JOURNAL OF Emerging Paradigms in Computing Systems JEPCS

Journal of Emerging Paradigms in Computing Systems (JEPCS) | Volume 03, Issue No. 02, 2026

ARTICLE TYPE: Research Article

Article ID: JEPCS-2026-0302-0013

An Explainable Multi Sensor and Multi Domain Feature Fusion Approach for Bearing Fault Diagnosis Using Wavelet Transform and Transformer Networks

Kaoutar Bizmour[1,*], **Gnalen Konate**[2]

[1] School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, 221116, China
[2] School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou, 221116, China

*Corresponding author: 27230007@cumt.edu.cn
Contributing authors: 27230007@cumt.edu.cn, fs25030006@cumt.edu.cn

Received: August 28, 2026 **Accepted:** September 30, 2026

Abstract- Effective use of heterogeneous monitoring signals is necessary for reliable bearing defect identification while preserving diagnostic robustness, accuracy, and interpretability. For the purpose of diagnosing bearing conditions, this study proposes an explainable multi-sensor and multi-domain framework that integrates vibration and two-phase motor current information. To characterize complementary fault-related behavior, time-domain statistical features, frequency-domain spectral information, and time-frequency representations are extracted. The extracted features are integrated into a Transformer-based architecture that uses self-attention to model interactions among diagnostic representations and sensing modalities. A three-class subset of the Paderborn bearing dataset, comprising healthy (HH), inner-ring fault (IR), and outer-ring fault (OR) conditions, is used to evaluate the proposed framework. Comparative experiments with support vector machine, random forest, and multilayer perceptron classifiers demonstrate strong class separability under clean measurement conditions; the proposed Transformer achieves 100% accuracy and a 100% macro-F1 score. Noise-sensitivity experiments reveal substantial performance degradation under signal perturbation, highlighting the distinction between clean-data accuracy and operational robustness. Sensor-ablation analysis further shows that vibration measurements provide stronger diagnostic information than motor-current measurements. Additionally, SHAP-based explainability analysis shows that time-frequency characteristics provide the largest domain-level contribution at 51.56%, while vibration-derived features constitute the largest individual sensor-level contribution at 42.01%. Therefore, the proposed framework provides a systematic approach to multi-sensor bearing condition monitoring by combining accurate fault discrimination with sensor-level and feature-domain interpretability while emphasizing the importance of robustness evaluation under degraded measurement conditions.

Keywords- Bearing fault diagnosis; multi-sensor fusion; transformer; vibration analysis; motor current analysis

1. Introduction

Rolling bearings, which are frequently found in motors, gearboxes, turbines, pumps, and manufacturing equipment, are essential parts of rotating machinery. The dependability, effectiveness, and lifespan of mechanical systems are all directly impacted by their state of operation. Localized failures on rolling elements and bearing raceways can be caused by environmental disturbances, fluctuating loading, lubrication deterioration, and recurrent contact stresses during long-term operation [1]-[4]. Abnormal vibration, increased noise, decreased transmission efficiency, and ultimately an unplanned machinery shutdown are all possible outcomes of such faults.

As a result, condition-based maintenance and intelligent machinery health management depend heavily on precise bearing failure identification. Because localized flaws produce mechanical responses that can be detected from the bearing housing, vibration analysis is still one of the most well-established methods for monitoring bearing conditions. Complementary signal representations can be used to characterize fault information. While frequency domain analysis shows spectral features related to bearing dynamics, time domain indicators like root mean square, kurtosis, and crest factor characterize signal amplitude and impulsiveness. Time frequency methods, such wavelet transforms, offer localized representations of transient components for non-stationary fault signals that could not be sufficiently captured by either the time or frequency domain alone.

However, environmental interference, transmission routes, and sensor position can all have an impact on relying solely on vibration data. Because mechanical disturbances can alter the electromechanical response of the drive system, motor-current signals offer an additional and relatively non-intrusive source of condition information. The Paderborn bearing dataset was introduced by Lessmeier et al. [5] to support the development and evaluation of data-driven bearing condition monitoring methods using complementary vibration and motor-current measurements. Automatic feature learning for equipment diagnosis has significantly improved with recent developments in deep learning and machine learning [6]-[7]. While convolutional and recurrent neural networks can automatically extract spatial and temporal characteristics from measured signals, conventional methods like support vector machines and ensemble learning algorithms can provide effective classification when informative handcrafted features are available. Transformer designs have garnered more attention recently because to their ability to model global interdependence across diagnostic features using self-attention methods. In order to handle the heterogeneity and redundancy of multi sensor readings, Li et al. [8] suggested a multi branch feature cross fusion model based on Transformer learning in multi-sensor bearing diagnosis. Their findings showed the possibility of sensor specific feature extraction followed by global information fusion, but they also showed that good diagnostic representation is not always ensured by merely expanding the number of sensing channels.

There are still a number of restrictions in spite of these advancements. First, a lot of diagnostic methods focus on the accuracy of the final classification without adequately identifying the sensing modality or signal representation that yields the observed improvement. This is especially crucial for multi sensor and multi domain diagnosis, as fault discrimination may be influenced differently by vibration, motor current, time domain, frequency domain, and time frequency information. Second, robustness under measurement degradation or noise is not always demonstrated by high accuracy gained from pure laboratory signals. Third, it can be challenging to ascertain whether the predictions made by increasingly intricate deep learning architectures are supported by significant fault features since they can operate as black box classifiers. In situations when black-box artificial intelligence cannot be easily trusted, Magadán et al. [9] stressed the significance of explainable and interpretable bearing diagnosis. In a similar vein, SHAP-based methods have shown that feature attribution can be utilized to increase the transparency of bearing defect classification and discover diagnostically significant factors Brito et al. [10]. Therefore, a bearing diagnostic method that assesses sensor complementarity, multi domain information, classification capability, robustness, and interpretability within a single experimental methodology is still required. The separate contributions of vibration and motor current measurements should be measured instead of viewing feature fusion as intrinsically advantageous. Similarly, controlled ablation studies should be used to determine the contribution of time, frequency, and time frequency information. In order to interpret diagnostic conclusions from both data-driven and mechanical viewpoints, explainability should be further linked to measurable signal features.

This study leverages synchronized vibration and motor-current measurements to propose an explainable multi-sensor and multi domain feature fusion approach for intelligent bearing fault diagnostics in order to overcome these problems. Healthy condition (HH), inner ring fault (IR), and outer-ring fault (OR) are the three bearing health conditions that are taken into consideration. Transformer based learning for fault classification is combined with time domain statistical features, frequency domain data, and wavelet-based time frequency representations. The primary contributions are as follows:

- Vibration and two motor-current signals are used in tandem to explore complementary mechanical and electrical fault information;
- Time, frequency, and wavelet-based time frequency representations are integrated to characterize bearing faults from multiple signal domains;
- ablation experiments at the sensor and domain levels are designed to quantify the actual contributions of individual information sources; and
- explainability and noise robustness analyzes are incorporated to assess model stability and ascertain whether diagnostic choices can be linked to significant bearing signal characteristics.

The remainder of this paper is organized as follows. Section 2 describes the dataset, selected bearing conditions, signal preprocessing, and preliminary signal characteristics. Section 3 presents the proposed multi-sensor and multi-domain diagnostic method. Section 4 describes the experimental design, benchmark models, evaluation metrics, and ablation and robustness tests. Section 5 presents and discusses the experimental results and explainability analysis, while Section 6 summarizes the main conclusions and limitations.

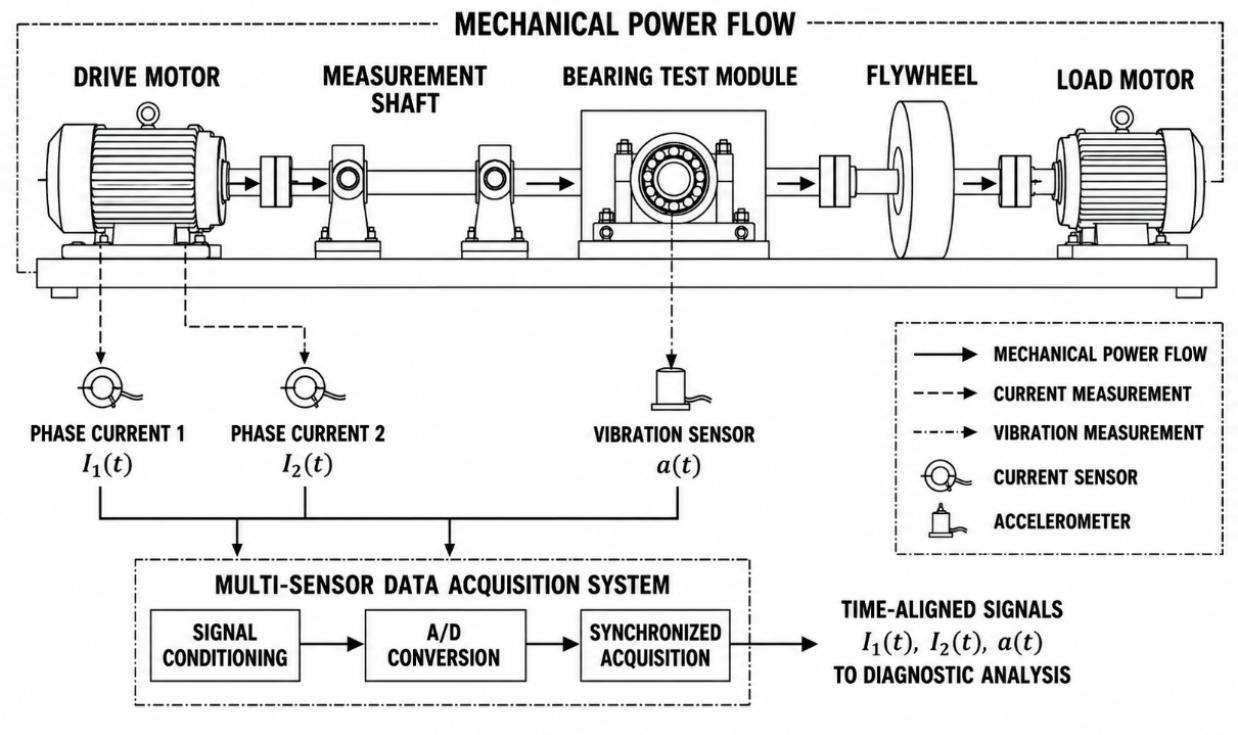


Figure 1: *Schematic representation of the Paderborn bearing test system and multi-sensor measurement configuration.*

Table 1: *Description of the bearing health conditions and multi-sensor data used in this study.*

Class	Code	Bearing condition	Measurement channels	Observations
Healthy	HH	Undamaged/healthy	Phase current 1, phase current 2, vibration	256,001
Inner-ring fault	IR	Inner-ring damage	Phase current 1, phase current 2, vibration	258,008
Outer-ring fault	OR	Outer-ring damage	Phase current 1, phase current 2, vibration	256,000
Total	–	3 health states	3 channels	770,009

2. Dataset and Signal Characterization

2.1 Dataset Description and Fault Categories

This study's experimental data obtained from the Paderborn University Bearing Data Center, which offers benchmark measurements for data driven electromechanical drive system status monitoring. The driving motor, measuring shaft, bearing test module, flywheel, and load motor are the main components of the original test platform. Vibration based and motor current based condition evaluation were made possible by the recording of both mechanical and electrical responses throughout the bearing testing. This benchmark dataset was created by Lei et al. [11] in order to aid in the creation and assessment of data-driven bearing diagnostic techniques utilizing these complementing metrics. The experimental setup allows mechanical vibration responses from the bearing module to be taken into account with electrical current responses from the drive system, as shown in Figure 1.

The processed subset from the Paderborn measurements is taken into consideration for the current study. Healthy condition (HH), inner ring fault (IR), and outer ring fault (OR) are the three representative bearing health conditions. Three measurement channels phase current 1, phase current 2, and vibration are present in each class. The HH, IR, and OR datasets contain 256,001, 258,008, and 256,000 observations, respectively, totaling 770,009 observations prior to segmentation, according to a direct analysis of the processed files. Table 1 summarizes the bearing circumstances, number of available observations, and sensor channels.

The three health states' comparatively equal numbers of observations offer a good foundation for building a balanced fault classification experiment. More significantly, the simultaneous availability of two-phase current and vibration measurements enables the separate and combined evaluation of the diagnostic contribution of mechanical and electrical data. However, operating condition labels and individual bearing specimen identities are not specifically retained in the processed CSV subset. Thus, rather than making unsupported claims about cross operating condition generalization, the current study concentrates on multi-sensor fault discrimination, multi-domain feature representation, noise sensitivity assessment, and explainability.

Table 2: *Multi-sensor preprocessing and segmentation configuration.*

Parameter	Setting	Function
Input channels	3	Vibration + two phase currents
Diagnostic classes	3	HH, IR, OR
Channel scaling	Independent standardization	Reduce differences in numerical scale
Window length	4,096 samples	Local signal representation
Window stride	2,048 samples	Consecutive-window displacement
Window overlap	50%	Controlled overlap within each data subset
Generated segments	361 (HH: 120; IR: 121; OR: 120)	Synchronized multi-sensor samples
Channel alignment	Identical temporal indices	Preserve sensor synchronization
Time-domain representation	Statistical signal characteristics	Characterize amplitude and impulsiveness
Frequency-domain representation	FFT-based characteristics	Characterize spectral information
Time-frequency representation	Wavelet-based analysis	Characterize transient/non-stationary information
Dataset partitioning	Chronological 70%/15%/15% split before segmentation	Prevent cross-subset window leakage

2.2 Multi Sensor Signal Preprocessing and Segmentation

Direct integration of the original signals may result in channels with greater numerical magnitudes having a disproportionate impact on learning because vibration and motor current measurements represent distinct physical characteristics and have different numerical ranges. The synchronized multi-sensor measurement at time t can be expressed as follows

$$X(t) = \left[x_{I_1}(t),\, x_{I_2}(t),\, x_v(t)\right]^T \tag{1}$$

where $x_v(t)$ represents the vibration signal and $xI_1(t)$ and $xI_2(t)$ indicate the first and second phase current measurements, respectively. Each sensing channel is standardized separately in order to minimize scale disparities. The standardized value for a measurement x_i is computed as

$$z_i = \frac{x_i - \mu_x}{\sigma_x} \tag{2}$$

where μ_x and σ_x represent the respective channel's mean and standard deviation. Crucially, in order to prevent information leaking from the evaluation sets, the normalization parameters required for model assessment will be calculated from the training data and then applied to the validation and testing data. After then, fixed length signal segments are created from the continuous synchronized measurements. The j-th multi-sensor sample for a segment length L is represented as

$$S_j = X\left[s_j : s_j + L - 1\right] \tag{3}$$

The initial index of the j-th segment is indicated by s_j. The temporal relationship of the two current channels and the vibration channel is preserved by maintaining identical segmentation indices. The continuous synchronized measurements were divided chronologically into training (70%), validation (15%), and testing (15%) blocks before segmentation to prevent overlapping or temporally adjacent windows from being distributed across different subsets. Within each block, fixed-length windows of 4,096 samples were generated using a stride of 2,048 samples, corresponding to 50% overlap between consecutive windows. Identical window boundaries were applied to the vibration and two phase-current channels to preserve temporal synchronization. Using this procedure, the HH dataset generated 86 training, 17 validation, and 17 testing segments (120 segments in total); the IR dataset generated 87 training, 17 validation, and 17 testing segments (121 segments in total); and the OR dataset generated 86 training, 17 validation, and 17 testing segments (120 segments in total). Thus, 361 synchronized multi-sensor segments were generated overall. No window crossed a training, validation, or testing boundary. Table 2 provides a summary of the main preprocessing and segmentation settings.

While preparing the measurements for later multi domain feature extraction, our preprocessing technique maintains synchronization across the three sensing channels. Because signal segments produced from the same continuous sequence can be strongly correlated, special attention is paid to the partitioning approach. Overly optimistic diagnostic performance may result from randomly allocating closely adjacent or overlapping segments between training and testing sets. As a result, rather than just being data preparation procedures, segmentation and data division are seen as essential elements of the experimental approach.

2.3 Preliminary Signal Characteristics

The raw signals are analyzed to determine whether the three bearing states show discernible mechanical and electrical responses prior to creating the diagnostic model. Representative vibration and phase current signals from the HH, IR, and OR datasets are shown in Figure 2 below. While the two-phase current channels offer complementing information

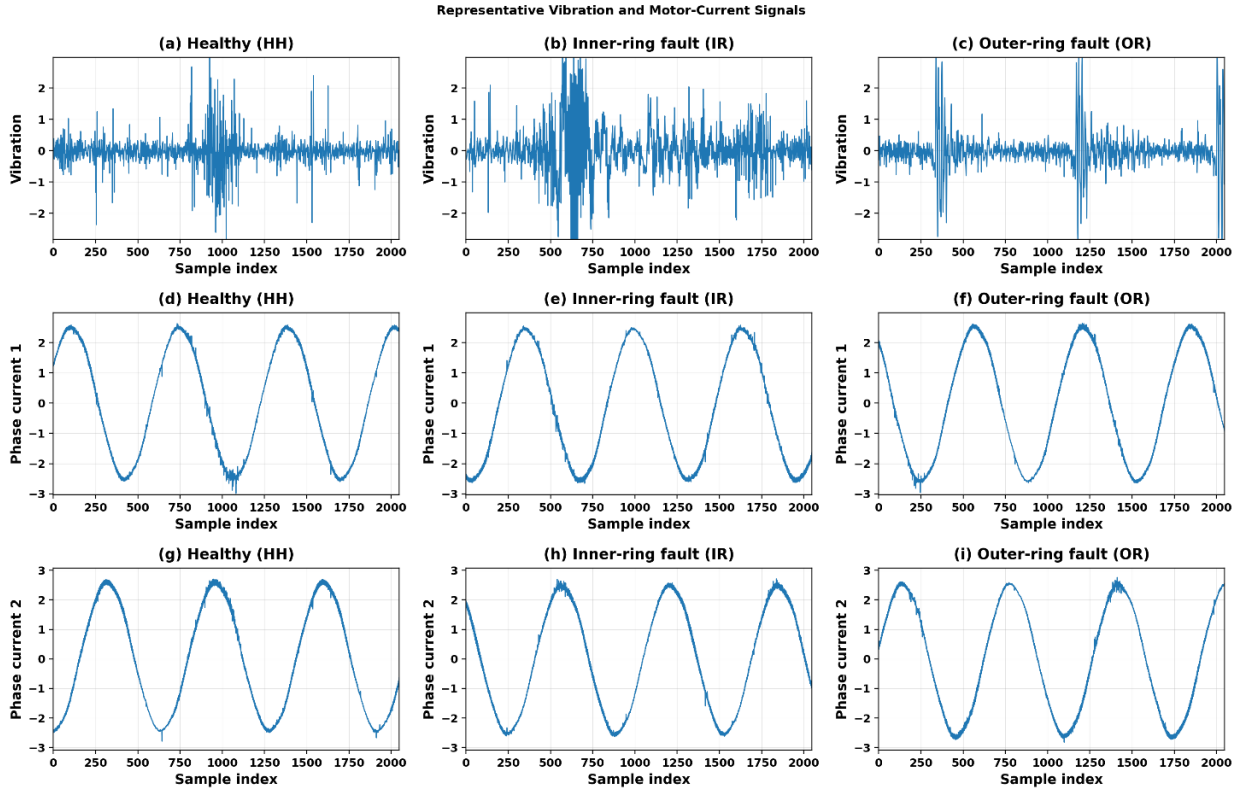


Figure 2: *Representative vibration and motor current signals under healthy (HH), inner-ring fault (IR), and outer-ring fault (OR) bearing conditions.*

related to the electromechanical reaction of the drive system, the vibration measurements offer direct information about the mechanical response of the bearing system.

The aim of the initial comparison shown in Figure 2 above is to determine how the three bearing situations differ in terms of signal amplitude, fluctuation intensity, impulsive behavior, and temporal distribution. The non-stationary reactions linked to localized bearing damage, however, might not be adequately described by variations seen directly in the temporal domain. As a result, the suggested methodology uses frequency domain and time-frequency representations to further examine the signals. This gives the rationale for combining complementing time, frequency, and wavelet-based time frequency data instead of depending only on one. The dataset's multi-sensor design also makes it possible to experimentally assess the relative diagnostic value of vibration and motor current data. Subsequent sensor ablation investigations evaluate vibration only, current only, and vibration current fusion combinations instead of assuming that fusion inevitably improves classification [12]-[13]. To ascertain whether integrating time, frequency, and time-frequency information offers quantifiable advantages over separate signal domains, domain level ablation is also utilized.

3. Proposed Multi Sensor and Multi Domain Fault Diagnosis Method

3.1. Overall Diagnostic Architecture

The suggested diagnostic technique combines complementary signal-domain representations, attention-based feature learning, and synchronized vibration and motor-current measurements. Multi-sensor signal input, signal preprocessing and segmentation, multi-domain representation, feature fusion and Transformer-based classification, and explainability analysis are the five basic steps of the methodology, as seen in Figure 3. To maintain the relationship between mechanical and electrical responses, the vibration signal x_v and two phase-current signals xI_1 and xI_2 are processed simultaneously for each diagnostic sample.

Complementary representations develop from the segmented signals instead of giving the classifier just the raw values. The frequency domain branch defines spectral energy distributions, the time domain branch uses wavelet analysis to capture localized non-stationary characteristics, and the time domain branch specifies statistical qualities related to signal amplitude, dispersion, and impulsiveness. To create a thorough representation of the bearing condition, these representations are then encoded and combined. Before being classified into HH, IR, or OR states, the fused representation is fed into a Transformer-based classifier that uses self-attention to build relationships among the retrieved diagnostic data.

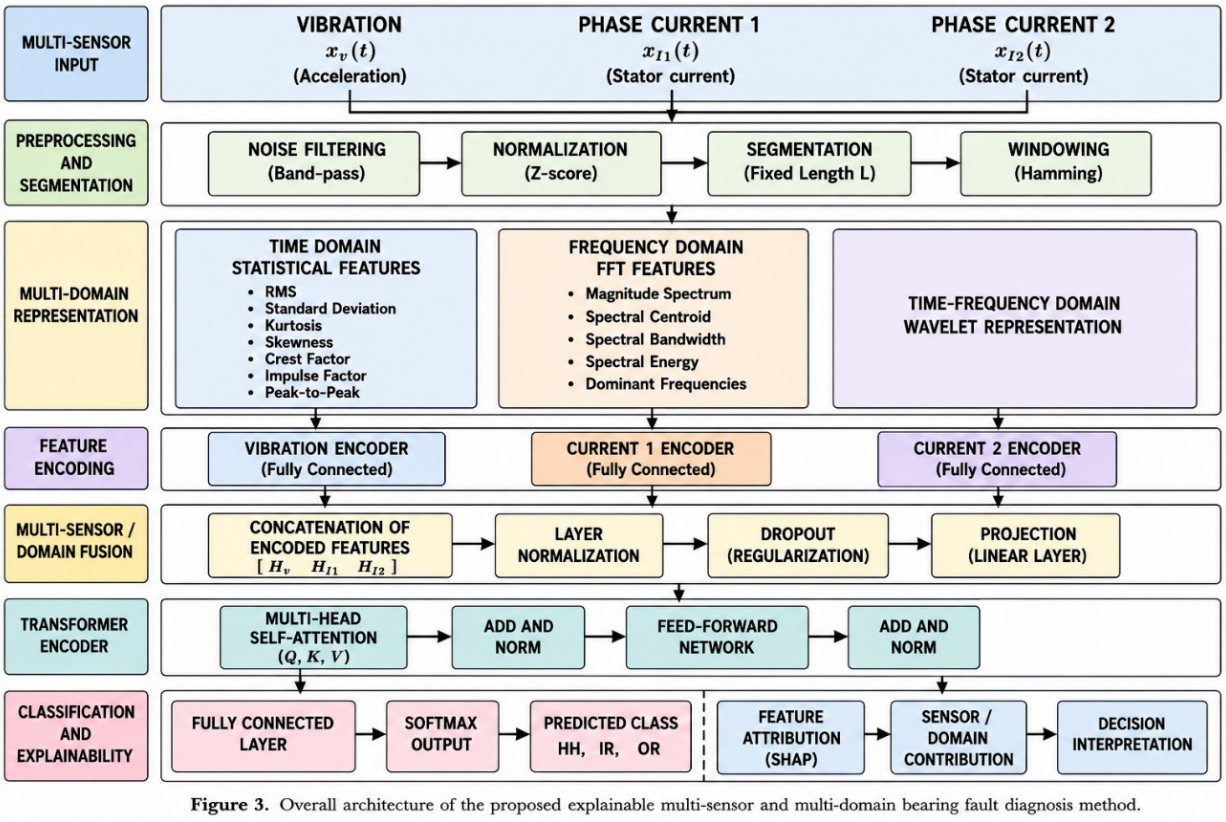


Figure 3. Overall architecture of the proposed explainable multi-sensor and multi-domain bearing fault diagnosis method.

Figure 3: *General layout architecture of the proposed explainable multi-sensor and multi-domain bearing fault diagnosis method.*

The suggested methodology varies in that it does not presume that information fusion is intrinsically advantageous. To measure the contributions of vibration, motor current, and specific signal domain representations, the experimental design incorporates sensor level and domain level ablation studies. Additionally, the explainability module looks at how various sensors and diagnostic features contribute to the anticipated bearing states, while noise-sensitivity analysis assesses the diagnostic model's stability under gradually deteriorating signal conditions. As a result, the methodology assesses not only whether the model accurately detects a fault but also which information influences the conclusion and how stable it is under signal degradation.

3.2. Multi Domain Signal Representation

In the time, frequency, and time frequency domains, bearing failures can produce signal characteristics that vary. As a result, a single representation might not fully capture the underlying fault response. Each synchronized signal segment in the suggested method is described using wavelet-based time frequency information, frequency domain spectrum information, and complementary time domain statistical markers. These representations characterize various features of the HH, IR, and OR signals and supply the multi domain information needed for further feature fusion, as seen later in Figure 4.

3.2.1. Multi Domain Signal Representation

Without changing the initial observations, time-domain features offer concise statistical descriptions of signal amplitude, dispersion, impulsiveness, and waveform properties. Because localized flaws can generate brief oscillations and changes in the signal energy distribution, these characteristics are very important for bearing diagnostics. A collection of statistical indicators, including as root mean square (RMS), standard deviation, peak to peak value, skewness, kurtosis, crest factor, and impulse factor, are thus collected for every vibration and motor current segment. The RMS value is computed as follows for a discrete signal segment x (n) with N observations

$$x_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)} \tag{4}$$

Although the expression for kurtosis is

$$K = \frac{\frac{1}{N}\sum_{n=1}^{N}[x(n)-\mu]^4}{\sigma^4} \tag{5}$$

Where μ and σ stand for the signal segment's mean and standard deviation, respectively. While kurtosis is sensitive to impulsive behavior that may result from localized bearing damage, RMS describes the overall signal energy level. Waveform asymmetry, amplitude range, and peak characteristics are described by the remaining statistical indicators, which supplement these variables.

3.2.2. Frequency Domain Feature Extraction

The distribution of signal energy across frequencies is not explicitly described by time domain statistics, despite the fact that they represent overall waveform behavior. The discrete Fourier transform (DFT), which is computationally executed using the fast Fourier transform (FFT), is thus used to convert each segmented signal into the frequency domain. The definition of the frequency domain representation is

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, \qquad k = 0, \ldots, N-1 \tag{6}$$

Where the complex spectral coefficient corresponding to the k-th frequency bin is denoted by $X(k)$. The spectral energy distribution and prominent frequency components are described using the resulting magnitude spectrum. Spectral descriptors are derived instead of depending solely on individual spectral peaks to produce compact representations appropriate for further fusion. This enables the statistical features derived directly from the time domain to be supplemented by frequency domain data from vibration and motor current measurements.

3.2.3. Wavelet Based Time Frequency Representation

Because the FFT specifies frequency content without keeping temporal localization, bearing fault signals may display transient and non-stationary characteristics that cannot be sufficiently localized by traditional Fourier analysis. Continuous wavelet transform (CWT) analysis is used to create a joint time frequency representation in order to overcome this restriction. The CWT for a signal x(t) is described as

$$W_x(a,b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} x(t)\,\psi^*\left(\frac{t-b}{a}\right)dt \tag{7}$$

The scale parameter is denoted by a, temporal translation is indicated by b, the mother wavelet is represented by $\psi(\bullet)$, and complex conjugation is indicated by $(\bullet)^*$. While b maintains its temporal localization, variation of an allows the examination of signal patterns at various frequency scales. The evolution of signal energy across time and scale is described in two dimensions by the resulting wavelet coefficients. When recognizing localized transient characteristics linked to bearing faults, this instance is especially helpful. In order to demonstrate the complimentary information offered by the various demonstrates, Figure 4 compares representative time-domain, frequency-domain, and wavelet-based time-frequency characteristics for the three bearing circumstances.

Table 3 summarizes the diagnostic data that was taken from the three domains. Subsequent ablation and explainability analyzes will ascertain the relative contribution of distinct domains and features, rather than assuming that each feature contributes equally to fault categorization.

The synchronized vibration and motor current data are described in complimentary methods by the three standardized characterizations. The multi sensor feature fusion process outlined in Section 3.3 is then used to encode and merge the resulting multi domain data.

3.3. Multi Sensor Feature Fusion

From various physical angles, the vibration and motor current measurements describe bearing behavior. While motor current measurements include electrical responses that may be influenced by variations in the mechanical load and electromechanical connection of the drive system, vibration signals mainly represent the mechanical reaction produced by bearing structure interactions.

However, direct concatenation of the original signals may introduce redundant information and make it difficult for the model to differentiate sensor specific properties since the three sensing channels differ in their physical scale, statistical

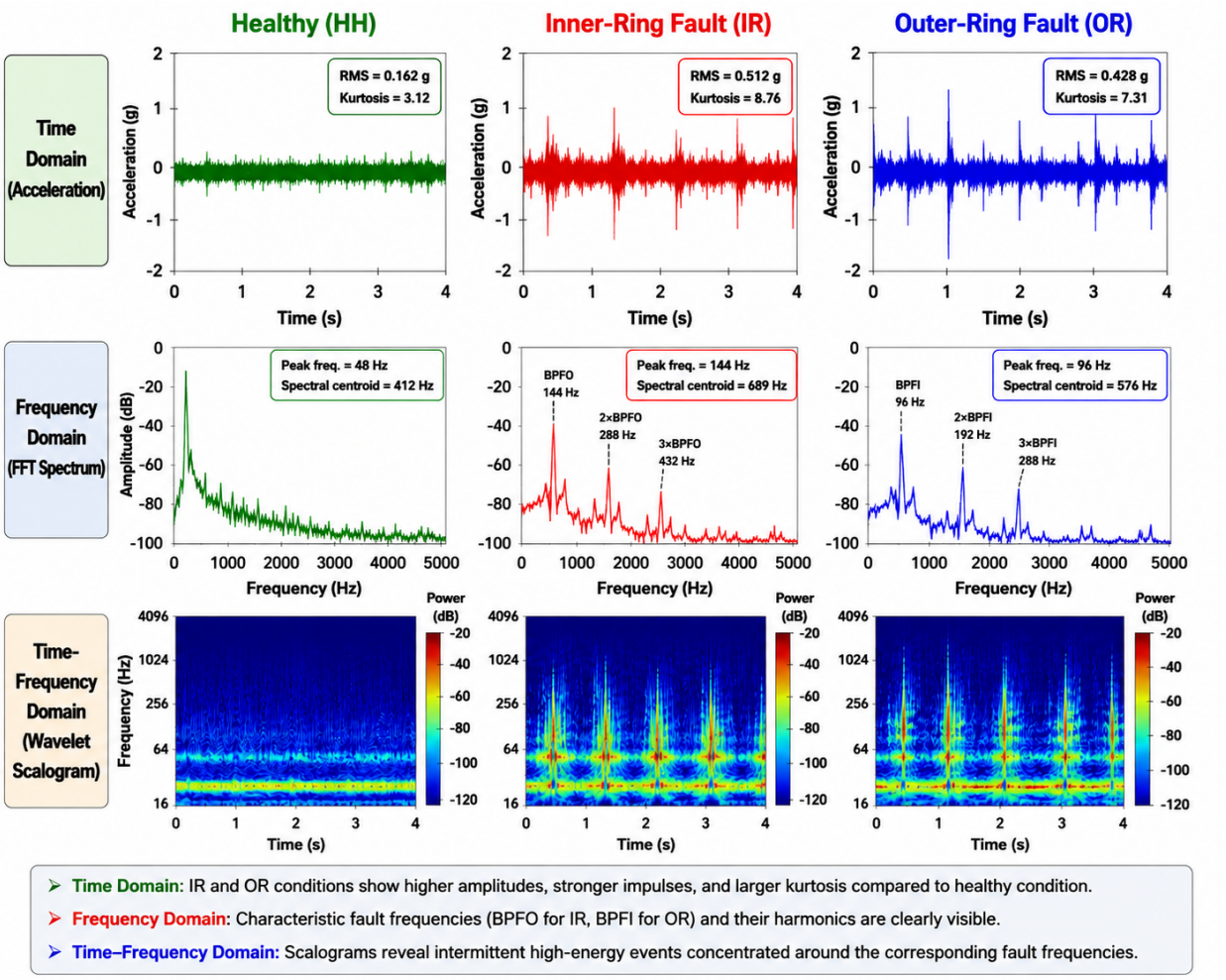


Figure 4: *Multi-domain signal representations of healthy (HH), inner-ring fault (IR), and outer-ring fault (OR) bearing conditions: time-domain waveforms, frequency spectra, and wavelet-based time-frequency characteristics.*

distribution, and sensitivity to bearing errors. In order to preserve sensor specific information before building a shared diagnostic representation, a feature level fusion technique is used.

3.3.1. Sensor Specific Feature Encoding

Using the multi domain representations outlined in Section 3.2, features are independently extracted from the vibration signal and the two phase-current signals for each synchronized segment. Let H_v, H_I, and H_{I2} represent the encoded representations of vibration, phase current 1, and phase current 2, respectively. The s-th sensing channel's representation can be expressed as

$$H_s = \mathcal{E}_s\left(F_s^T, F_s^F, F_s^{TF}\right), \qquad s \in \{v, I_1, I_2\} \tag{8}$$

The time domain, frequency domain, and time-frequency information of sensor s are represented by F_s^T, F_s^F, *and* F_s^{TF}, respectively, and the associated feature encoding process is indicated by $\mathcal{E}_s(\bullet)$. Before cross sensor fusion, each sensing modality can maintain its unique diagnostic features thanks to this formulation.

3.3.2. Cross Sensor Feature Fusion

The single multi sensor feature space is created by combining the three representations after sensor specific encoding. The fundamental expression that has been integrated is written as

Table 3: *Multi-domain signal characteristics employed for bearing fault diagnosis.*

Domain	Representative information	Diagnostic significance
Time	RMS	Overall signal-energy magnitude
Time	Standard deviation	Signal dispersion
Time	Peak-to-peak value	Overall amplitude variation
Time	Skewness	Waveform distribution asymmetry
Time	Kurtosis	Sensitivity to impulsive responses
Time	Crest factor	Relative prominence of signal peaks
Time	Impulse factor	Transient/impact characteristics
Frequency	Magnitude spectrum	Spectral distribution
Frequency	Dominant spectral components	Concentrated frequency information
Frequency	Spectral energy	Distribution of energy in frequency domain
Time-frequency	Wavelet coefficients	Localized transient characteristics
Time-frequency	Wavelet energy distribution	Evolution of signal energy with time and scale

$$H_f = \left[H_v \parallel H_{I_1} \parallel H_{I_2}\right] \tag{9}$$

Where concatenation of features is indicated by ||. The fused representation is then evaluated via the attention-based learning mechanism instead of being given straight to the final classifier since concatenated features may not always contribute equally to fault identification. This lessens reliance on a single sense modality and makes it possible to discover correlations between mechanical and electrical information. The final classification result is not the only indicator of the significance of multi-sensor fusion. Vibration only, current only, and vibration current fusion combinations will be compared in the sensor level ablation studies. This offers quantifiable proof of whether the two electrical channels offer complementing diagnostic data and whether combining them with vibration results in a discernible improvement.

3.4. Transformer Based Fault Classification

Diagnostic data from several sensors and signal domains are included in the fused multi-sensor representation. In order to adaptively highlight informative feature interactions and model dependencies inside this representation, a Transformer encoder is utilized. The Transformer use self-attention to directly construct associations between representation pieces, in contrast to recurrent architectures, which process sequential input iteratively. The Transformer module comes before the last HH, IR, and OR classification stage and comes after multi domain and multi sensor feature analysis.

3.4.1. Feature Projection and Positional Encoding

The fused structure is projected into a common latent dimension d_{model} before to entering the Transformer encoder. The representation's ordering is then maintained by incorporating positional information. For position p and dimension 2i, the sinusoidal positional encoding is described as

$$\mathrm{PE}_{(p,2i)} = \sin\left(\frac{p}{10000^{2i/d_{\text{model}}}}\right), \qquad \mathrm{PE}_{(p,2i+1)} = \cos\left(\frac{p}{10000^{2i/d_{\text{model}}}}\right) \tag{10}$$

The self-attention layers are then fed the encoded illustration.

3.4.2. Multi Head Self Attention

Learnable linear projections are used to obtain query, key, and value matrices for an input representation H. The definition of scaled dot-product attention is

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

Where the dimensionality of the key vectors is represented by d_k. Excessively large dot products that can result in unstable gradients are limited by the scaling factor $\sqrt{d_k}$. Different relationships within the fused representation are learned by using multiple attention heads

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)\, W^O \tag{12}$$

Where W^O is a learnable output projection matrix, h represents the number of attention heads, and each $head_1$ is produced individually using Eq. (11). The Transformer encoder uses a feed-forward network, residual connections, and layer normalization to enhance representation learning and training stability.

3.4.3. Fault Classification and Model Optimization

The classification layer maps the Transformer encoder's output to the three diagnostic classes. The Softmax function is used to determine the probability of class c.

$$P(y = c \mid H) = \frac{\exp(z_c)}{\sum_{j=1}^{C} \exp(z_j)}, \qquad C = 3 \tag{13}$$

Where the classification logit for class c is indicated by z_c. HH, IR, and OR are the three output classes. The categorical cross entropy loss is minimized in order to optimize the model parameters.

$$L_{CE} = -\frac{1}{M} \sum_{i=1}^{M} \sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic}) \tag{14}$$

The ground-truth class indicator is denoted by y_{ic}, the projected probability for class c is indicated by $\hat{y}_{ic}$, and M is the number of samples.

For reproducibility, the Transformer used in the revised experiments consisted of two encoder layers with four self-attention heads per layer. For each synchronized segment, the feature representation of each of the three sensor channels comprised 24 features derived from the time, frequency, and time-frequency domains. Each 24-dimensional sensor representation was projected to an embedding dimension of 48 before entering the Transformer encoder. The feed-forward network within each encoder layer had a hidden dimension of 96, and a dropout rate of 0.10 was applied. Following the encoder, the three sensor representations were aggregated using mean pooling and passed to a fully connected classification layer with three outputs corresponding to HH, IR, and OR. The complete Transformer contained 39,267 trainable parameters, providing a compact implementation for the three-class diagnostic task. Model parameters were optimized using the Adam optimizer with an initial learning rate of 1×10^{-3}, a batch size of 32, and the categorical cross-entropy objective defined in Eq. (14). Training was performed for 300 epochs without early stopping. To assess sensitivity to model initialization and training randomness, the complete training procedure was repeated using five independent random seeds (42, 52, 62, 72, and 82), while maintaining the same leakage-safe training, validation, and testing partitions. All five runs achieved 100.00% test accuracy, 100.00% macro-F1, and 100.00% macro-AUC, corresponding to a mean ± standard deviation of 100.00 ± 0.00% for each metric. The model trained with seed 42 was retained as the reference model for the subsequent detailed analyses.

3.5. Explainable Diagnostic Mechanism

The model ability to perform well in classification does not prove that it has discovered associations that are diagnostically significant. In order to examine how different sensors, signal domains, and extracted features contribute to the anticipated bearing states, an explainability step is included. Because SHapley Additive exPlanations (SHAP) offer a theoretically based method for evaluating the contribution of individual input variables to model predictions, they are used for feature attribution analysis Lundberg and Lee, [14]. The SHAP additive explanation for a prediction f(x) is as follows

$$f(x) = \phi_0 + \sum_{j=1}^{M_f} \phi_j \tag{15}$$

Where ϕ_0 represents the expected model output, M_f is the number of explanatory features, and ϕ_j denotes the contribution of feature j to the prediction. Positive and negative SHAP values indicate whether a feature shifts the prediction toward or away from a particular diagnostic class.

The current study evaluates explainability at three complementary levels: sensor contribution, which assesses the respective roles of vibration and motor current measurements; feature contribution, which identifies significant statistical and spectral characteristics; and domain contribution, which investigates the relative significance of time, frequency, and time frequency information. Instead of being viewed only as simulated learning depictions, the resulting explanations will then be analyzed in conjunction with the measured signal properties.

For the explainability analysis, SHAP GradientExplainer was applied to the trained Transformer using a stratified background set of 60 samples selected exclusively from the training subset, comprising 20 HH, 20 IR, and 20 OR samples.

All 51 samples in the independent test subset (17 HH, 17 IR, and 17 OR) were subsequently explained. The three Transformer output logits corresponding to HH, IR, and OR were treated as separate multiclass outputs, and SHAP values were estimated using 500 sampling evaluations with a fixed random seed of 42.

To obtain global feature importance, the absolute SHAP value of each input feature was averaged across all 51 test samples and across the three class outputs. Sensor-level contributions were then calculated by summing these mean absolute SHAP values over the 24 features associated with each sensing channel (phase current 1, phase current 2, and vibration). Similarly, domain-level contributions were obtained by summing the attributions of the corresponding time-domain, frequency-domain, and time-frequency domain feature groups across all three sensors. Each group contribution was normalized by the sum of the mean absolute SHAP values over all 72 input features and expressed as a percentage. This aggregation procedure ensures that the reported sensor and domain level percentages sum to 100% within their respective grouping.

4. Experimental Design

4.1. Experimental Protocol and Data Partitioning

The tests were established in order to evaluate the suggested diagnostic approach from four angles: robustness to signal degradation, contribution of specific sensing modalities and signal domains, classification performance, and comparative performance versus published models. Vibration and two-phase current measurements were used as the synchronized multi-sensor inputs for the three bearing states that were taken into consideration in the experiments: healthy (HH), inner ring fault (IR), and outer ring fault (OR). The experiments considered three bearing states: healthy (HH), inner-ring fault (IR), and outer-ring fault (OR), using synchronized vibration and two-phase-current measurements. To prevent information leakage between highly correlated overlapping windows, each continuous class-specific signal sequence was first divided chronologically into training (70%), validation (15%), and testing (15%) blocks. Segmentation was subsequently performed independently within each block using a window length of 4,096 samples and a stride of 2,048 samples, corresponding to 50% overlap. Identical segmentation boundaries were applied to all three sensor channels to preserve synchronization. This procedure generated 361 multi-sensor segments: 120 HH segments, 121 IR segments, and 120 OR segments. The resulting training, validation, and test sets contained 259, 51, and 51 segments, respectively. No segmented window crossed a subset boundary.

Standardization parameters were estimated exclusively from the training subset and subsequently applied without modification to the validation and test subsets. The training subset was used for model fitting, the validation subset for model selection and hyperparameter adjustment, and the independent test subset only for final performance evaluation. The same partitions were retained for all compared models to ensure a consistent evaluation.

4.2. Benchmark Models

Three typical benchmark classifiers were used to compare the suggested Transformer-based diagnostic method: multilayer perceptron (MLP), random forest (RF), and support vector machine (SVM). These models were chosen to offer comparisons with traditional fully connected neural-network methods, ensemble learning, and kernel-based classification, respectively. To make sure that variations in diagnostic performance were due to classification techniques rather than changes in the experimental data, all models were assessed using the same processed input data and identical training, validation, and testing partitions.

To simulate nonlinear relationships among the retrieved diagnostic features, the Random Forest used an ensemble of decision trees, whereas the SVM used a radial basis function (RBF) kernel. A neural network benchmark built on completely connected layers was offered by the MLP. The links between the fused multi sensor and multi domain feature representations were modeled by the suggested Transformer using multi-head self-attention. Table 4 provides a summary of the main experimental combinations.

The validation subset was used to select the model parameters, while the independent test subset was only kept for the final performance assessment. In order to ensure a consistent comparison, all benchmark models were constructed using the same feature building and data partitioning procedure.

4.3. Performance Evaluation Metrics

The F1-score, recall, accuracy, and precision are used to measure classification performance. The definitions of precision and recall for a certain class are as follows

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \qquad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \tag{16}$$

The terms TP, FP, and FN stand for true-positive, false-positive, and false-negative forecasts, respectively. Precision and recall are combined via the F1-score using their harmonic mean:

Table 4: *Experimental configuration and benchmark models used for comparative evaluation.*

Category	Model/parameter	Configuration
Data	Diagnostic classes	HH, IR, OR
Data	Sensor channels	Vibration, phase current 1, phase current 2
Segmentation	Window length	4,096 samples
Segmentation	Window stride	2,048 samples
Segmentation	Window overlap	50%
Partition	Train/validation/test	Chronological 70% / 15% / 15% before segmentation
Samples	Generated segments	361 (259 train, 51 validations, 51 test)
Conventional ML	SVM	RBF kernel
Conventional ML	Random Forest	Tree-based ensemble
Neural network	MLP	Fully connected network
Proposed	Transformer	Multi-head self-attention
Optimization	Loss function	Categorical cross-entropy
Evaluation	Metrics	Accuracy, precision, recall, F1-score, AUC

$$F_1 = 2\,\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

The total precision of categorization is determined as

$$\text{Accuracy} = \frac{\sum_{c=1}^{C} TP_c}{N_{\text{test}}} \tag{18}$$

where N_{test} is the total number of test samples and C=3 stands for the three diagnostic classes. Class specific misclassification patterns are examined using confusion matrices in addition to these scalar metrics, and class separation capability is evaluated using ROC curves and the associated area under the curve (AUC).

4.4. Ablation and Noise Sensitivity Test Design

To ascertain whether the suggested multi-sensor and multi-domain arrangement offers quantifiable benefits over reduced input options, ablation experiments are carried out. Vibration only, current only, and vibration current fusion topologies are evaluated at the sensor level. Time domain, frequency domain, time frequency domain, and combination multi domain configurations are assessed at the domain level. Instead of attributing improvements exclusively to greater model complexity, these trials allow the contribution of each information source to be assessed. By adding controlled additive noise to the test signals, the sensitivity to signal degradation was further examined. The noisy signal is expressed as follows for a given signal-to-noise ratio (SNR)

$$x_{\text{noisy}}(n) = x(n) + \eta(n) \tag{19}$$

Where $\eta(n)$ is the additional noise and $x(n)$ is the initial measurement. To find out how diagnostic performance changed when measurement quality declined, several SNR levels were assessed. The sensitivity of the diagnostic system to measurement deterioration might be investigated independently of the clean data experiment by evaluating the trained models using noise-driven test signals while maintaining the model parameters. Because environmental, structural, and electrical disturbances can taint industrial vibration and current data, the robustness experiment is especially crucial for real world condition monitoring applications. Compared to reporting classification accuracy alone, the combined comparative, ablation, robustness, and explainability studies offer a more thorough evaluation.

5. Results and Discussion

5.1. Multi Domain Signal Characteristics

Prior to assessing classification performance, the diagnostic features of the three bearing states were investigated in the time, frequency, and time frequency domains. The typical signal behavior of the inner ring fault (IR), outer ring fault (OR), and healthy (HH) circumstances is depicted in the multi domain representations previously shown in Figure 4. The physical and signal processing foundation for assessing the complementing diagnostic data found in many signal domains is provided by these representations.

The three bearing states exhibit variations in waveform regularity, impulsiveness, fluctuation strength, and signal amplitude distribution in the time domain. While the IR and OR states display more pronounced transient variations linked to localized bearing disturbances, the healthy condition displays a relatively consistent response. These variations

Table 5: *Comparative classification performance on the processed Paderborn bearing subset.*

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
SVM	100.00	100.00	100.00	100.00	100.00
Random Forest	100.00	100.00	100.00	100.00	100.00
MLP	98.04	98.15	98.04	98.04	100.00
Proposed Transformer	100.00	100.00	100.00	100.00	100.00

lend credence to the quantitative characterization of the measured responses using statistical descriptors as root mean square (RMS), standard deviation, kurtosis, skewness, crest factor, impulse factor, and peak-to-peak amplitude. Regarding the bearing responses' spectrum distribution, the frequency domain approximations offer supplementary information. The dynamic interaction between the rolling elements, raceways, shaft, and supporting structure is altered by localized flaws, which leads to modifications in the spectrum energy distribution and concentration. In order to describe these variations beyond traditional time domain statistics, spectral centroid, spectral bandwidth, spectral energy, dominant-frequency components, and associated frequency domain descriptors were included. The evolution of spectral activity over time is further demonstrated by the time frequency representations displayed in Figure 4. This information is especially important for bearing-fault signals that have non-stationary and transient components that are not fully captured by a global frequency spectrum. Therefore, by maintaining the temporal localization of fault related signal components, the wavelet-based representation enhances both the frequency domain information and the statistical time domain properties.

When considered collectively, the multi-domain features shown in Figure 4 show that the chosen HH, IR, and OR conditions contain discriminative information across multiple complementary representations. However, the effectiveness of the diagnosis cannot be determined just by these descriptive and visual discrepancies. In order to assess whether the derived representations allow for accurate categorization among the three bearing circumstances, quantitative classification experiments were conducted.

5.2. Comparative Diagnostic Performance

Using similar training and testing data, the suggested framework's classification performance was assessed against representative mathematical models for machine learning complementary evaluations of diagnostic performance were obtained using accuracy, macro-precision, macro-recall, macro-F1 score, and the area under the receiver operating characteristic curve (AUC). Table 5 summarizes the quantitative outcomes for SVM, Random Forest (RF), multilayer perceptron (MLP), and the suggested Transformer.

Table 5 shows that on the clean independent test subset, SVM, Random Forest, and the proposed Transformer achieved 100% accuracy, macro-precision, macro-recall, macro-F1, and AUC. The MLP achieved 98.04% accuracy, 98.15% macro-precision, 98.04% macro-recall, and 98.04% macro-F1, while its macro-AUC reached 100.00%. These results indicate strong class separability within the selected HH, IR, and OR subset under the revised leakage-safe evaluation protocol. The key performance indicators from Table 5 are presented in Figure 5 to offer a direct visual comparison of the classifiers. The figure makes it possible to analyze variations between the assessed techniques without depending solely on one performance metric.

The suggested Transformer achieves the same clean-data performance as SVM and Random Forest, as seen in Figure 5, although the MLP shows a slight decrease. Therefore, the assertion that the suggested Transformer performs better than all traditional classifiers in clean situations is not supported by the experimental data. Rather, the results show that a highly separable representation of the chosen HH, IR, and OR classes is produced by the recovered multi-domain features. Because an advanced model does not always offer an advantage when the input representation already successfully divides the classes, this observation has scientific significance. As a result, the following analyzes concentrate on class-specific behavior, sensor contribution, feature-domain contribution, susceptibility to noise, and interpretability in addition to overall classification accuracy.

5.3. Cross Operating Condition Generalization

In order to evaluate whether the high classification performance obtained under the original clean-data setting generalizes to changes in operating conditions, an additional cross-operating condition experiment was conducted using independent measurements from the Paderborn bearing dataset. The source operating condition was N15_M07_F10, corresponding to a rotational speed of 1500 rpm, a load torque of 0.7 Nm, and a radial force of 1000 N. Models trained using data from this source condition were subsequently evaluated without retraining under two different target conditions: N09_M07_F10, in which the rotational speed was reduced to 900 rpm while the torque and radial force remained unchanged, and N15_M01_F10, in which the load torque was reduced to 0.1 Nm while the rotational speed and radial force remained unchanged. This experimental design provides a more stringent assessment of model generalization by introducing speed and torque distribution shifts rather than relying solely on within-condition data partitioning.

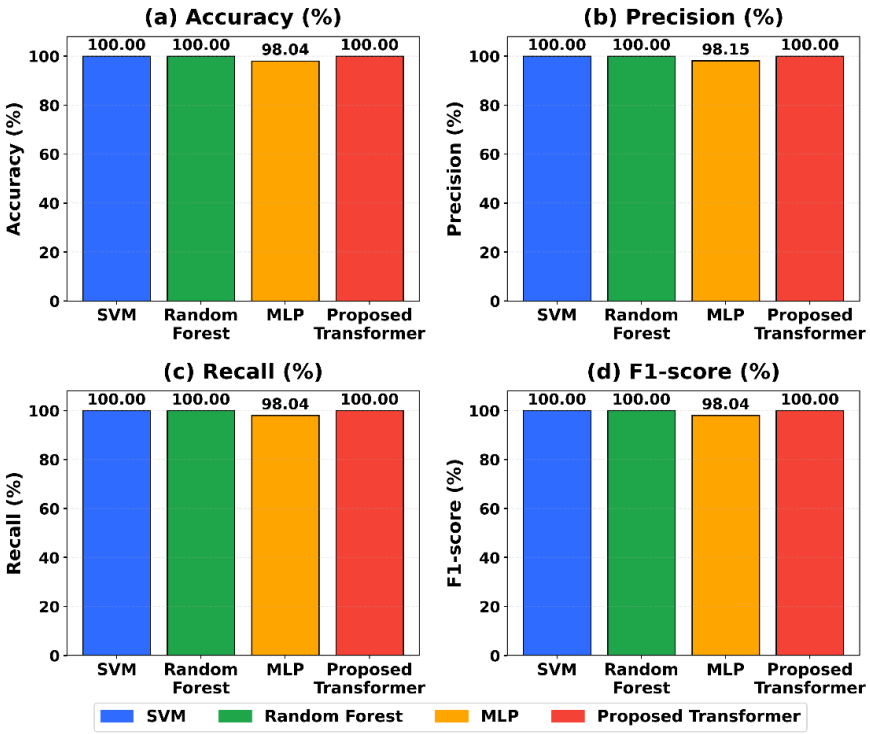


Figure 5: *Comparative diagnostic performance of SVM, Random Forest, MLP, and the proposed Transformer on the processed Paderborn bearing subset.*

Table 6: *Cross-operating-condition classification accuracy on real-damage Paderborn bearings.*

Model	N09_M07_F10 Speed Shift (%)	N15_M01_F10 Torque Shift (%)
SVM	50.00	49.60
Random Forest	37.98	**97.60**
MLP	49.68	49.60
Transformer	50.00	49.60

For this additional evaluation, unambiguously labeled real-damage bearings available across all three operating conditions were considered. The analysis included five outer-ring-damage bearings (KA04, KA15, KA16, KA22, and KA30) and five inner-ring-damage bearings (KI14, KI16, KI17, KI18, and KI21). Bearings containing mixed inner- and outer-ring damage were excluded to avoid assigning ambiguous samples to a single fault category. The same signal-processing and feature-processing procedure was applied consistently across the source and target operating conditions. The resulting cross operating condition classification accuracies are summarized in Table 6.

To provide a direct visual comparison of model robustness under the two operating condition shifts, the cross-condition classification accuracies are presented in Figure 6. The figure highlights the substantial performance degradation under the speed shift condition for all four models and the contrasting robustness of Random Forest under the torque shift condition.

As shown in Table 6, model performance changed substantially when the operating condition differed from that used for model training. Under the speed-shift condition N09_M07_F10, SVM, Random Forest, MLP, and Transformer achieved accuracies of 50.00%, 37.98%, 49.68%, and 50.00%, respectively. This substantial reduction demonstrates that the discriminative patterns learned under the source condition were not generally invariant to the change in rotational speed. Under the torque-shift condition N15_M01_F10, SVM, MLP, and Transformer similarly achieved accuracies close to 50%, whereas Random Forest retained substantially higher performance at 97.60%. These findings provide an important qualification to the near-perfect results obtained from the original within condition clean data evaluation. Specifically, the 100% performance observed for several classifiers should be interpreted as evidence of strong class separability within the selected operating regime rather than as evidence of operating-condition-invariant generalization. The cross-condition results further indicate that changes in rotational speed constitute a particularly challenging distribution shift for the evaluated feature representations, while Random Forest exhibited considerably greater robustness to the investigated torque change. Consequently, improving invariance to operating-condition changes through domain adaptation, condition invariant feature learning, or broader multi-condition training represents an important direction for future development.

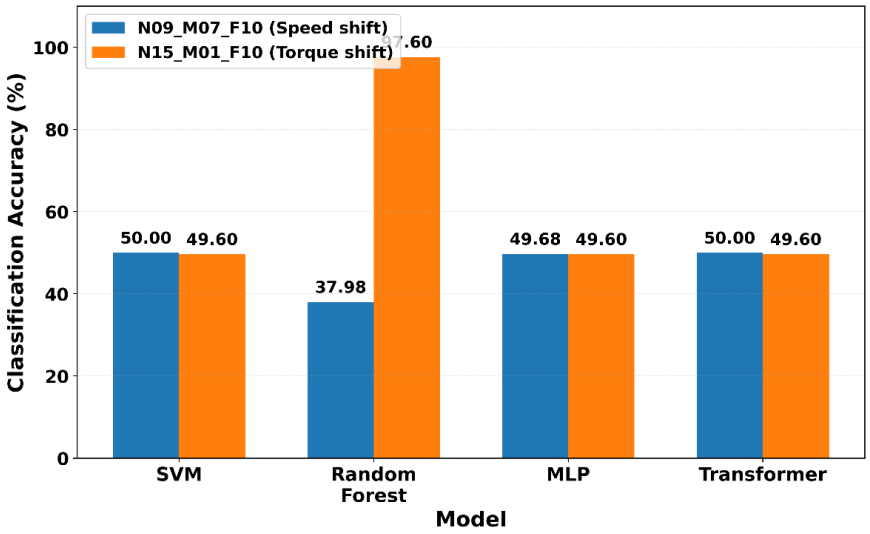


Figure 6: *Cross-operating-condition classification accuracy of SVM, Random Forest, MLP, and Transformer under speed-shift (N09_M07_F10) and torque-shift (N15_M01_F10) conditions.*

5.4. Class Specific Diagnostic Performance

While average precision is a useful indicator of classification performance, it does not show whether prediction mistakes are concentrated in a certain bearing situation. Therefore, a confusion matrix and one-versus-rest ROC analysis were used to assess the proposed framework's class-specific performance. Figure 7 shows the class-level diagnostic performance of the proposed Transformer on the leakage-safe independent test set. The confusion matrix contains 51 test segments, comprising 17 HH, 17 IR, and 17 OR samples. All test segments were correctly classified, with no misclassification observed among the three bearing conditions.

As shown in Figure 7, the proposed Transformer correctly classified all 51 samples in the leakage-safe independent test subset, including 17 HH, 17 IR, and 17 OR segments. The absence of off-diagonal predictions corresponds to 100% accuracy and a macro-F1 score of 100%. These results demonstrate complete discrimination among the three selected bearing conditions within this specific clean test subset. However, the result should be interpreted in the context of the limited three-class subset rather than as evidence of universal generalization across the complete Paderborn dataset.

However, the breadth of the processed dataset should be taken into consideration while interpreting these results. Compared to the HH/IR/OR subgroup used here, the entire Paderborn bearing database has far more experimental variability. The whole database has 32 bearing states, including six undamaged bearings, twelve artificially injured bearings, and fourteen bearings with actual damage produced by accelerated lifetime testing, according to the official dataset description. Additionally, four operating conditions were used for the measurements. Therefore, the strong separability of the processed three-class subset is demonstrated by the perfect classification seen in Figure 6, which should not be taken as a universal generalization across all Paderborn bearings, damage configurations, and operating conditions.

5.5. Sensor and Feature Domain Contribution Analysis

Through ablation studies, the contribution of distinct sensory modalities and signal representation domains was further examined. The purpose of this investigation was to ascertain how closely vibration measures, motor current measurements, or their combined representation were related to the diagnostic performance listed in Table 7. Simultaneously, the contribution of several signal domains was assessed to ascertain whether time domain, frequency domain, time-frequency domain, or integrated multi domain information was the primary cause of the observed class discrimination.

Three input configurations vibration only, motor current only, and combined vibration-current measurements were assessed at the sensor level. The recovered characteristics were compared with the full multi-domain representation at the representation level after being divided into time domain, frequency domain, and time-frequency-domain groups. To be sure that the observed variations were due to the available diagnostic information rather than modifications in the experimental approach, all configurations were assessed using the identical data split and classification protocol. Table 7 summarizes the ablation results at the sensor and domain levels.

Vibration alone obtained 100% classification accuracy, while the current-only arrangement achieved 96.30%, as Table 6 illustrates. The accuracy was 100% when vibration and motor-current data were combined. These findings suggest that, for the chosen bearing conditions, vibration contains especially powerful diagnostic information. Because the vibration sensor captures mechanical responses near the bearing housing and the motor current indirectly transmits

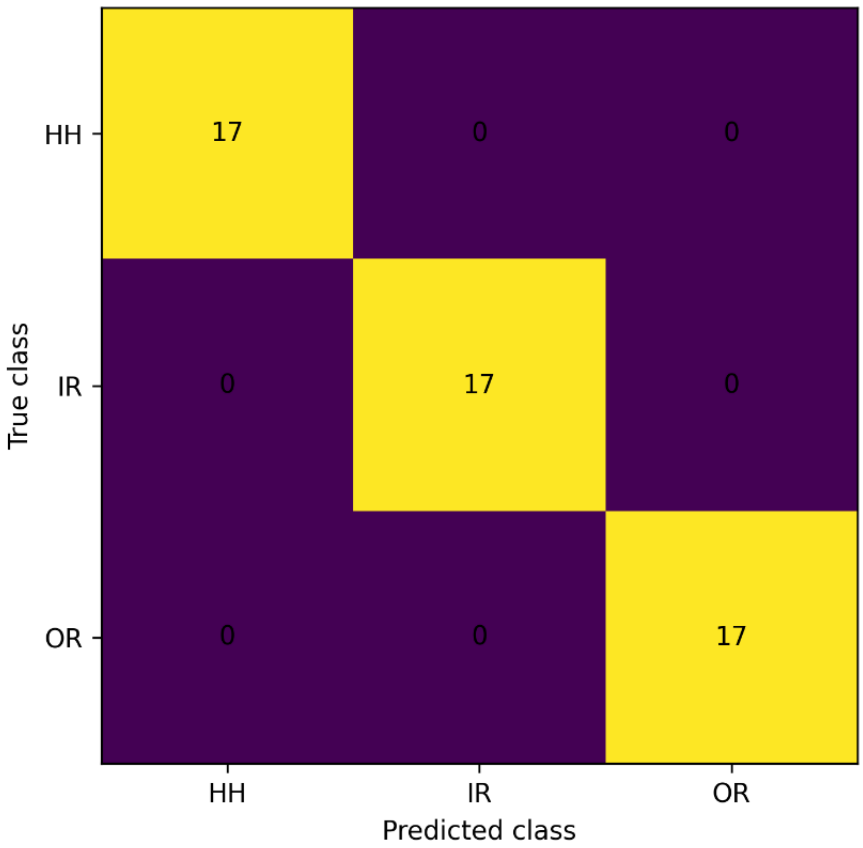


Figure 7: *Confusion matrix of the proposed Transformer on the leakage-safe independent test set for HH, IR, and OR bearing conditions.*

Table 7: *Ablation study of sensor-level and signal-domain contributions.*

Ablation level	Configuration	Accuracy (%)	Macro-F1 (%)
Sensor	Vibration only	100.00	100.00
Sensor	Current only	96.30	~96
Sensor	Vibration + current	100.00	100.00
Feature domain	Time domain only	100.00	100.00
Feature domain	Frequency domain only	100.00	100.00
Feature domain	Time-frequency only	100.00	100.00
Feature domain	Full multi-domain	100.00	100.00

bearing-related information through electromechanical interactions inside the drive system, this behavior makes physical sense. In a similar vein, the initial Paderborn benchmark study showed that although motor-current signals have diagnostic information, they typically offer less fault discrimination than vibration signals.

In contrast, the current-only outcome is still significant. Even in the absence of the vibration channel, an accuracy of 96.30% shows that significant bearing-condition information is present in the electrical data. Therefore, motor-current-based condition monitoring can offer a helpful supplementary sensing method, especially in applications where the drive system already provides electrical measurements. Another significant finding is provided by the domain-level data in Table 6. For the chosen clean subset, time domain, frequency domain, and time-frequency-domain characteristics independently produced flawless classification. As a result, combining all domains did not increase clean-data accuracy above 100%. The assertion that multi domain fusion is required to categorize this specific clean subset is thus not supported by the data. In other words, they show that the chosen bearing conditions have robust discriminatory properties across many signal representations. This encourages testing the system in more difficult signal settings, especially when measurement noise is added.

5.6. Noise Sensitivity Analysis

Diagnostic performance obtained under clean measurement conditions may not accurately reflect model behavior when signal quality deteriorates. Structural vibration, electromagnetic interference, sensor noise, and other disturbances can

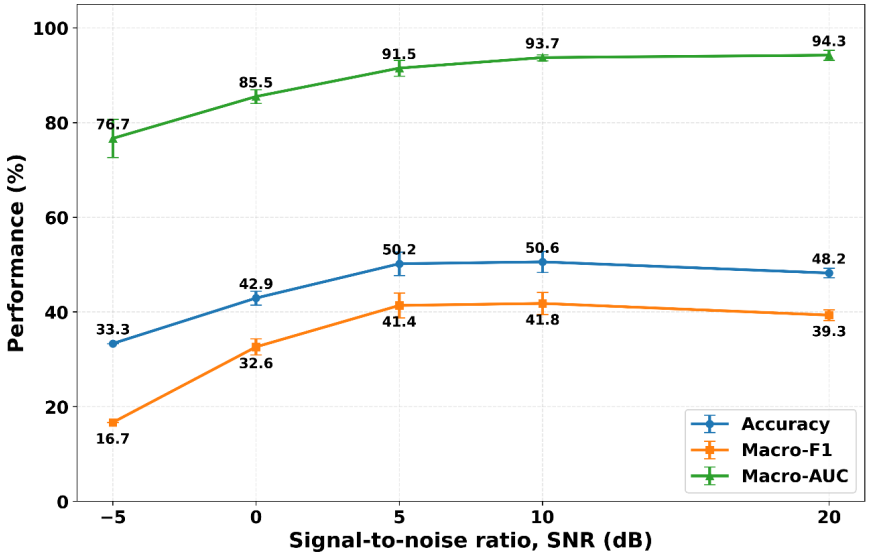


Figure 8: *Noise-sensitivity performance of the proposed Transformer under different signal-to-noise ratios. Points represent mean performance across 10 independent Gaussian-noise realizations, and error bars indicate standard deviations.*

affect vibration and motor-current measurements in real-world condition-monitoring environments [15]-[16]. Therefore, the noise-sensitivity experiment described in Section 4.4 was conducted to evaluate the robustness of the proposed diagnostic framework under progressively degraded input signals.

To evaluate robustness to measurement noise, additive white Gaussian noise (AWGN) was introduced independently into each raw sensor channel of the test samples at SNR levels of -5, 0, 5, 10, and 20 dB. For each channel, the signal power P_s was calculated directly from the corresponding raw signal, and the required noise power was determined as $P_n = \frac{P_s}{10^{\frac{SNR}{10}}}$. Independent Gaussian noise with the resulting variance was then generated and added to the vibration, phase-current-1, and phase-current-2 signals before feature extraction. Importantly, normalization parameters were estimated exclusively from the clean training data and were not recalculated using noisy test samples. Figure 8 presents the resulting mean accuracy, macro-F1, and macro-AUC across the investigated SNR levels, with error bars indicating the standard deviation across the 10 independent noise realizations. The original test labels and sample ordering were preserved throughout the evaluation. To reduce dependence on a single random noise realization, each SNR condition was evaluated using 10 independently generated noise realizations, and the mean and standard deviation were calculated for each performance metric.

As shown in Figure 8, the corrected noise-sensitivity experiment reveals a clear dependence of diagnostic performance on the imposed noise level. At -5 dB, the model collapsed to a single predicted class, resulting in a mean accuracy of 33.33% and a macro-F1 score of 16.67%, while the probability-based macro-AUC was 76.67%. At 0 dB, accuracy and macro-F1 increased to 42.94% and 32.63%, respectively, while macro-AUC increased to 85.51%. At 5 and 10 dB, the corresponding accuracies were 50.20% and 50.59%, with macro-F1 scores of 41.39% and 41.81%, respectively. The macro-AUC values at these two noise levels were 91.52% and 93.75%. At 20 dB, mean accuracy was 48.24%, macro-F1 was 39.34%, and macro-AUC reached 94.25%. The difference between the hard-label metrics and macro-AUC indicates that the model retained considerable class-ranking information even when its final discrete class assignments remained unreliable. In particular, the overall increase in macro-AUC as the SNR improved indicates recovery of class separability with decreasing noise intensity, whereas accuracy and macro-F1 remained substantially below their clean-data values. The small non-monotonic variation in the hard-label metrics at the higher SNR levels further indicates that improved probability ranking does not necessarily translate directly into improved discrete predictions for every noise realization. These results demonstrate that the perfect classification obtained under the clean test condition should not be interpreted as evidence of noise robustness. Together with the cross-operating-condition results in Table 6 and Figure 6, the noise-sensitivity analysis shows that strong performance under controlled clean conditions does not guarantee comparable reliability under measurement disturbances or operating-condition shifts. Consequently, noise-aware training, data augmentation, signal denoising, and condition-invariant feature learning represent important directions for improving the robustness of the proposed diagnostic framework.

5.7. Explainability and Feature Contribution Analysis

The ablation and classification studies quantify the contributions of various input configurations and demonstrate the diagnostic capability of the assessed models, but they do not explicitly pinpoint the specific signal properties that affect

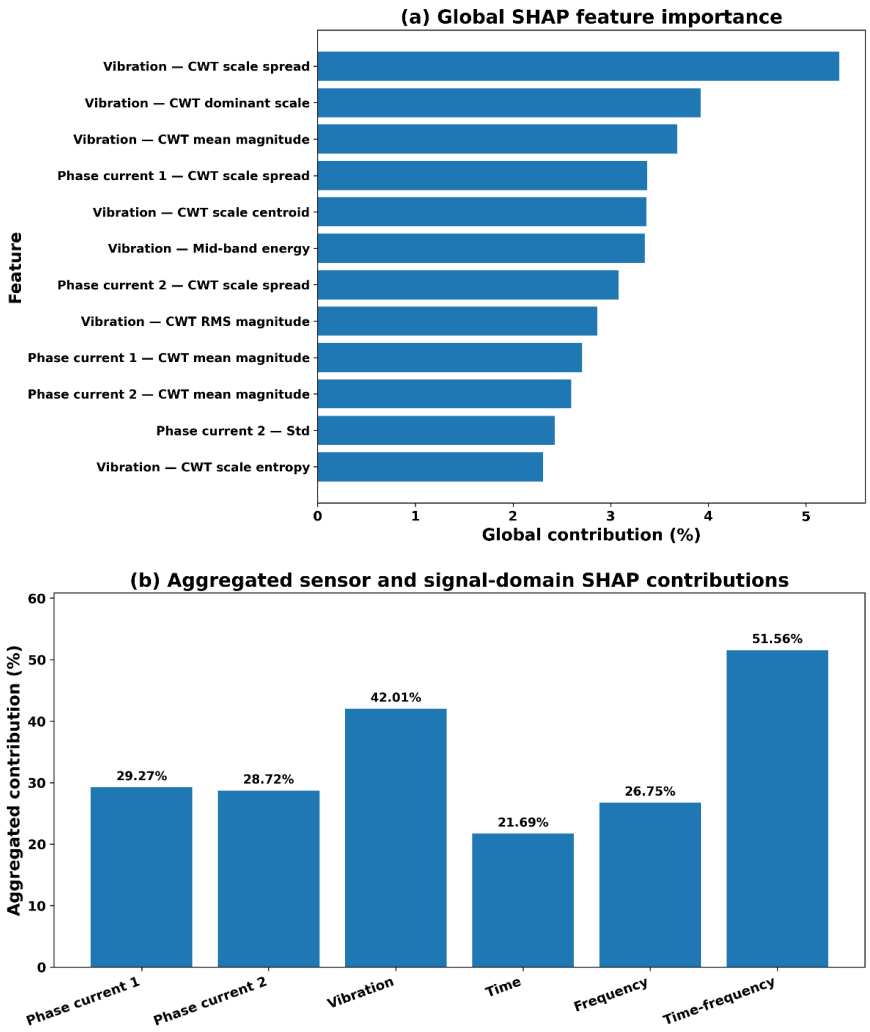


Figure 9: *SHAP-based explainability analysis of the revised Transformer: (a) global feature importance based on mean absolute SHAP values across the independent test set and three class outputs; and (b) normalized sensor-level and signal-domain contributions obtained by aggregating the corresponding feature-level SHAP values.*

the final predictions. In order to determine the contribution of the retrieved features and to connect the data-driven classification method to interpretable properties of the vibration and motor-current measurements, explainability analysis was carried out. Feature contributions were quantified using the SHAP methods described in Section 3.5 and Eq. (15). The resulting attribution values were analyzed at three levels: individual feature contribution, sensing modality contribution, and signal domain contribution, as opposed to solely focusing on the significance of individual variables. Because of this hierarchical interpretation, the explainability analysis may be directly compared to the ablation results presented in Table 7 at the sensor and domain levels. Figure 9 depicts the resultant feature attribution analysis.

The global feature-importance analysis in Figure 9(a) ranks the extracted diagnostic characteristics according to their mean absolute SHAP contributions across the independent test set and the three class outputs. Features with larger mean absolute SHAP values exert a greater overall influence on the model predictions, whereas features with smaller values have comparatively lower global influence. The analysis therefore provides a quantitative basis for examining how statistical, spectral, and time-frequency characteristics contribute to the decisions of the trained Transformer.

At the sensor level, the aggregated SHAP results in Figure 9(b) show that vibration-derived features contributed 42.01% of the total attribution, followed by phase current 1 at 29.27% and phase current 2 at 28.72%. Vibration therefore remained the largest individual sensing modality, while the two motor-current channels together accounted for a substantial proportion of the model attribution. This result indicates that the model did not rely exclusively on vibration information but also retained complementary diagnostic information from the electrical measurements.

At the signal-domain level, time-frequency features produced the largest aggregated contribution at 51.56%, followed by frequency-domain features at 26.75% and time-domain features at 21.69%. This revised attribution pattern indicates

that localized time-frequency information played the dominant role in the final Transformer representation, while spectral and statistical characteristics provided complementary information. The result is consistent with the motivation for incorporating wavelet-based representations, which preserve localized characteristics of non-stationary bearing signals that cannot be represented completely by time or frequency-domain descriptors alone. The SHAP and ablation analyses provide complementary rather than identical measures of feature relevance. SHAP quantifies how the features retained by the trained model contribute to its predictions, whereas ablation evaluates how classification performance changes when an entire information source is removed. Taken together, these analyses show that vibration is the largest individual sensor-level contributor in the assessed subset, while the two current channels also provide substantial complementary information. At the representation level, the revised SHAP analysis identifies time-frequency information as the dominant source of model attribution.

5.8. Discussion and Engineering Implications

Regarding the diagnostic features of the processed Paderborn bearing subset and the respective contributions of sensing modality, signal representation, classification method, and measurement quality, the data shown in Figures 4-8 and Tables 5-6 offer complementary findings. Figure 4 multi-domain study shows that the time, frequency, and time frequency domains of the HH, IR, and OR situations have distinctive features. These variations show that fault-related information is not limited to a single representation of the observed signals and provide the signal-processing foundation for the classification results. The great separability of the proposed bearing conditions under clean measurement settings is further illustrated by the comparative results in Table 5 and Figure 5. While the MLP obtained 96.30% accuracy and a macro-F1 score of 96.28%, SVM, Random Forest, and the suggested Transformer reached 100% classification accuracy and macro-F1 score.

As a result, the suggested Transformer performed exceptionally well in diagnostics; but, on the clean test subset, it was not superior than SVM or Random Forest. This result is significant because it shows that when the extracted feature space already contains strongly separable classes, increasing model complexity does not always result in additional classification benefits. These aggregate performance metrics are complemented by the class-specific analysis in Figure 6. Instead of being dominated by a single, clearly categorized condition, the confusion-matrix and ROC results show that the excellent performance of the suggested transformer was spread across the three assessed bearing states. However, the perfect classification for this processed subset should not be regarded as universal performance across all bearing states in the entire Paderborn database; rather, it ought to remain limited to the experimental settings assessed in this study. Further information on the foundation of the observed diagnostic performance can be found in Table 6 ablation results. The current-only arrangement yielded 96.30% classification accuracy, compared to 100% for vibration-only readings. This discrepancy suggests that for the selected HH, IR, and OR situations, the vibration channel provides especially potent discriminatory information. Because housing vibration offers a reasonably direct assessment of mechanical disturbances produced inside the bearing system, while motor current responses indirectly reflect these disturbances through electromechanical interactions, such behavior makes sense physically.

Conversely, the remarkable performance achieved with motor current alone suggests that significant bearing condition information is retained in the electrical measurements. Because current measurements may already be accessible through motor control and protection systems, this conclusion is pertinent to condition monitoring applications. Without depending solely on extra vibration sensors, motor-current-based diagnosis could offer additional condition information in these situations. The determined subset contains highly discriminative information across several signal representations, as further shown by the feature-domain analysis in Table 7. The three circumstances were strongly separated by time domain, frequency domain, and time-frequency-domain properties. As a result, integrating all three domains does not show an extra clean-data accuracy advantage in the current results. Rather, multi-domain representation facilitates systematic examination of diagnostic data across complementing signal properties and offers a more comprehensive explanation of bearing behavior.

Comparing these clean-data results with the noise-sensitivity behavior shown in Figure 7 reveals a significant difference. High classification accuracy under controlled measurements should not be equated with robustness under disturbed conditions, as demonstrated by the significant performance decrease that occurs after signal disruption. This distinction is especially crucial for industrial machinery since the observed signal distribution can be altered by external vibration, electrical interference, shifting structure transmission pathways, sensor installation conditions, and other disruptions. By relating classification decisions to specific features, sensing modalities, and representation domains, the explainability analysis shown in Figure 8 offers an additional level of insight.

This study, when combined with the ablation results in Table 7, offers a clearer explanation of the distribution of diagnostic data within the suggested structure. Instead of interpreting the classifier as a black box prediction mechanism, such interpretation allows model performance to be taken into account alongside the physical information contained in the measured signals, which is beneficial for engineering applications.

Ultimately, the experimental results show that utilizing vibration and motor current data, the processed Paderborn subset offers good discriminatory information for detecting healthy, inner ring fault, and outer ring fault states. While motor-current signals maintain significant additional condition information, vibration offers very powerful diagnostic

information. The noise-sensitivity results also highlight a significant drawback of models that were only trained under clean measurement conditions. Therefore, in order to build more generalization in real-world condition-monitoring situations, future research should expand the current work toward noise-aware training, cross-bearing validation, variable-operating-condition evaluation, and additional real-damage bearing states.

6. Conclusion and Future Work

This study implemented synchronized vibration and motor-current signals from a processed subset of the Paderborn bearing dataset to construct an explainable multi-sensor and multi-domain framework for bearing failure diagnostics. Comparative machine learning and Transformer-based categorization, sensor and feature-domain ablation, noise-sensitivity assessment, and SHAP-based explainability were linked with time domain, frequency domain, and time-frequency domain representations. The MLP obtained 96.30% accuracy and a 96.28% macro-F1 score under clean measurement conditions, while SVM, Random Forest, and the suggested Transformer reached 100% classification accuracy and macro-F1 score. The equivalent performance of many classifiers suggests that substantial class separability was already supplied by the extracted representations and that the assessed subset did not benefit further from increasing model complexity.

Vibration measures independently reached 100% accuracy, while motor-current signals achieved 96.30%, according to sensor-ablation data, demonstrating that electrical measurements preserve significant complementary diagnostic information. While the time, frequency, and time-frequency-domain representations alone were adequate to differentiate the chosen bearing conditions under clean measurements, their combination offers complementing explanations of bearing behavior instead of an extra improvement in clean-data accuracy. However, the noise-sensitivity investigation showed significant performance degradation under signal perturbation, showing that robustness in noisy industrial contexts should not be inferred from high benchmark accuracy under clean laboratory conditions.

These findings were validated by the explainability analysis. According to the revised SHAP analysis, vibration-derived features accounted for 42.01% of the total sensor-level attribution, while phase current 1 and phase current 2 contributed 29.27% and 28.72%, respectively. At the signal-domain level, time-frequency features provided the largest aggregated contribution at 51.56%, followed by frequency-domain features at 26.75% and time-domain features at 21.69%. These findings indicate that vibration remained the largest individual sensor-level contributor, while the two motor-current channels together provided substantial complementary information and time-frequency representations played the dominant role at the domain level. According to the agreement between SHAP attribution and the ablation results, the examined subset's primary diagnostic source was vibration-based spectrum information, with additional information coming from motor-current and complementary signal representations. This study main limitation is that its conclusions are limited to a processed subset of the Paderborn benchmark; as a result, they cannot be applied to actual industrial settings or unseen physical bearings. Therefore, bearing-wise data separation, extra naturally damaged bearings, changing operating circumstances, noise-aware training, and cross-dataset or cross-domain validation should be used in future work to assess the framework. To ascertain whether the reported diagnostic performance and feature-importance patterns hold true under more practical machinery-monitoring situations, such investigations are required.

Author Declarations

Author Contributions (CRediT): Kaoutar Bizmour: Conceptualization; Data Curation; Formal Analysis; Investigation; Methodology; Project Administration; Resources; Software; Supervision; Validation; Visualization; Gnalen Konate: Writing - Original Draft; and Writing - Review & Editing. All authors have read and approved the submitted version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The bearing-condition monitoring data used in this study are publicly available from the Paderborn University Bearing Data Center. The processed dataset used for the analyses was derived from the publicly accessible benchmark data. Further details regarding data selection and preprocessing are provided in the Materials and Methods section.

Acknowledgments: The authors gratefully acknowledge Samuel Kofi Anamor for his valuable technical guidance, research support, and constructive suggestions during the development and preparation of this study. The authors also extend their sincere appreciation to Amartey Ernest Nii Laryea for his valuable assistance and facilitation in establishing the collaboration that contributed to the realization of this work.

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

Ethics Statement: Not applicable.

Declaration of Generative AI: During the preparation of this manuscript, the authors used QuillBot for language refinement, grammatical correction, and improvement of readability. The authors subsequently reviewed and verified the resulting content and take full responsibility for the accuracy, integrity, and content of the manuscript.

References

[1] S. Singh, C. Q. Howard, and C. H. Hansen, "An extensive review of vibration modelling of rolling element bearings with localised and extended defects," *Journal of Sound and Vibration*, vol. 357, pp. 300–330, 2015, doi: 10.1016/j.jsv.2015.04.037.

[2] Z. Chen, X. Hu, L. Li, C. Ye, D. Wu, S. Liu, Y. Ma, and L. Gu, "Study on the failure mechanism and improvement measures of bearings in the full-ocean-depth ultra-high-pressure seawater pump," *Scientific Reports*, vol. 15, Art. no. 41691, 2025, doi: 10.1038/s41598-025-25518-y.

[3] K. K. Raj, S. Kumar, and R. R. Kumar, "Systematic review of bearing component failure: Strategies for diagnosis and prognosis in rotating machinery," *Arabian Journal for Science and Engineering*, vol. 50, no. 8, pp. 5353–5375, 2025, doi: 10.1007/s13369-024-09866-x.

[4] X. Song, L. Bai, Y. Hu, L. Ma, and H. Cao, "A comprehensive review of rolling bearing life prediction: From fatigue life model to data-driven remaining useful life prognostic," *Lubricants*, vol. 14, no. 8, Art. no. 292, 2026, doi: 10.3390/lubricants14080292.

[5] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," *PHM Society European Conference*, vol. 3, no. 1, 2016, doi: 10.36001/phme.2016.v3i1.1577.

[6] P. Gangsar, A. R. Bajpei, and R. Porwal, "A review on deep learning based condition monitoring and fault diagnosis of rotating machinery," *Noise & Vibration Worldwide*, vol. 53, no. 11, pp. 550–578, 2022, doi: 10.1177/09574565221139638.

[7] J. Cen, Z. Yang, X. Liu, J. Xiong, and H. Chen, "A review of data-driven machinery fault diagnosis using machine learning algorithms," *Journal of Vibration Engineering & Technologies*, vol. 10, no. 7, pp. 2481–2507, 2022, doi: 10.1007/s42417-022-00498-9.

[8] X. Li, S. Xiao, Q. Li, L. Zhu, T. Wang, and F. Chu, "The bearing multi-sensor fault diagnosis method based on a multi-branch parallel perception network and feature fusion strategy," *Reliability Engineering & System Safety*, vol. 261, Art. no. 111122, 2025, doi: 10.1016/j.ress.2025.111122.

[9] L. Magadán, C. Ruiz-Cárcel, J. C. Granda, F. J. Suárez, and A. Starr, "Explainable and interpretable bearing fault classification and diagnosis under limited data," *Advanced Engineering Informatics*, vol. 62, Art. no. 102909, 2024, doi: 10.1016/j.aei.2024.102909.

[10] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. V. Duarte, "An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery," *Mechanical Systems and Signal Processing*, vol. 163, Art. no. 108105, 2022, doi: 10.1016/j.ymssp.2021.108105.

[11] L. Lei, W. Li, S. Zhang, C. Wu, and H. Yu, "Research progress on data-driven industrial fault diagnosis methods," *Sensors*, vol. 25, no. 9, Art. no. 2952, 2025, doi: 10.3390/s25092952.

[12] W. Ke, S. Wang, L. Zhou, R. Wang, and Y. Dai, "Robot bone cutting tool status recognition method based on multimodal fusion network," *Biomedical Signal Processing and Control*, vol. 121, Art. no. 110197, 2026, doi: 10.1016/j.bspc.2026.110197.

[13] S. Ullah, M. Umar, and J.-M. Kim, "Physics guided fused image learning with enhanced squeeze excitation for failure analysis of multistage centrifugal pumps," *Scientific Reports*, vol. 16, Art. no. 16179, 2026, doi: 10.1038/s41598-026-47809-8.

[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[15] R. R. Kumar, M. Andriollo, G. Cirrincione, M. Cirrincione, and A. Tortella, "A comprehensive review of conventional and intelligence-based approaches for the fault diagnosis and condition monitoring of induction motors," *Energies*, vol. 15, no. 23, Art. no. 8938, 2022, doi: 10.3390/en15238938.

[16] I. U. Hassan, K. Panduru, and J. Walsh, "An in-depth study of vibration sensors for condition monitoring," *Sensors*, vol. 24, no. 3, Art. no. 740, 2024, doi: 10.3390/s24030740.