and methodological:

(1) a modality-specific dual-branch architecture maps six-channel vibration and one-channel AE signals into aligned 128-dimensional representations while preserving their distinct physical characteristics;

(2) parameter-free feature concatenation retains complementary information without introducing a large attention or gating module;

(3) the evaluation uses source-group-disjoint splitting before windowing, training-only normalization, five matched random seeds, validation-based checkpoint selection, independently retrained single-modality baselines, missing-modality stress testing, quantitative feature-space indices, and complexity reporting. This combination provides a reproducible benchmark for heterogeneous slewing-bearing diagnosis and separates predictive gain from potential window-level leakage.

2. Preliminaries

To precisely characterize the feature-fusion method and the strict source-group-disjoint evaluation protocol, the mathematical notation system is summarized in Table 1.

Table 1. Key Mathematical Notations and Definitions

Category	Symbol	Definition and Description
Input	$\mathbf{X}_{raw}$	Raw multi-channel time-series signal tensor
	$\mathbf{X}_{vib} \in \mathbb{R}^{L\times C_v}$	Vibration signal matrix of a single sample (Length L，Channel $C_v = 6$)
	$\mathbf{X}_{ae} \in \mathbb{R}^{L\times C_a}$	Acoustic Emission signal matrix of a single sample (Length L，Channel $C_a = 1$)
	$y \in \{1,2,3,4\}$	Label (0: Normal, 1: Inner, 2: Outer, 3: Ball)
Preprocess	L	Sliding window segment length
	μ, σ	Channel-wise mean and standard deviation for Z-score normalization
	$\mathbf{x}_{vib}, \mathbf{x}_{ae}$	Normalized vibration and acoustic emission sample matrices
Dataset	N_{tr}, N_{val}, N_{te}	Number of samples per fault category for training, validation, and testing, respectively
	D_{bal}	Dataset
Model	$\mathbf{f}_v \in \mathbb{R}^{D_v}$	Deep feature vector extracted by the vibration branch (Dimension $D_v = 128$)
	$\mathbf{f}_a \in \mathbb{R}^{D_a}$	Deep feature vector extracted by the acoustic emission branch (Dimension $D_a = 128$)
	$\mathbf{f}_{fused} \in \mathbb{R}^{D_f}$	Concatenated multimodal feature vector (Dimension $D_f = 256$)
	$\mathbf{z} \in \mathbb{R}^{K}$	Unnormalized score vector from the classifier (Logits，Class K=4)
	$\hat{y}$	Predicted label
Training and evaluation	L_{CE}	Cross-Entropy loss function
	Acc	Accuracy
	$F1_{macro}$	F1 score

This table provides a centralized listing of the key symbols and their definitions referenced in subsequent chapters, along with explicit correspondence to their roles within the methodology. This notation system will be strictly adhered to throughout the methodological implementation and experimental analysis that follow.

3. Methodology

The proposed approach is designed as a compact and reproducible heterogeneous-signal fusion framework rather than as a new generic fusion operator. This section first describes source-group-disjoint dataset construction and leakage-controlled preprocessing, and then presents the modality-specific feature extractors, parameter-free feature concatenation, and final classification layer.

A. Brief Overview of Dataset Construction

To evaluate the proposed method without class-prior bias or source-record leakage, the raw recordings are divided at the source-group level before any sample window is generated. Complete source groups are assigned to training, validation, or test subsets, and no raw CSV file or acquisition group is allowed to appear in more than one subset. In the dataset organization used in this study, each CSV file represents one independent acquisition and is therefore treated as one source group.

After source-group assignment, the synchronized six-channel vibration and one-channel AE sequences are divided into fixed-length windows of L = 1024 samples with a stride of 1024; therefore, adjacent windows share no sampling points. The same temporal boundaries are used for both modalities. Channel-wise Z-score statistics [31] are fitted exclusively from complete non-overlapping windows belonging to the training source groups and are then applied unchanged to the validation and test groups. For channel j, the normalization is given by Equation (1), where μ_j and σ_j are computed from the training groups and epsilon is a small numerical constant:

$$\hat{x}_{ij} = (x_{ij} - \mu_j) / (\sigma_j + \varepsilon) \quad (1)$$

where μ_j and σ_j denote the training-group mean and standard deviation of channel j, respectively, and epsilon is set to a small value to avoid numerical division by zero.

This ordering - source-group assignment, non-overlapping segmentation, and training-only normalization - ensures that test recordings do not contribute samples or preprocessing statistics to model development.

The study covers normal, inner-race fault, outer-race fault, and rolling-element fault states (y in {0, 1, 2, 3}). For each run, a constrained source-group assignment provides 600 training, 200 validation, and 200 test windows per class. Five seeds (42, 7, 2024, 3407, and 1234) generate five matched group-disjoint splits. Within a given seed, the fusion, vibration-only, and AE-only models use exactly the same source groups and selected windows, enabling paired statistical comparison. The balanced protocol controls class-prior effects; it is not described as a correction for an originally imbalanced dataset.

A split audit is saved for every seed, including the source manifest, selected-window manifest, training normalization statistics, and automated checks confirming pairwise-disjoint source groups/files, mutually exclusive sample keys, and zero within-file window overlap.

10.37965/jdmd.2026.1583

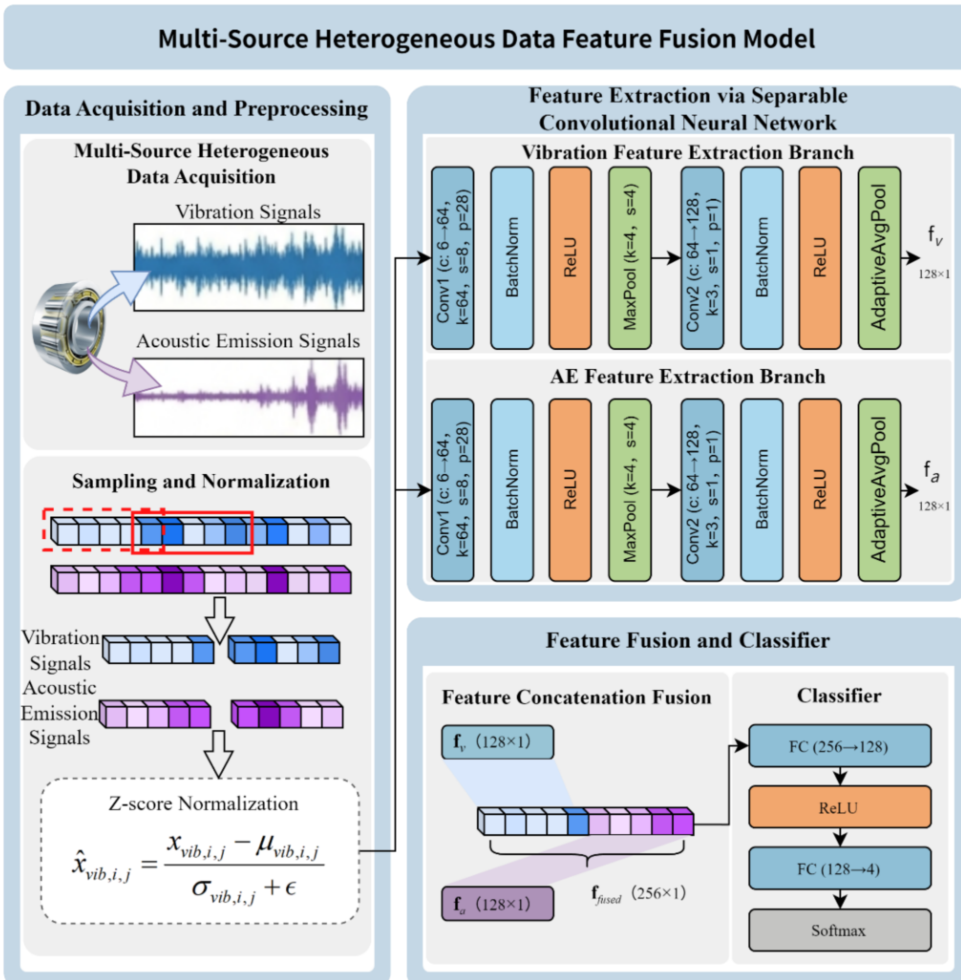


Fig. 1. Schematic diagram of the multi-source heterogeneous data feature fusion model structure.

B. Multi-Source Heterogeneous Data Feature Fusion Network Architecture

To integrate complementary information while keeping the architecture compact, the model uses two modality-specific one-dimensional CNN branches followed by direct feature concatenation. Unlike data-level fusion, which mixes signals with different physical scales before representation learning, the proposed design allows each branch to learn modality-appropriate filters before the embeddings are combined. Unlike attention-based or multi-stage fusion networks, concatenation introduces no additional fusion parameters and reduces the risk of overfitting on the limited slewing-bearing dataset.

(1) Heterogeneous Feature Extraction Branch Design

The first component of the network consists of two parallel heterogeneous feature extraction branches, which process the normalized vibration samples and acoustic emission samples , respectively.

The vibration branch uses an initial kernel width of 64 and stride of 8 to obtain a broad receptive field and rapidly reduce temporal

10.37965/jdmd.2026.1583

resolution. A max-pooling layer with kernel and stride of 4 further suppresses local temporal shifts, after which a kernel-3 convolution refines short-range patterns. Before global pooling, the effective receptive field is approximately 152 input samples. Adaptive average pooling then produces a 128-dimensional global vibration vector. This configuration balances sensitivity to low-speed impact patterns with computational cost.

The AE branch uses the same depth and output dimensionality to make the two embeddings directly comparable, but its first convolution accepts one input channel rather than six. Keeping the branch topology symmetric avoids introducing a capacity advantage for either modality, while the separate weights allow the AE branch to learn high-frequency elastic-wave patterns independently. The output is a 128-dimensional AE vector.

(2) Multi-Source Heterogeneous Data Feature Fusion

The output vectors $\mathbf{f}_v$ and $\mathbf{f}_a$ from the heterogeneous feature extraction branches encapsulate deep discriminative features derived from vibration and acoustic emission signals, respectively, each reflecting distinct physical perspectives and semantic information. To effectively integrate this heterogeneous information and overcome the inherent limitations of single-modality representations in characterizing health states, this study adopts a dedicated feature concatenation fusion mechanism. Following adaptive average pooling, the extracted vibration feature vector $\mathbf{f}_v$ and AE feature vector $\mathbf{f}_a$ are directly concatenated along their feature dimensions to generate a new fused feature vector $\mathbf{f}_{fused}$, as expressed in Equation (2):

$$\mathrm{f}_{fused} = [\mathrm{f}_v, \mathrm{f}_a], dim(f_{fused}) = 256 \quad (2)$$

where $\mathbf{f}_v$ and $\mathbf{f}_a$ denote the deep features from vibration and acoustic emission signals, respectively, and is the fused feature vector.

Given that dim() = 128 and dim() = 128, the dimensionality of equals 256. By integrating health-state discriminative information from two distinct dimensions—vibration and acoustic emission—within a unified deep semantic space, $\mathbf{f}_{fused}$ establishes a more comprehensive feature representation with enhanced discriminative capacity.

(3) Classifier and State Decision Layer

The fused feature vector $\mathbf{f}_{fused}$ is fed into a subsequent classifier composed of a multilayer perceptron (MLP). In this classifier, the first fully connected layer reduces the feature dimensionality from 256 to 128, followed by a Dropout layer to mitigate overfitting. The final fully connected layer maps the feature dimension to the number of fault categories, outputting an unnormalized score vector . Ultimately, the model yields the predicted fault category label $\hat{y}$, as expressed in Equation (3):

$$\hat{y} = argmax_i(z_i), i = 0,1,2,3 \quad (3)$$

where z_i denotes the score for category i.

4. Experiments and Discussion

A. Dataset and Experimental Setup

This study uses the JUST slewing-bearing dataset [32] and evaluates four conditions: normal, inner-race fault, outer-race fault, and rolling-element fault. The main modality comparison is conducted under a strict source-group-disjoint protocol. Raw recording groups are assigned to training, validation, and test subsets before windowing; each selected window has length and stride equal to 1024, and the channel-normalization statistics are fitted from training groups only. Each run contains 600 training, 200 validation, and 200 test windows per class.

The experiments are implemented in PyTorch on Windows 11 using an Intel Core i7-12700K CPU, an NVIDIA GeForce RTX 3060 12-GB GPU, and 32 GB RAM. Adam optimization with a fixed learning rate of 0.001 and cross-entropy loss is used for 100 epochs; no learning-rate scheduler or weight decay is applied. Five random seeds (42, 7, 2024, 3407, and 1234) are used. For each model and seed, the checkpoint with the highest validation Macro-F1 is retained; ties are resolved using validation accuracy and then validation loss. The test set is never used for checkpoint selection. Paired two-sided t-tests are conducted across matched seeds, and Cohen's dz is reported as the paired effect size. Because only five pairs are available, p-values are interpreted together with the effect direction and magnitude.

Table 2. Experimental hyperparameter settings

Hyperparameter / protocol	Value / setting
Input length	1024
Batch size	32
Optimizer	Adam
Dropout	0.3
Learning rate	0.001
Epochs	100
Normalization	Training groups only
Seeds	5 seeds
Repeated trials	5 runs

Overall Accuracy (Acc) and Macro-F1 are used as the principal predictive metrics. Accuracy is the proportion of correctly classified test samples. Macro-F1 is the unweighted arithmetic mean of the four class-wise F1 scores, so every bearing condition contributes equally. To quantify feature geometry in the original learned feature space, the Davies-Bouldin index (DBI) and the ratio of mean intra-class distance to mean inter-class centroid distance (Rdist) are also calculated after L2 normalization. Lower DBI and Rdist values indicate more compact and separated class representations, although these geometric metrics are interpreted jointly with classification performance.

$$F1_{macro} = (1/K) sum_i F1_i \quad (4)$$

where K= 4 denotes the total number of categories, and $F1_i$represents the F1 score for the i-th fault category.

B. Comparative Experimental Results and Analysis

Table 3 reports five matched source-group-disjoint runs. The vibration-only and AE-only baselines are independently initialized and retrained using the same source groups, selected windows, optimization settings, and validation-based checkpoint rule as the fusion model. Fusion achieves the highest mean Accuracy (99.10%) and Macro-F1 (0.9910). Its gains over vibration-only are statistically significant (p = 0.0224 and p = 0.0227), while the numerical gains over AE-only do not reach the 0.05 level with five runs (both p approximately 0.063).

Table 3. Mean ± standard deviation over five strict source-group-disjoint runs

Metric	Vibration	AE	Fusion
Accuracy	0.9738 ± 0.0090	0.9803 ± 0.0127	**0.9910 ± 0.0051**
Macro-F1	0.9738 ± 0.0089	0.9802 ± 0.0127	**0.9910 ± 0.0051**
DBI ↓	1.0913 ± 0.0860	1.0631 ± 0.1225	**0.9631 ± 0.0602**
R_{dist} ↓	0.4864 ± 0.0227	**0.3192 ± 0.0288**	0.3774 ± 0.0148

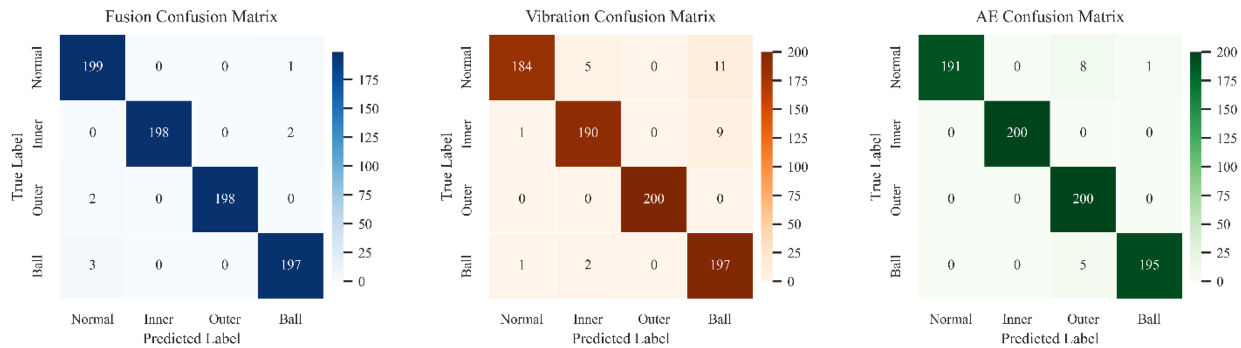


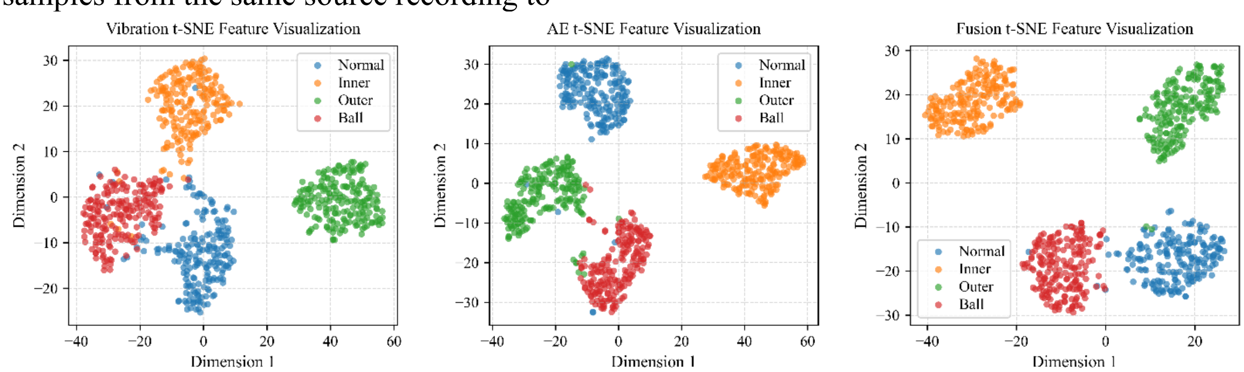
Fig. 2. Representative confusion matrices under the strict source-group-disjoint protocol: (a) vibration + AE fusion, (b) vibration-only, and (c) AE-only.

The representative confusion matrices in Fig. 2 illustrate the class-wise error patterns for one matched split, whereas the primary conclusions are based on the five-run means in Table 3 and the paired statistics reported in the text. The fused model has the highest mean predictive performance, but individual class errors should be interpreted together with run-to-run variation rather than from a single confusion matrix alone.

The independently trained vibration-only and AE-only networks remain strong baselines under the strict split. This result is important because it shows that the reported fusion gain is not obtained by comparing against zero-masked inputs or by allowing samples from the same source recording to enter different subsets. The AE-only result is also substantially stronger under the revised protocol than in the original single-run table, reinforcing the need to report matched repeated trials rather than rely on one favorable split.

C. Feature Visualization Analysis

The test-set representations are visualized using t-SNE only for qualitative inspection. All quantitative feature-separation calculations are performed in the original penultimate-layer feature space, not in the nonlinear two-dimensional t-SNE coordinates.



10.37965/jdmd.2026.1583

Fig. 3. t-SNE visualization of learned features: (a) vibration-only, (b) AE-only, and (c) vibration + AE fusion.

The visualization provides an intuitive view of the learned class organization, but no superiority claim is based on t-SNE alone. The quantitative indices reveal a more nuanced pattern: fusion achieves the lowest mean DBI, whereas the global intra/inter distance ratio does not rank all modalities identically.

Across five runs, the mean DBI values are 1.091 for vibration-only, 1.063 for AE-only, and 0.963 for fusion. Fusion is numerically best, although its paired DBI differences do not reach the 0.05 significance level versus vibration ($p = 0.055$) or AE ($p = 0.089$). The mean Rdist values are 0.486, 0.319, and 0.377, respectively. Fusion significantly reduces Rdist relative to vibration ($p = 0.0011$), but AE yields a lower global ratio than fusion ($p = 0.0058$). This disagreement is not contradictory: DBI emphasizes the most similar cluster pairs, whereas Rdist averages global within-class and between-centroid distances. Consequently, feature geometry and predictive performance are reported jointly rather than reduced to a single ranking.

D. Missing-Modality Stress Test of the Fusion Model

A missing-modality stress test evaluates the selected fusion checkpoint after setting one input stream to zero, without retraining or parameter updating. This experiment is distinct from the independently trained single-modality baselines in Table 3: it tests deployment behavior when a sensor stream is absent from a model trained for complete bimodal input.

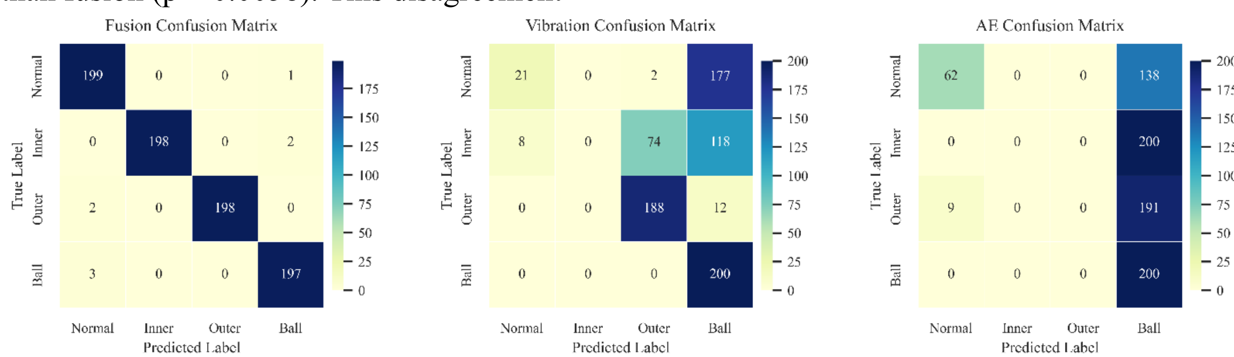


Fig. 4. Missing-modality stress test of a selected fusion checkpoint: (a) complete bimodal input, (b) vibration input with AE set to zero, and (c) AE input with vibration set to zero.

The missing-modality experiment shows that zeroing either stream changes the prediction distribution and degrades performance, confirming that the trained fusion classifier is not interchangeable with an independently trained single-modality model. The degradation is asymmetric, indicating that the two branches play different roles in the joint decision function. In particular, retaining one stream while replacing the other with zeros introduces an input distribution that was not observed during training. Therefore, the result should be interpreted as a sensor-loss stress test, not as an estimate of the standalone diagnostic capability of vibration or AE. Standalone capability is assessed only by the separately trained models in Table 3.

The stress test provides supporting evidence that both inputs influence the learned decision function, but zero masking cannot establish equal causal contribution or exclude partial modality dominance. For deployment, a sensor-health check and a fallback independently trained single-modality model are advisable.

Accordingly, missing-sensor robustness and multimodal complementarity are treated as separate properties: the current model benefits from complete synchronized input but is not designed as a missing-modality architecture.

E. Exploratory Evaluation under Different Conditions

Table 4 summarizes the previously conducted condition-wise comparison across different rotational speeds and overturning forces. Because these results were obtained as an exploratory within-dataset analysis and are not part of the five paired strict-split runs used for the main statistical claims, they are interpreted only as preliminary evidence of condition sensitivity rather than as proof of cross-dataset or cross-machine generalization.

Table 4. Exploratory results under different operating conditions.

Condition	Rotational Speed/rpm	Overturning Force/N	Vibration ACC/%	AE ACC/%	Fusion ACC/%
1	2	0	97.38	98.03	99.10
2	2	30	83.25	71.50	91.37
3	2	60	90.00	81.87	97.88
4	6	0	99.00	59.13	99.25

Under Condition 2, the fusion model retains a larger accuracy margin over vibration-only than under the nominal condition. This trend is physically plausible because low-speed loaded operation can weaken periodic vibration impacts while AE remains sensitive to local elastic-wave activity. Nevertheless, the condition-wise table should be treated as exploratory until all operating conditions are rerun using the same source-group-disjoint repeated-seed protocol.

The exploratory condition results and the missing-modality test address different questions. The former examines changes in modality informativeness across operating conditions, whereas the latter imposes complete sensor loss. Neither analysis is used to overstate cross-domain generalization.

F. Complexity Analysis

Model complexity is estimated for a 1024-point input. Multiply-accumulate operations (MACs) include convolutional and fully connected layers and exclude batch normalization, activation, and pooling overhead. The complete fusion model has 112,388 trainable parameters, approximately 5.28 million MACs per sample, and an FP32 parameter storage of about 0.45 MB. Thus, the improved mean predictive performance is obtained with a moderate computational increase and without a parameterized fusion module.

Table 5. Model complexity for one 1024-point sample.

Model	Parameters	MACs/sample
Vib-only	66,756	3.95 M
AE-only	46,276	1.33 M
Fusion	112,388	5.28 M

Conclusion

This study develops a compact vibration-AE feature-fusion framework for slewing-bearing diagnosis and evaluates it using source-group-disjoint splitting, non-overlapping windows, training-only normalization, validation-based checkpoint selection, and five matched random-seed runs. The principal conclusions are as follows:

(1) Fusion provides the highest mean predictive performance under the strict split. It achieves 99.10% Accuracy and 0.9910 Macro-F1, compared with 97.38%/0.9738 for vibration-only and 98.03%/0.9802 for AE-only. The gain over vibration-only is statistically significant, whereas the gain over AE-only is numerical but not statistically conclusive at the 0.05 level with five runs.

(2) The feature-space analysis supports a qualified, rather than universal, separation advantage. Fusion has the lowest mean DBI and significantly improves the intra/inter ratio relative to vibration-only, but AE-only obtains a lower global intra/inter ratio. The two geometric indices and the classification metrics therefore provide complementary evidence and should be interpreted together.

(3) The exploratory multi-condition results suggest that modality complementarity may become more valuable when operating conditions reduce the informativeness of vibration signals.

(4) The selected fusion model is sensitive to missing input streams. Zero masking changes the prediction distribution and degrades performance, so complete bimodal input, sensor-health monitoring, or a fallback independently trained single-modality model is required for reliable deployment.

In summary, the proposed network is best viewed as a compact and reproducible baseline for synchronized vibration-AE diagnosis. Its contribution lies in modality-specific representation learning, parameter-free fusion, strict leakage-controlled evaluation, paired statistical analysis, and explicit reporting of limitations. Future work will include strict repeated evaluation under all operating conditions, same-dataset implementation of advanced multimodal baselines, and explicit missing-modality and modality-attribution mechanisms.

References

[1] Guan, J., Wang, Y. "Fault diagnosis strategy for wind turbine bearings based on improved high-dimensional Bayesian optimized CNN." *J Mech Sci Technol*, vol. 39, 6655–6667, 2025. https://doi.org/10.1007/s12206-025-1012-9

[2] Liu, Q., et al. "A Novel Method of Bearing Fault Diagnosis for Train Bogie Transmission System Based on MpResCNN-BiLSTM Model with Attention Mechanism." *IEEE Trans. Intell. Transp. Syst.*, 2025. doi: 10.1109/TITS.2025.3557063

[3] Zhou, M., Yu, Y., Qu, C., Xiang, G., Yang, Z. "A Fault Diagnosis Method for Aerospace Bearings Based on CS-WGAN-PSO." *2025 Global Reliability and Prognostics and Health Management Conference (PHM-Xian)*, Xian, China, 2025, pp. 1–7. doi: 10.1109/PHM-Xian66756.2025.11427439

[4] Gawde, S., Patil, S., Kumar, S., et al. "A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion." *Artif. Intell. Rev.*, vol. 56, no. 5, 4711–4764, 2023.

[5] Xu, Y., Tang, X., Feng, G., Wang, D., Ashworth, C., Gu, F., Ball, A. "Orthogonal On-Rotor Sensing Vibrations for Condition Monitoring of Rotating Machines." *J. Dyn. Monit. Diagn.*, vol. 1, no. 1, 29–36, 2021. https://doi.org/10.37965/jdmd.v2i2.47

[6] Zhang, W., Li, X., Ding, Q. "Deep residual learning-based fault diagnosis method for rotating machinery." *ISA Trans.*, vol. 95, 295–305, 2019.

[7] Hou, L., Yi, H., Jin, Y., Gui, M., Sui, L., Zhang, J., Chen, Y. "Inter-shaft Bearing Fault Diagnosis Based on Aero-engine System: A Benchmarking Dataset Study." *J. Dyn. Monit. Diagn.*, vol. 2, no. 4, 228–242, 2023. https://doi.org/10.37965/jdmd.2023.314

[8] Kim, J., Kim, J.-M. "Bearing Fault Diagnosis Using Grad-CAM and Acoustic Emission Signals." *Appl. Sci.*, vol. 10, no. 6, 2050, 2020. https://doi.org/10.3390/app10062050

[9] Zhang, Y., Yu, Y., Yang, Z., Liu, Q. "Rolling bearing fault identification with acoustic emission signal based on variable-pooling multiscale convolutional neural networks." *Sci Rep*, vol. 15, 15644, 2025. https://doi.org/10.1038/s41598-025-00573-7

[10] Fan, Z., Gao, R. X., He, Q., Huang, Y., Jiang, T., Peng, Z., Thévenaz, L., Xiong, Y., Zhong, S. "New Sensing Technologies for Monitoring Machinery, Structures, and Manufacturing Processes." *J. Dyn. Monit. Diagn.*, vol. 2, no. 2, 69–88, 2023. https://doi.org/10.37965/jdmd.2023.322

[11] Yang, Z., Wu, B., Shao, J., Lu, X., Zhang, L., Xu, Y., Chen, G. "Fault detection of high-speed train axle bearings based on a hybridized physical and data-driven temperature model." *Mech. Syst. Signal Process.*, vol. 208, 111037, 2024. https://doi.org/10.1016/j.ymssp.2023.111037

[12] Abtane, M., Dahi, K., Martinez, H., Sedki, M., El Kimi, H., Dahhassi, C., Hlib, C., Borges, L. F. "Axle bearing fault diagnosis for high-speed trains: A comprehensive review of methodologies, technologies, challenges and emerging trends." *Measurement*, vol. 251, 117098, 2025. https://doi.org/10.1016/j.measurement.2025.117098

[13] Yang, G., Li, M., Jiao, D., Shi, L., Zhang, K., Xu, Y. "Fault Feature Enhancement and Diagnosis Method for UAV Rotor Bearings Based on Amplitude-Modulated Graph Fourier Spectrum." *J. Dyn. Monit. Diagn.*, 2026. https://doi.org/10.37965/jdmd.2026.1382

[14] Helm, D., Timusk, M. "Wavelet Denoising Applied to Hardware Redundant Systems for Rolling Element Bearing Fault Detection." *J. Dyn. Monit. Diagn.*, vol. 2, no. 2, 102–114, 2023. https://doi.org/10.37965/jdmd.2023.231

[15] He, K., Zhang, X., Ren, S., Sun, J. "Deep residual learning for image recognition." *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[16] Vaswani, A., Shazeer, N., Parmar, N., et al. "Attention is all you need." *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[17] Fukushima, K. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." *Biol. Cybern.*, vol. 36, no. 4, 193–202, 1980.

[18] Peng, D., Wang, H., Desmet, W., Gryllias, K. "RMA-CNN: A Residual Mixed-Domain Attention CNN for Bearings Fault Diagnosis and its Time-Frequency Domain Interpretability." *J. Dyn. Monit. Diagn.*, vol. 2, no. 2, 115–132, 2023. https://doi.org/10.37965/jdmd.2023.156

[19] Li, X., Zhang, W., Ma, H., Luo, Z. "Domain generalization in rotating machinery fault diagnostics using deep neural networks." *Neurocomputing*, vol. 403, 409–420, 2020.

[20] Ji, Y., Gao, J., Shao, X., Wang, C. "Multi-scale quadratic convolutional neural network for bearing fault diagnosis based on multi-sensor data fusion." *Nonlinear Dyn.*, vol. 113, 14223–14244, 2025. https://doi.org/10.1007/s11071-025-10918-6

[21] Liu, H., Zhou, J., Zheng, Y., Jiang, W., Zhang, Y. "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders." *ISA Trans.*, vol. 77, 167–178, 2018. https://doi.org/10.1016/j.isatra.2018.04.005

[22] An, Z., Li, S., Wang, J., Jiang, X. "A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network." *ISA Trans.*, vol. 100, 155–170, 2020. https://doi.org/10.1016/j.isatra.2019.11.010

[23] Sinkittiyanont, S., Ratanasumawong, C. "Fault Diagnosis of Rolling Element Bearing from Vibration Signal Using One Dimensional Convolutional Neural Networks." *2024 21st Int. Jt. Conf. Comput. Sci. Softw. Eng. (JCSSE)*, 2024, pp. 390–397. doi: 10.1109/JCSSE61278.2024.10613660

[24] Tang, L., Wu, X., Wang, D., Liu, X. "A Comparative Experimental Study of Vibration and Acoustic Emission on Fault Diagnosis of Low-Speed Bearing." *IEEE Trans. Instrum. Meas.*, vol. 72, 1–11, 2023. https://doi.org/10.1109/TIM.2023.3312761

[25] Le, V.-L., Nguyen, H.-A.-H., Kim, C.H. "Efficient Dual-Stream Network with Soft-Gated Fusion for Bearing Fault Diagnosis Using Acoustic Emission Signals." *Machines*, vol. 14, no. 4, 414, 2026. https://doi.org/10.3390/machines14040414

[26] Tan, H., Xie, S., Ma, W., Yang, C., Zheng, S. "Correlation feature distribution matching for fault diagnosis of machines." Reliability Engineering & System Safety, vol. 231, 108981, 2023. https://doi.org/10.1016/j.ress.2022.108981

[27] Tan, H., Su, Y., Xie, S., et al. "Standardized collaborative multiscale Manhattan entropy: A nonlinear time series metric for train bearing fault diagnosis." Digital Signal Processing, vol. 182, 106322, 2026. https://doi.org/10.1016/j.dsp.2026.106322

[28] Yan, J., Liao, J.-B., Gao, J.-Y., Zhang, W.-W., Huang, C.-M., Yu, H.-L. "Fusion of Audio and Vibration Signals for Bearing Fault Diagnosis Based on a Quadratic Convolution Neural Network." Sensors, vol. 23, no. 22, 9155, 2023. https://doi.org/10.3390/s23229155

[29] Jiang, M., Luo, M., Zhang, C., Shu, M., Sun, G. "Rolling bearing fault diagnosis based on acoustic-vibration data fusion and mode decomposition combined with the crested porcupine optimization algorithm." Heliyon, vol. 10, no. 22, e40351, 2024. https://doi.org/10.1016/j.heliyon.2024.e40351

[30] Yang, J., Han, H., Dong, X., Wang, G., Zhang, S. "Bearing Fault Diagnosis Grounded in the Multi-Modal Fusion and Attention Mechanism." Applied Sciences, vol. 15, no. 3, 1531, 2025. https://doi.org/10.3390/app15031531

[31] Shalabi, L. A., Shaaban, Z., Kasasbeh, B. "Data Mining: A Preprocessing Engine." *J. Comput. Sci.*, vol. 2, no. 9, 2006. doi: 10.3844/jcssp.2006.735.739

[32] Ren, X. "JUST Slewing bearing datasets-1." *Mendeley Data*, V1, 2023. [Online]. Available: https://data.mendeley.com/datasets/hwg8v5j8t6/1. doi: 10.17632/hwg8v5j8t6.1