ITC 2/55

Information Technology and Control
Vol. 55 / No. 2/ 2026
pp. 953-974
DOI 10.5755/j01.itc.55.2.44748

Noise-Robust Multi-Domain Feature Integration for Bearing Fault Diagnosis

Received 2025/12/13 | Accepted after revision 2026/06/26

HOW TO CITE: Tastimur, C. (2026) Noise-Robust Multi-Domain Feature Integration for Bearing Fault Diagnosis. *Information Technology and Control*, 55(2), 953-974. https://doi.org/10.5755/j01.itc.55.2.43823

Noise-Robust Multi-Domain Feature Integration for Bearing Fault Diagnosis

Canan Tastimur

Department of Computer Engineering; Faculty of Engineering-Architecture; Erzincan Binali Yildirim University, Erzincan, 24100, Turkey; e-mail: ctastimur@erzincan.edu.tr

Corresponding author: ctastimur@erzincan.edu.tr

In this study, a hybrid deep learning architecture is proposed for robust vibration-based fault diagnosis in industrial machinery by jointly modeling time-domain, frequency-domain, and temporal dynamics. In the time domain, convolutional layers combined with an Enhanced Gated Attention (EGA) mechanism emphasize informative signal components while suppressing noise. Temporal evolution is modeled using Neural Ordinary Differential Equations (Neural ODEs), enabling smooth and stable continuous-time feature representations. In parallel, a Fourier Neural Operator (FNO) extracts frequency-domain characteristics, augmented with gated attention to focus on fault-related spectral patterns. Long Short-Term Memory (LSTM) layers capture long-range dependencies, while Squeeze-and-Excitation (SE) blocks adaptively recalibrate channel-wise feature responses. A multi-scale attention-based fusion module integrates domain-specific representations and auxiliary features to enhance discrimination under varying operating conditions. The proposed model is evaluated on the SUBFv1.0 dataset through extensive ablation studies and experiments under multiple noise levels, achieving 98.41% accuracy in noise-free conditions and maintaining performance above 91% even at 5 dB SNR. Unlike existing multi-path approaches that combine heterogeneous features in a loosely coupled or discrete manner, the proposed architecture uniquely integrates multi-domain feature learning with bidirectional attention mechanisms and continuous-time temporal dynamics, enabling coherent cross-domain interaction and robust fault characterization.

KEYWORDS: Bearing fault diagnosis; Feature fusion; Hybrid deep learning; Neural ODE; Self-attention.

1. Introduction

Bearings are essential mechanical elements commonly employed in rotating machinery to facilitate the seamless transmission of rotational motion and minimize friction. In industrial systems, bearing failures can cause sudden stoppage, significant maintenance costs, and safety hazards. As a result, early and precise detection of bearing failures is critical for the industry's predictive maintenance systems to function effectively [38], [29]. However, traditional defect diagnostic approaches rely heavily on feature extraction and statistical analysis based on engineering knowledge. These approaches fail to capture the dynamic nature of the signal in complex and noisy situations, and they have poor generalization performance [30].

One of the most difficult issues in diagnosing bearing failures is that they typically emit unclear and low-amplitude signals in their early phases. These signals are easily obscured by vibrations and external noises that occur during normal machine operation. This circumstance limits the utility of conventional signal processing methods, particularly in detecting low-frequency or localized damage [45]. Furthermore, because different types of bearing defects (such as inner race, outer race, ball, or cage faults) arise at different rates and patterns, accurately identifying this variety requires a sophisticated analysis procedure. Although time-frequency analyses and various filtering algorithms are used in this area [31], they often need fault diagnosis skills and are dependent on the quality of manually constructed features. Furthermore, manual decisions must be made about which window to assess the signal in and which features to extract. This decreases the system's flexibility and reusability in varied situations. Given the varied loads, speeds, and noise sources encountered in field circumstances, the demand for autonomous and learning systems is growing [35].

In recent years, deep learning approaches have achieved amazing success, particularly in signal processing and defect diagnostics. Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Self-Attention, and other architectures can outperform previous approaches by automatically learning complicated patterns in vibration signals [21], [62]. Deep learning models provide more powerful and generalizable representation capabilities by learning from huge datasets without requiring manual feature engineering. This allows for more accurate, rapid, and reliable diagnostics of bearing issues [58], [42].

In literature, time-frequency analysis, statistical feature extraction, support vector machines (SVM), random forests (RF), and conventional deep learning-based approaches (CNN, LSTM, etc.) have largely been selected for bearing problem diagnosis. Although these studies have had some success, they have limitations such as working with a limited number of features, failing to adequately reflect the entire dynamic structure of the signal, having difficulty distinguishing similarities between different fault types, and being sensitive to different operating conditions (load, speed, and environmental noise). Furthermore, numerous methods in the literature rely solely on one-way information flow, with insufficient inclusion of mechanisms for feeding back and re-evaluating retrieved features.

This study proposes a hybrid model that simultaneously exploits time-domain and frequency-domain features to achieve high-accuracy vibration-based fault diagnosis. The proposed hybrid model integrates the strengths of Neural ODE for dynamic system modeling, the Fourier Operator for frequency-aware feature extraction, and Cross-Domain Attention for efficient feature fusion. This design enables the model to effectively capture both short-term transient faults and long-term degradation patterns, resulting in superior diagnostic performance compared to conventional approaches. The time-domain path begins with convolutional layers extracting fundamental features, followed by an Enhanced Gated Attention (EGA) mechanism. Unlike conventional attention methods, Gated Attention performs simultaneous gating and weighting operations, selectively highlighting critical signal components while suppressing noise. The model can successfully capture complicated and dynamic patterns present in vibration signals because of its adaptive focus. Furthermore, the time-domain path incorporates Neural Ordinary Differential Equations (Neural ODEs) to precisely characterize temporal dynamics. To better capture realistic temporal transitions and fault growth within vibration signals, neural ODEs offer a continuous-depth framework that gradually learns smooth and stable feature representations. A Fourier Neural Operator (FNO)-based design examines the frequency components of the signal in the frequency-domain path. Additionally, this method uses Gated Attention to highlight important frequency bands associated with fault characteristics. Furthermore, Squeeze-and-Excitation (SE) blocks adjust feature maps to adaptively prioritize significant features, while Long Short-Term Memory (LSTM) layers capture long-term dependencies in the time series. Ultimately, a multi-scale attention fusion mechanism fuses the outputs from all three channels, enabling the model to incorporate complimentary data from various scales and domains. The robustness and accuracy of vibration fault classification are greatly increased by this all-inclusive and flexible feature learning approach. In contrast to existing multi-path architectures that combine domain-specific features in a sequential or loosely coupled manner, the proposed framework introduces a unified design that jointly optimizes multi-domain representations through bidirectional attention and continuous-time temporal modeling, explicitly capturing cross-domain dependencies and smooth fault evolution.

The main contributions of this study are:

- Multi-domain feature integration enables a more holistic and discriminative representation of vibration signals, allowing the system to capture complementary fault-related information that single-domain methods typically miss.
- The proposed attention-driven feature enhancement framework improves model focus and resilience, ensuring that critical fault characteristics are emphasized while irrelevant variations and noise are effectively suppressed.
- Continuous-time temporal modeling provides smoother and more interpretable feature evolution, offering greater insight into fault progression compared to traditional discrete-time approaches.
- The multi-scale cross-domain fusion strategy strengthens the model's ability to distinguish subtle and complex defect patterns, improving generalization across varying operating conditions and machinery states.
- Extensive evaluations under severe noise conditions demonstrate the model's high robustness and practical deployability, confirming its suitability for real-world industrial monitoring environments where signal contamination is unavoidable.

2. Related Works

Identifying rotational faults in bearing machines through the analysis of vibration signals is a common issue discussed in literature as shown in Table 1. In recent years, a wide range of methods have been proposed to address this problem, spanning from traditional signal processing and machine learning techniques to modern deep learning-based approaches. Moreover, studies in related engineering domains have shown that hybrid and multi-strategy learning frameworks are particularly effective in handling complex, noisy, and dynamic systems. Motivated by these developments, existing research on vibration-based fault diagnosis can be broadly categorized into two main groups: conventional methods and deep learning-based approaches.

2.1. Conventional Methods for Vibration Fault Diagnosis

Traditional methods for identifying machine faults primarily depend on features extracted manually from frequency, time, or time-frequency domains. To identify faults, these features are input into standard classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees, or Random Forests. In particular, the requirement for expert knowledge in feature engineering restricts their ability to adapt to various operational settings. Recent research has proposed various improvements to address the limitations of conventional methods.

Time-frequency analysis, statistical feature extraction, and signal processing methods are frequently the core components of conventional vibration-based approaches for diagnosing bearing faults. In addition, cepstrum-based signal processing techniques have been widely studied as effective tools for extracting periodic fault-related information from vibration signals, particularly in the context of structural and modal analysis applications [39]. Smith et al. [43] employed Wavelet Transform for the detection of localized bearing faults and showed that this approach is effective even in the presence of noise. Wang et al. [46] proposed a deep separable convolutional network for remaining useful life prediction of machinery components, demonstrating the effectiveness of deep learning-based feature extraction methods in handling complex and non-linear degradation signals in rotating machinery.

Antoni [3] introduced the concept of spectral kurtosis as an effective tool for characterising non-stationary vibration signals and enhancing the detection of early-stage bearing faults under noisy conditions. To address mechanical big data challenges, Lei et al. [28] employed an unsupervised feature learning approach for intelligent fault diagnosis of rotating machinery, enabling automatic feature extraction without manual signal processing. Shao et al. [41] proposed an optimized deep belief network (DBN) for rolling bearing fault diagnosis, demonstrating that deep learning-based feature representation can achieve high classification accuracy even without complex handcrafted feature extraction. Janssens et al. [22] proposed a thermal image-based fault diagnosis framework for rotating machinery and highlighted the effectiveness of machine learning-based feature analysis in improving fault classification performance. Ayankoso et al. [1] proposed a multisensory data-based fault diagnosis approach for induction motors using 1D and 2D convolutional neural networks, demonstrating robust classification performance under complex and uncertain operating conditions.

2.2. Deep Learning Methods for Vibration Fault Diagnosis

The ability of deep learning algorithms to automatically generate hierarchical feature representations from raw input data has made them prominent in the identification of machine defects. While LSTM networks are helpful for modeling temporal dependencies in vibration signals, CNNs are frequently used to capture local patterns and frequency domain features. Additionally, to capitalize on the features of different architectures, hybrid models that combine CNNs with LSTMs or attention modules have been proposed.

Deep learning algorithms are becoming increasingly used in bearing failure detection due to their automatic feature extraction and high accuracy. Wang et al. [50] employed multi sensor fusion model to detect problems directly from raw vibration data, stressing that this method avoids the requirement for manual feature extraction. Zhao et al. [60] proposed a few-shot cross-domain fault diagnosis framework based on model-agnostic meta-learning combined with a genetic algorithm (MAML-GA), enabling robust bearing fault identification under limited labeled data and varying operating conditions. Che et al. [6] created an excellent dimensionality reduction and denoising method using transfer learning, demonstrating that it works even with low SNR inputs. Cui et al. [11] detected bearing faults using the Attention-based CNN model, concluding that this method enhances accuracy by focusing on critical frequency bands. Janssens et al. [23] proposed a Transfer Learning-based model that achieves good generalization performance despite varied operating conditions. Lei et al. [27] used Capsule Networks to detect bearing problems, claiming that their method outperformed standard CNNs. Haidong et al. [15] discovered bearing faults in real time using a Deep Reinforcement Learning-based methodology and demonstrated that this system is stable even in dynamic conditions. Zhao et al. [61] used Graph Neural Networks to detect bearing problems, demonstrating the method's effectiveness in modeling complex interactions. Li et al. [32] suggested a Federated Learning-based approach, claiming that it gives excellent diagnostic accuracy while maintaining data privacy.

A multi-parallel graph neural convolutional network architecture was created in [59] that parallelly processes data in the frequency and temporal domains. Both domains underwent independent analysis, and a late fusion classification was carried out. However, due to the complexity of the structure, the frequency domain offers little information and demands a lot of computing power. An et al. [2] pointed out that temporal relationships are learned via LSTM networks and significant moments in the attention mechanism, which operates on time series data. However, the model may perform worse in noisy environments due to the attention mechanism's increased computational effort. The FNO was employed in this frequency domain-based model created in [33]. The model's drawback is that it ignores time information and is only reliant on frequency information. In [47], which works with noisy inputs, the joint learning attention mechanism was combined with the CNN architecture to highlight key feature components.

In [57], the model was extracted spatial features with CNN, temporal relationships with LSTM, and did adaptive channel weighting with SE blocks. However, LSTM may be inefficient in extended sequences. Huang et al. [18] employed a multi-scale CNN structure and the channel attention module to extract characteristics of various dimensions. The special dataset yielded an accuracy of 99.26%. However, parameter modifications were complicated, and the model was highly sensitive. In [53], time and frequency data were processed in distinct CNN-BiLSTM modules and categorized using multi-scale cross attention. 99.25% accuracy was attained with the CWRU dataset. It is unsuitable for real-time applications. Chen et al. [7] proposed a 3DCNN-LSTM model with an attention mechanism for fault diagnosis, where 3DCNN extracts spatial features, LSTM models temporal dependencies, and attention highlights important features. The method achieved 98.05% accuracy, but its computational complexity is relatively high due to the combined architecture. The computational burden is slightly high. Yang et al. [19] extracted features using CNN, highlighted important information using the attention mechanism, and visualized the model output. 95% accuracy was achieved with the CWRUdataset.

3. Methodology

3.1. Overview of The Proposed Methodology

This paper introduces a hybrid multi-domain neural architecture that unifies time-domain, frequency-domain, and temporal feature extraction pathways for vibration-based defect diagnosis. The impetus for this approach arises from the intrinsic complexity of vibration signals, which frequently contain both transient and stationary components, necessitating simultaneous analysis across many domains to adequately capture discriminative fault characteristics.

The suggested model contains three parallel paths:

– **Time Domain approach (Path-1):** Fundamental temporal features are extracted from raw vibration data using convolutional layers, and the EGA

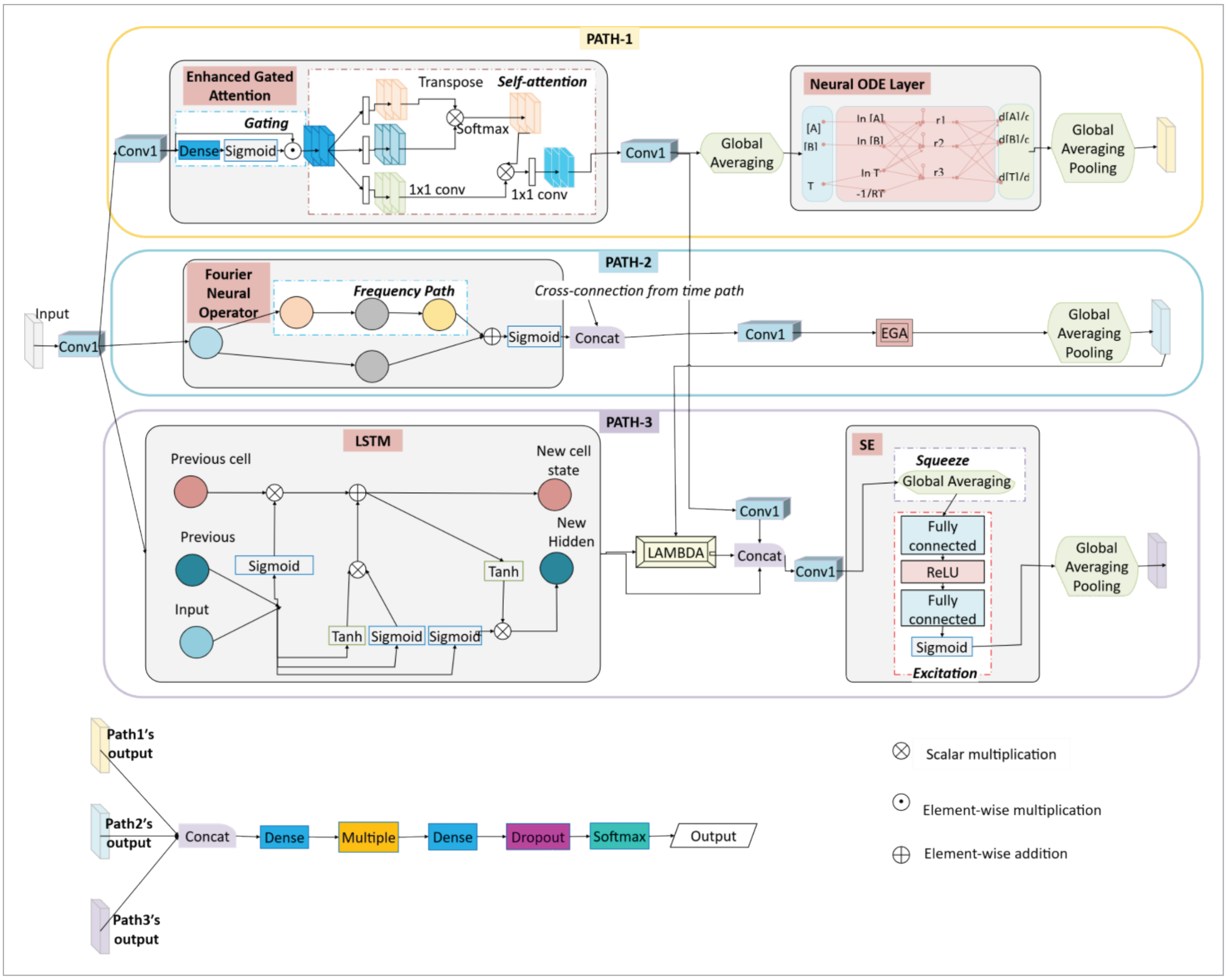
Table 1

Literature overview of vibration-based defect classification in rotating bearing machinery.

Refs.	Method	Description	Dataset	Metric	Disadvantages
[59]	Multi Parallel Graph Neural Network	A structure that processes time and frequency domains with parallel CNN paths and performs late merging.	CWRU	Acc: 0.9813	Frequency information is limited, deep structure calculation is intensive.
[2]	Attention-LSTM	LSTM based model highlighting important moments with periodic sparse attention mechanism in time domain.	CWRU	Acc: 0.9730	Combined failures and the need for numerous examples during training.
[33]	FNO-based Network	Model using Fourier Neural Operator in frequency domain.	CWRU	Acc: 0.9750, F1: 0.9750	Requires big data, time information is neglected.
[47]	CNN + Joint Learning Attention	A gated attention-enhanced CNN model that selectively highlights important signal components with joint learning	CWRU	Acc: 0.9324	Gated structure is complex, real-time implementation is difficult.
[57]	SE-LSTM-CNN	Feature extraction with CNN, temporal relationship learning with LSTM, and adaptive channel weighting with SE blocks.	SDJU	Acc: 0.8304	LSTM part may be inefficient for long sequences.
[18]	Multi-Scale CNN + Channel Attention	Multi-scale feature extraction and channel attention mechanism highlighting important channels.	Special dataset	Acc: 0.9926	Requires detailed parameter adjustments.
[53]	Fused CNN-BiLSTM with multi scale attention	The structure processes and combines the frequency information obtained from the raw signal and multi-scale attention with separate CNN-BiLSTM modules.	CWRU	Acc: 0.9925	Not optimized for real-time application.
[7]	CNN + Attention + Visualization	Feature extraction with CNN, combined structure with attention mechanism and visualization.	CWRU	Acc: 0.9500	The visualization module adds additional computational load.

Figure 1

The overview of recommended approach.



mechanism highlights important information while suppressing noise. Neural ODE component models continuous-time feature evolution for a more stable and smooth representation.

– **Frequency Domain Path (Path-2):** The frequency components of the vibration signal are analyzed with an FNO structure to better capture fault-related frequency bands. This path also uses a gated attention mechanism to emphasize significant frequency regions and suppress irrelevant ones.

– **Temporal Connection and Feature Recalibration (Path-3):** LSTM layers model long-term temporal dependencies in the data. SE blocks adaptively scale feature maps and highlight important channels for improved representation.

The EGA modules in PATH-1 and PATH-2 share the same structural design but operate on different representation spaces, enhancing feature diversity. In the final stage, outputs from all three paths are fused using a multi-scale attention fusion module, which adaptively integrates complementary information from different domains and scales. This strategy improves the model's discriminative ability, enabling robust and accurate fault classification across various operating conditions. The overall architecture and interactions between the parallel paths and fusion module are illustrated in Figure 1. To ensure the reproducibility and transparency of the proposed framework, Table 2 presents a comprehensive architectural description of the model, including each layer and module with their corresponding input/output dimensions, filter and unit configurations, kernel sizes, activation functions, normalization layers, dropout settings, and parameter counts.

3.2. Time-Domain Feature Extraction and Time Modeling with Neural ODE

The simplest region of investigation is the time domain, where vibration signals can be immediately observed to exhibit abrupt shifts, transitory behaviors akin to shocks, and short-term fault effects. Consequently, the time domain features are crucial for identifying the initial indications of malfunctions, evaluating transient irregularities, and differentiating important signal components even in noisy settings. In this study, low-level time information is abstracted and transformed into meaningful representations by processing the fundamental features extracted from the time domain signal through a series of CNN layers.

However, the time-dependent evolution of the signal might not be sufficiently represented by static time features exclusively retrieved using convolutional layers. A time-spreading change is observed in vibration signals. For these reasons, the Neural ODE approach overcomes this limitation by learning the continuous-time change of latent features through differential equations. Thanks to its continuous-depth structure, failure progression can be represented more smoothly and with greater resistance to noise. It can represent long-term deterioration trends more consistently. Therefore, Neural ODE offers a theoretically consistent and physically meaningful modeling framework for vibration-based fault diagnosis.

However, the time-dependent evolution of the signal might not be sufficiently represented by static time features that are exclusively retrieved using convolutional layers. By concurrently carrying out gating and attention weighting operations, the EGA module, in contrast to traditional attention methods, concentrates on information-rich portions of the signal. Neural ODE is used into the proposed system to enable more intricate time-domain modeling. Neural ODE is a dynamic structure that mimics the evolution of characteristics in a continuous time stream, providing an alternative to the discrete nature of conventional artificial neural networks. This approach makes it possible to comprehend the structural features of signals that change over time in a more practical and trustworthy manner.

By integrating Neural ODE into the time path, feature transitions across time can be modeled more smoothly and continuously. To ensure stable training of the combined EGA and Neural ODE framework, the Neural ODE module employs a fixed-step Runge–Kutta solver with an adaptive step size bounded within predefined limits. The ODE function is parameterized using a lightweight multilayer perceptron with smooth activation functions to avoid gradient instability. Gradient clipping and weight decay are applied during training to improve numerical stability. The key hyperparameters of the EGA mechanism and Neural ODE module are summarized in the supplementary materials. The key hyperparameters and implementation details are provided in Appendix A. There is less chance of problems like exploding or gradient fading that might occur with traditional temporal models like LSTM. Starting from the beginning, the model can more precisely track how defects have changed over time and use this evolutionary process to guide classification decisions. Therefore, when CNN-based basic feature extraction in the time domain route, adaptive attention with EGA, and continuous time modeling with Neural ODE are combined, the model is more sensitive and stable against transient, noisy, and dynamic fault patterns.

Vibration signals have a structure that varies with time. Neural ODE interprets the signal as a differential equation and learns it as a continuous function of time. Because vibration faults arise gradually, Neural ODE allows for the modeling of this progressive shift, which aids in understanding the fault generation process. The EGA's gated structure has an advantage over traditional attention in that it evaluates and filters signal components. With more robust and accurate feature extraction, EGA increases the model's reliability by highlighting the crucial fault-related components in the signals while suppressing the noisy and irrelevant regions.

The Neural ODE expresses the function that best describes the data by calculating its derivative or by approximating the function using an analytical technique. The Neural ODE technique uses an ordinary differential equation to calculate the function's estimated value. The function's derivative is parameterized using a neural network. A residual network [16], which all have the identical hidden units, generates output by conducting many transformations on a hidden layer [8].

$$h_{t+1} = h_t + f(h_t, \theta_t). \tag{1}$$

In Equation (1), h_t is the layer output, f is the neural layer, and t is the layer depth. In addition to the ability to build an endless number of layers, a neural network ODE is utilized to derive the continuous dynamics of the hidden units [37].

$$\frac{dh(t)}{dt} = f(h(t), t, \theta_t). \tag{2}$$

In (2), $h(t)$ is the hidden state value for t. $f(h(t),t,\theta_t)$ is a neural network layer parameterized by θ_t at layer t. We obtain the hidden layer output value at depth T by solving the integral in Equation (3).

$$h(t) = h(t_0) + \int_{t_0}^{T} f(h(t), t, \theta_t)dt. \tag{3}$$

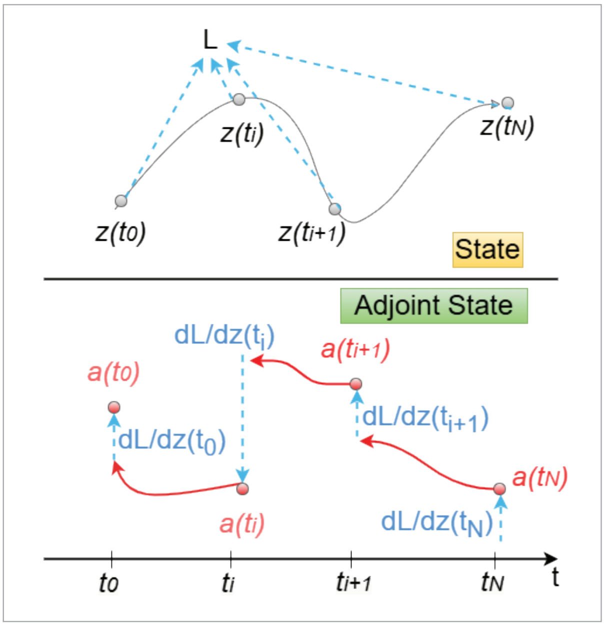
Neural ODEs take a novel way to backpropagating the loss, using continuous hidden state dynamics. In a Neural ODE, the parameters contain not just θ_t, but also t_0 and t_1. Any differentiable loss function that we select can be defined using these parameters in Equation (4).

$$L(t_0, t_1, \theta_t) = L\big(ODESolve(h(t_0), t_0, t_1, \theta, f)\big). \tag{4}$$

It is necessary to transition the hidden states in the opposite direction to compute the loss regarding time. Chen et al. [8] suggested calculating this gradient using the adjoint method [37]. The adjoint approach is the gradient computation method for neural ODEs. The gradient of the loss function L with respect to the parameters and time values is determined in neural ODEs using the adjoint approach. This technique uses a backpropagation that works well for dynamics in continuous time. The temporal change of the hidden state is the $\frac{dz}{dt} = f(z(t),t,\theta)$ solution of the original ODE dynamics in the opposite direction, whereas the adjoint state is $a(t)$ and the sensitivity of the loss with respect to the hidden state is

Figure 2

Illustration of Neural ODE solution.



$z(t)$. The initial condition comes from the gradient at the last point t_1. The gradient with respect to θ is in Equation (6). The gradients with respect to t_0 and t_1 are in Equations (7)-(8). The advantage of neural ODE is that ODE solvers work independently of the number of layers [25].

Beginning with the initial state $z(t_0)$, the ODE solver solves the system dynamics in a forward direction. The states at one or more time points may have an impact on the loss function value. The gradient is computed by solving, as shown in Figure 2, red representations in the opposite time direction [14].

Important areas of the input data are highlighted using the attention method. Because they assign equal weight to each input, traditional self-attention techniques struggle to capture long-range dependencies and extraneous information noise. To address these issues, EGA was created for dynamic gate mechanisms and global context integration [56]. The content gate and the context gate are the two fundamental parts of EGA [55]. By adjusting conventional attention weights, the content gate concentrates on local patterns. By scaling it into the model, the context gate incorporates global information. EGA better catches long-range interactions and filters noisy data in this way.

As shown in Equation (5), the content gate (G_c) modifies conventional attention weights, considering $W_c \in R^{d\times d}$ and $b_c \in R^d$, where $Q \in R^{n\times d}$ represents the query feature matrix. The gate is computed as:

$$G_c = \sigma(W_c . Q + b_c). \quad (5)$$

As shown in Equation (6), the context gate (G_x) captures global patterns, where $V \in R^{n\times d}$, n is the number of time steps, and d is the feature dimension. Global average pooling is applied along the temporal dimension:

$$G_x = \sigma(W_x . GAP(V) + b_x). \quad (6)$$

In Equation (6), $GAP(V) \in R^d$ and $W_x \in R^{d\times d}$. The multi-head attention mechanism is defined by projecting the input into query, key, and value spaces using learned linear projections: $Q=XW_Q$, $K=XW_K$, and $V=XW_V$. W_Q, W_K, and W_V are learnable linear projection matrices that map input features into query, key, and value subspaces. The attention of each head is computed as:

$$head_i = softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) . V_i . \alpha_i, \quad (7)$$

where K_i^T explicitly denotes the transpose of the key matrix, and α_i is a learnable adaptive scaling factor defined as:

$$\alpha_i = \sigma\left(W_\alpha . (Q_i ; K_i)\right), \quad (8)$$

where [· ; ·] denotes concatenation along the feature dimension. The standard scaled dot-product attention is given by:

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right). \quad (9)$$

Finally, the EGA is formulated as:

$$EGA = G_c \odot A + G_x \odot GAP(V), \quad (10)$$

where $\odot$ denotes element-wise multiplication. In this formulation, $A \in R^{n\times n}$, while $GAP(V) \in R^d$, which is broadcasted to match the attention output dimension during fusion.

The content gate G_C is applied element-wise to the attention matrix $A \in R^{n\times n}$, ensuring attention re-weighting across temporal positions. The context gate G_x is broadcast along the temporal dimension to match the shape of $V \in R^{n\times d_k}$, enabling element-wise modulation of global features. The final EGA output preserves the same dimensionality as the value space, i.e., $R^{n\times d_k}$, and is directly passed to subsequent transformer layers without dimensional mismatch. No additional residual or normalization is applied within the EGA block, as these operations are handled in the subsequent encoder stage.

3.3. Frequency Domain Feature Extraction and FNO-Based Modeling

The frequency domain contains important information for identifying periodic structures in vibration signals and defect-related frequency components. In this context, the proposed technique employs FNO-based frequency feature extraction. Unlike standard convolutional structures, FNO learns directly in the spectral domain and captures global frequency interactions across the signal [36].

In the proposed model, the frequency path processes feature maps through a convolutional block and then feeds them into the FNO, which transforms the representations into Fourier space and models key frequency components [26]. The frequency-domain features are then fused with the time-domain features, enabling cross-domain information exchange. As a result, the model achieves a more comprehensive representation by integrating temporal and frequency information.

In the proposed implementation, the 1D vibration signal is processed using a 1D Fast Fourier Transform (FFT) applied along the temporal axis, converting the input feature map $X \in R^{n\times d}$ into the frequency domain representation $X_f \in C^{n\times d}$. Only the first K low-frequency Fourier modes are retained (mode truncation), while high-frequency components are discarded to reduce noise sensitivity and improve robustness in fault-related frequency extraction.

The FNO operator is a deep learning architecture intended for partial differential equation solutions and function operator learning [12]. It considerably reduces the amount of data needed to produce a high-resolution solution, resulting in faster numerical analysis [49]. The goal of FNO is to learn an operator or external condition and model a specific type of a function [54].

The signal is divided into frequency components using the Fourier transform [10]. A nonlinear mapping $G{:}H_1 \rightarrow H_2$ is obtained to model G_θ in Equation (11).

In this implementation, complex-valued spectral convolution is performed in the Fourier domain using learnable complex weights, where real and imaginary parts are parameterized separately and jointly optimized during training. This allows the model to capture both amplitude and phase variations of vibration-induced fault signatures.

$$G_\theta = H_1 \times \theta \rightarrow H_2. \tag{11}$$

The neural network's trained parameter set is called θ. Equation (12) formulates the neural network operator with a recursive layer design.

$$\begin{cases} z^0 = Px \\ z^{l+1} = L^l\left(z^l\right) = \sigma\left(W^l z^l + b^l + K_\theta^l\left(z^l\right)\right) \\ for\ l = 1,2,3,\ldots,L-1 \\ s = Qz^L \end{cases} \tag{12}$$

In Equation (12), two shallow neural networks, P and Q, encode functions with lower dimensions into higher dimensions and vice versa. $\sigma{:}R \rightarrow R$ is a nonlinear activation function that is based on points. The kernel integral operator is denoted by K_θ^l. K_θ^l is modeled with FNO utilization (12).

In the frequency domain, the Fourier transform F and inverse Fourier transform F^{-1} are applied along the temporal dimension, and only K retained Fourier modes are learned through complex weight matrices $\{R^l\}$. The resulting feature dimensionality is preserved to ensure compatibility with subsequent fusion layers, i.e., $X_f \in R^{K \times d}$. The weight matrices in the Fourier domain are denoted by $\{R^l | l=1,2,3,\ldots,L\text{-}1\}$.

3.4. Strengthening Temporal Dependencies: LSTM + SE

Long-term correlations play a crucial role in the modeling of time-dependent dynamic changes in vibration signals. In this context, LSTM can analyze sequential patterns over time, thereby discerning the order of occurrence and progression of fault symptoms. Furthermore, the integration of an SE block with LSTM introduces a channel-based attention mechanism to the temporal representation. This mechanism allows for the adaptive scaling of the significance of each channel, thereby directing the model's attention towards the most salient features.

By capturing long-term sequential relationships that conventional CNNs are unable to learn, LSTM offers the capability of modeling the evolution of defects over time. As a result, irrelevant features are suppressed and channels with a high information density are reinforced. Complementary information for the model's decision-making process is obtained by learning the channel-based importance order and temporal evolution.

Input gates, forget gates, and output gates are the three parts of an LSTM. These gates decide which information will be provided in pairs, retained, and forgotten [51], [34], [64]. The expression for the LSTM components' forget gate is (13).

$$f_t = \sigma(W_f \cdot \left[h_{t-1}, x_t\right] + b_f). \tag{13}$$

In Equation (13), f_t is the output of the forget gate, which determines which components to forget. W_f is the weights, while b_f is the bias. x_t is the current input, and h_{t-1} is the hidden state at the previous time step.

$$i_t = \sigma(W_i \cdot \left[h_{t-1}, x_t\right] + b_i). \tag{14}$$

In Equation (14), i_t is the input gate's output that determines the relevance of fresh information. The value obtained for adding new information to memory for the contribution to the memory cell is shown in Equation (15). The memory cell is updated in the same way as described in Equation (16). At this point, it is possible to forget some of the previously learned knowledge while retaining new information. Thus, long-term memory is formed.

$$\tilde{c}_t = tanh(W_c \cdot \left[h_{t-1}, x_t\right] + b_c) \tag{15}$$

$$c_t = (f_{t.} \cdot c_{t-1.} + i_t \cdot \tilde{c}_t). \tag{16}$$

The output gate h_t determines how much information is produced, as seen in Equation (17). In this stage, the LSTM cell generates an output based on the updated cell state [63], [9]. This output is used in the following step as well as to control the model's output.

$$h_t = \tanh(c_t) \odot \sigma(W_i . \left[h_{t-1}, x_t\right] + b_i). \quad (17)$$

The SE block improves the performance of deep learning models, particularly in the fields of image and signal processing, by implementing the attention mechanism at the channel level [17], [48]. It teaches how important each channel is and assigns weight accordingly. Deep neural networks' performance can be improved via channel-based attention methods [33]. It reweighs the channels of the feature maps using channel-based attention, allowing the model to concentrate on significant characteristics [13]. This block comprises two stages. The first stage, squeeze, does global pooling on the supplied data. This procedure is followed for each channel of the input. In Equation (18), the channel-level vector is calculated by taking the average of each channel [20].

$$z_c = \sigma(W_2 . \delta(W_1 . F + b_1) + b_2). \quad (18)$$

In Equation (18), weights (W_1 and W_2) F input data includes b_1 and b_2 bias values. δ is the relu activation function. σ represents the sigmoid function. The second stage, Excitation, rescales the output based on the relevance of each channel like in Equation (19).

$$\hat{F} = F . z_c. \quad (19)$$

In Equation (19), $\hat{F}$ is the rescaled output, z_c is the weight vector learned for each channel, and F is the input data feature map. The compressed information obtained by Squeeze is routed over a two-layer fully linked network, with significance scores generated for each channel. This structure learns the relationship between the channels and decides which ones should contribute to the output [33], [13], [20], [52].

3.5. Multi-Scale Attention-Based Feature Fusion

In the suggested approach, the feature maps obtained from the three branches (time-domain, frequency-domain, and LSTM+SE representation) exhibit different representation strengths and semantic scales. Therefore, direct fusion may lead to information loss or dominance of specific branches. To address this issue and appropriately weight the information provided by each branch, a multi-scale attention-based feature fusion module was designed. Since each branch processes the signal using different structures and filtering strategies, the extracted features contain complementary spatial and semantic characteristics.

The proposed attention mechanism highlights temporal continuity captured by the LSTM+SE branch, anomaly-sensitive frequency-domain responses, channel-wise importance, and signal variations in the time domain. The attention network adaptively assigns importance coefficients to each branch feature, enabling content-aware fusion. Consequently, the fusion module reduces redundant information, preserves discriminative fault-related representations, and improves both the accuracy and generalization capability of the proposed model

Table 2

Complete implementation details of the proposed lightweight hybrid architecture.

Module / Branch	Output Shape	Configuration	Parameters
Input Layer	(1024, 1)	Raw vibration signal	-
Shared Feature Extraction	(1024, 12)	Conv1D (6, k=3) + Conv1D (12, k=3) + BatchNorm + GELU	300
Time-Domain Branch	(1024, 12)	Enhanced Gated Attention (2-head self-attention)	325
Frequency-Domain Branch	(1024, 12)	Fourier Neural Operator (Modes=4) + Conv1D Fusion	828
Temporal Modeling Branch	(12)	LSTM (12 units) + Neural ODE Block	1224
Channel Attention Branch	(1, 12)	SE Block (reduction ratio=8)	37
Multi-Branch Fusion	(36)	Concatenation + Attention Scaling	1332
Classification Head	(3)	Dense (24, GELU) + BatchNorm + Dropout(0.3) + Dense (12, GELU) + Dropout(0.3) + Softmax	1323

4. Experimental Results

In this section, we thoroughly analyze the performance of our proposed hybrid model using the SUBF v1.0 industrial vibration dataset. The noise robustness of the model is investigated by evaluating it at different SNR levels (5, 10, 15, and 20 dB). The experiments demonstrate that the multi-attention mechanism of our model, which combines features from the frequency and temporal domains, can identify intricate and variable patterns in vibration signals. According to ablation studies, the EGA module and the Neural ODE components significantly improve model performance. The model's processing speed and parameter count show that it is appropriate for real-time applications. These results show that hybrid multi-domain architecture is a practical and effective way to identify industrial vibration problems as shown in Table 4.

4.1. Experimental Setup

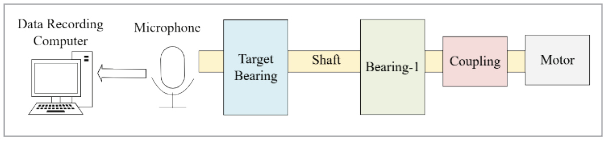
Model training and testing were performed using Python 3.12 and the TensorFlow 2.18 deep learning framework. Experiments were conducted on a computer equipped with a 13th generation Intel Core i9-13900H processor (2.60 GHz) and 32 GB of RAM. The training parameters included 50 epochs a batch size of 256, and the Adam optimizer. The learning rate was set at 0.0001. These hardware specs and training circumstances facilitated the effective learning of complicated multi-domain features.

4.2. Dataset Description

The SUBF Dataset v1.0 is used in this investigation, a large-scale dataset designed primarily for analyzing and diagnosing mechanical bearing faults using vibration signals [4]. The experimental setup consists of a motor, carrier base, bearings, and a shaft, and is designed to simulate various operating conditions including inner race faults, outer race faults, and healthy states. The SUBF v1.0 dataset was introduced by Aziz et al. [4], who proposed a diagnostic framework based on autoregressive features and automated relative energy-based empirical mode decomposition (EMD) for intelligent fault detection. In addition to labeled fault categories, the dataset provides long-duration vibration recordings under controlled operating conditions, making it suitable for analyzing temporal fault evolution and continuous-time modeling approaches.

Figure 3

Block diagram of the in-house SUBF v1.0 data acquisition setup [4].



In the experimental setup, a 0.25 HP 3-phase AC motor was tuned at 1440 RPM and 50 Hz frequency as shown in Figure 3. Three scenarios were simulated by replacing the bearing on the left side: healthy bearing, inner race fault, and outer race fault. Vibration data were recorded using a BeanDevice 2.4 GHz AX-3D wireless sensor, collected via BeanGateway, and stored on a PC. The sampling rate was determined to be 1000 Hz. The availability of long, uninterrupted vibration recordings enables the proposed Neural ODE module to effectively learn smooth temporal transitions and progressive fault dynamics, which are difficult to capture using short or heavily segmented datasets. A total of 18 hours of data were collected, with 6 hours of information accessible for each defect category. The recordings were separated into 10-second pieces, yielding 2160 signals for each fault category. The data set is divided into three categories: healthy, inner, and outer race faults.

4.3. Performance Evaluation Metrics

Several performance metrics are employed to comprehensively evaluate the effectiveness of the proposed model, including accuracy, precision, recall, F1-score, Area Under the Curve (AUC), confusion matrix, and t-SNE visualizations. In Equation (19), accuracy measures the overall proportion of correctly classified samples across all classes, providing a global performance indicator. Precision, defined in Equation (20), measures the proportion of correctly predicted instances for each class relative to all predicted instances of that class. Recall, defined in Equation (21), represents the proportion of correctly predicted instances for each class relative to all actual instances of that class. These metrics quantify class-wise predictive performance. Precision, recall, and F1-score are computed using macro-averaged scheme to ensure balanced evaluation across the three classes (healthy, inner race fault, and outer race fault), mitigating the effects of class imbalance. The AUC metric evaluates the model's discriminative capability across classes. In the multiclass setting, AUC is calculated using a one-vs-rest (OvR) strategy, where each class is sequentially treated as the positive class against the remaining classes. A higher AUC value indicates stronger separability among classes. A confusion matrix provides a detailed class-wise representation of correct and incorrect predictions, enabling a comprehensive analysis of inter-class misclassification behavior. Finally, t-SNE is used to project high-dimensional feature representations into a lower-dimensional space for qualitative visualization of class separability.

In the multiclass formulation, True Positive Rate (TPR) in Equation (22) and False Positive Rate (FPR) in Equation (23) are computed under a one-vs-rest scheme for each class.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{20}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{21}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{22}$$

$$TPR = \frac{TP}{(TP + FN)} \tag{23}$$

$$FPR = \frac{FP}{(FP + TN)} \tag{24}$$

$$AUC = \int_0^1 TPR(t)d(FPR(t)) \tag{25}$$

4.4. Overall Performance of the Proposed Model

The suggested hybrid model has demonstrated higher performance in fault identification using industrial vibration data. The model successfully incorporates multidimensional information taken from the time and frequency domains while also accurately capturing complicated signal dynamics. As a result, it can accurately and precisely classify various defect types and health statuses. When Figure 4 and Table 3 are examined together, it is evident that the proposed model not only converges efficiently during training but also maintains stable and robust performance on the validation data. As shown in Table 3, the training accuracy reaches 98.59%, and the validation accuracy remains high at 98.92%, indicating minimal overfitting. Figure 4 visually confirms this consistency, where training and validation curves closely align, suggesting a well-optimized learning process. These results collectively underline the model's strong generalization ability and reliability for real-world fault diagnosis tasks.

Figure 5 illustrates the confusion matrix of the proposed model, demonstrating its strong classification performance across the three classes. All normal samples were correctly classified, indicating perfect discrimination of healthy conditions. Minor misclassifications occurred between the inner race and outer race fault classes, which may be attributed to the similarity in their vibration characteristics. Overall, the model exhibits robust and reliable fault diagnosis capability, effectively distinguishing between normal and faulty bearing states. Figure 6 presents the ROC curves for the three classes: inner race, normal, and outer race.

Table 3

Overall classification performance of proposed model.

Metrics	Training	Validation	Test
Accuracy	0.9859	0.9892	0.9841
Precision	0.9853	0.9896	0.9881
Recall	0.9867	0.9888	0.9807
F1-Score	0.9860	0.9892	0.9843

Table 4

Computational complexity of proposed model.

Metric	Value
Parameters	6,042
FLOPs	19,161
Model Size	0.02 MB
Inference Latency	8.18 ms
Throughput	110.60 samples/s

Figure 4

The training and validation accuracy curves of the proposed model.

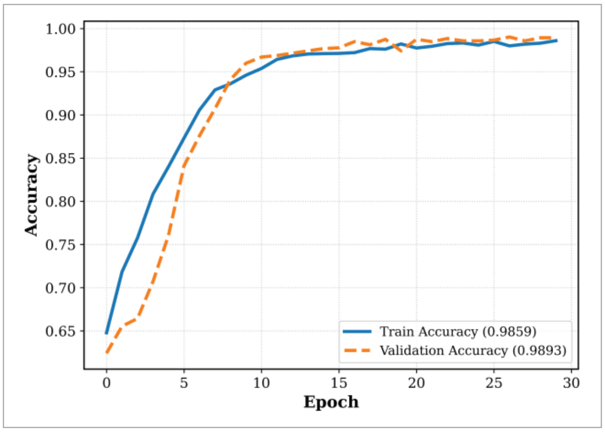


Figure 5

Confusion matrix of the proposed method.

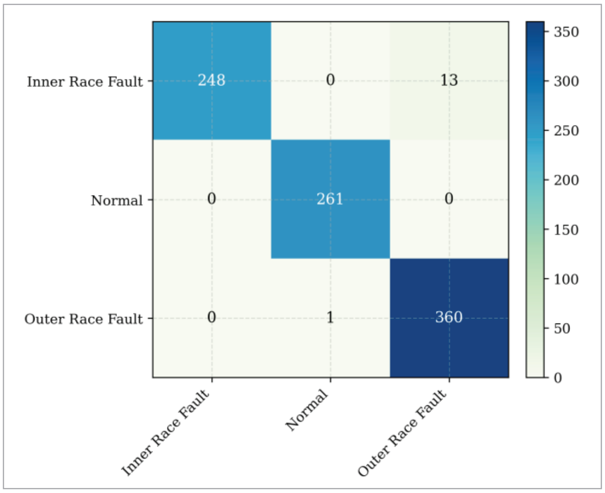


Figure 6

ROC curve for each class of the proposed method.

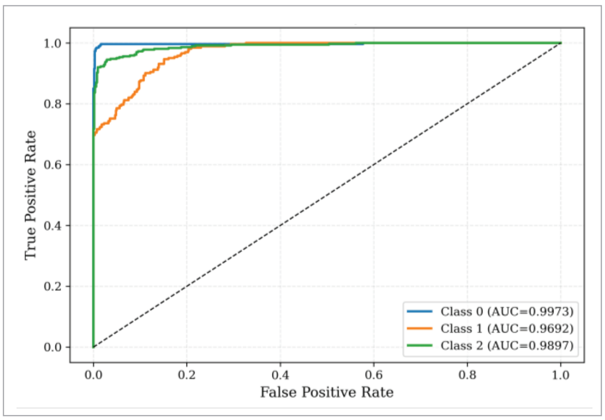


Table 5

Performance comparison with state-of-the-art models.

Model	Accuracy (%)
Dual CNN (Time + Freq) [59]	94.23
Attention-LSTM [24]	95.18
FNO-based Network [33]	96.10
CNN + Gated Attention [47]	96.42
SE-LSTM-CNN [57]	96.75
Multi-Scale CNN + Channel Attention [18]	97.12
Fused CNN-LSTM with FFT [53]	97.50
CNN-LSTM + Attention [7]	97.81
CNN + Attention + Visualization [19]	96.80
Proposed model	98.41

Table 5 compares the proposed model to the present state-of-the-art models used in literature. These results clearly highlight the advantages of our model's hybrid structure, as well as the sophisticated attention and continuous time modeling components. This complete performance review demonstrates that the model is a strong, dependable, and practical solution for industrial machinery defect diagnostics.

4.5. Noise Robustness Analysis

In real-world industrial scenarios, vibration signals are frequently contaminated by various sources of noise, such as mechanical interference, electromagnetic interference, environmental vibrations, and sensor inaccuracies. To simulate realistic industrial environments, Gaussian white noise was systematically added to the original vibration signals at various Signal-to-Noise Ratio (SNR) levels: 20 dB, 15 dB, 10 dB, and 5 dB. These levels represent a range from relatively clean to severely noisy signal conditions commonly encountered in machinery monitoring applications.

The classification performance of the proposed hybrid model was assessed at each noise level and compared with several state-of-the-art baseline models. As shown in Figure 7, the proposed model maintains high accuracy across all noise levels, with only a gradual decline as SNR decreases. Notably, even at 5 dB, where the signal is heavily corrupted, the model achieves robust performance. This resilience is primarily attributed to the EGA mechanism, which

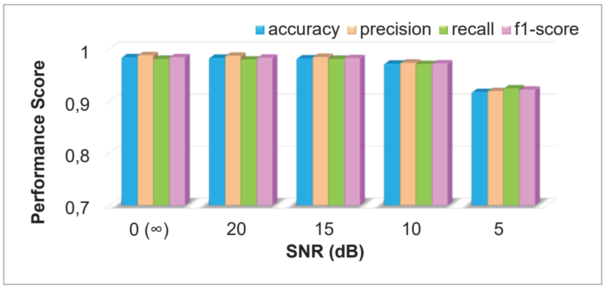
Table 6

Test samples performance of the proposed model under different SNR levels.

SNR(dB)	Acc	P	R	F1
0 (∞)	0.9841	0.9881	0.9807	0.9843
20	0.9830	0.9869	0.9796	0.9832
15	0.9819	0.9845	0.9807	0.9825
10	0.9717	0.9734	0.9711	0.9722
5	0.9173	0.9192	0.9244	0.9218

Figure 7

Performance comparison under varying SNR levels.



selectively suppresses irrelevant and noisy features, and the Neural ODE module, which effectively models the smooth temporal evolution of signal dynamics. Furthermore, Table 6 quantitatively presents the classification metrics, including accuracy, precision, recall, and F1-score, under each SNR level. In a noise-free setting (SNR = ∞), the model achieves nearly perfect accuracy (98.41%) along with high precision (98.81%) and recall (98.07%), ensuring both minimal false alarms and reliable fault detection. As noise increases, a gradual performance drop is observed; however, even at 5 dB SNR, the model maintains over 91% across all metrics, including accuracy, precision, recall, and F1-score.

4.6. Ablation Studies

A comprehensive ablation study was conducted to evaluate the impact of each fundamental component of the proposed hybrid model on the overall performance. In this analysis, the four main modules of the model were sequentially replaced with simpler structures, and each variant was evaluated separately. Specifically, the EGA module was substituted with Vanilla Self-Attention, the Neural ODE component was replaced by a simple 1D convolutional layer, the FNO module was similarly replaced by a simple 1D convolutional layer, and the LSTM + SE block was substituted with a basic convolutional layer. This method seeks to identify and quantify each component's contribution to the classification performance of the model. Notably, the EGA module's selective feature focus and the Neural ODE's continuous-time temporal modeling capacity are essential to preserving the model's high accuracy levels. Additionally, by eliminating each of the three parallel lines separately and examining the performance impact that resulted, the importance of each path within the suggested model was examined. The outcomes shown in Table 7, Figures 8-9 show that each suggested module significantly enhances the efficacy of the model.

When essential model components were systematically removed or replaced, the ablation results reveal

Table 7

Test performance comparison between the proposed framework and ablation models.

Ablation No	Model variant	Accuracy	Precision	Recall	F1-Score
1	EGA replaced with Vanilla Self-Attention	0.8992	0.8995	0.9094	0.9044
2	Neural ODE replaced with Simple Conv	0.8505	0.8640	0.8748	0.8693
3	FNO replaced with Simple Conv	0.8641	0.8771	0.8794	0.8782
4	LSTM + SE replaced with Simple Conv	0.8754	0.8835	0.8911	0.8873
5	No Path1	0.8279	0.8355	0.8539	0.8445
6	No Path2	0.8324	0.8543	0.8631	0.8586
7	No Path3	0.8573	0.8668	0.8808	0.8737
8	Proposed model	0.9841	0.9881	0.9807	0.9843

noticeable performance degradations, highlighting the contribution of each module to the overall framework. In the first ablation study, replacing the critical time-domain EGA mechanism with Vanilla Self-Attention resulted in an accuracy of 89.92%, precision of 89.95%, recall of 90.94%, and an F1-score of 90.44%. This confirms that the EGA mechanism plays a crucial role in enhancing feature discrimination by selectively emphasizing informative temporal patterns while suppressing noise.

In the second variant, the Neural ODE component was replaced with a simple convolutional operation, leading to a significant performance drop, with an accuracy of 85.05%, precision of 86.40%, recall of 87.48%, and an F1-score of 86.93%. This demonstrates that Neural ODE contributes effectively to continuous feature evolution and improved representation stability.

The third variant, where the frequency-domain FNO module was replaced with a simple convolution, achieved an accuracy of 86.41%, precision of 87.71%, recall of 87.94%, and an F1-score of 87.82%. These results indicate that the FNO plays an important role in capturing global frequency-domain characteristics relevant to fault signatures.

In the fourth variant, replacing the LSTM and SE blocks with a simple convolution resulted in an accuracy of 87.54%, precision of 88.35%, recall of 89.11%, and an F1-score of 88.73%. Although the performance drop is moderate, these results confirm that modeling long-term temporal dependencies and adaptive channel recalibration contributes positively to the overall learning process.

Variants five through seven evaluate the impact of removing each parallel path individually. In the fifth variant, eliminating Path 1 (time-domain EGA and Neural ODE components) caused a substantial decline, yielding an accuracy of 82.79%, precision of 83.55%, recall of 85.39%, and an F1-score of 84.45%. This confirms that Path 1 is the most critical component of the architecture, providing fundamental temporal feature extraction.

In the sixth variant, removing Path 2 (frequency-domain FNO branch) resulted in an accuracy of 83.24%, precision of 85.43%, recall of 86.31%, and an F1-score of 85.86%, showing that frequency-domain learning significantly enhances discriminative capability.

The seventh variant, where Path 3 (LSTM + SE blocks) was removed, achieved an accuracy of 85.73%, precision of 86.68%, recall of 88.08%, and an F1-score of 87.37%. This indicates that while Path 3 contributes meaningfully, its impact is less dominant compared to Path 1 and Path 2.

Finally, the proposed full model integrating all three parallel paths achieves the best performance, with an accuracy of 98.41%, precision of 98.81%, recall of 98.07%, and an F1-score of 98.43%. These results clearly demonstrate that the complementary interaction between time-domain, frequency-domain, and temporal-channel modeling pathways leads to superior feature representation and classification performance.

In summary, the ablation study clearly demonstrates that each architectural component contributes to the overall performance of the proposed framework. Among them, Path 1 exhibits the most critical role, while Path 2 and Path 3 provide complementary improvements. The full integration of all components yields the highest performance, confirming the effectiveness of the proposed multi-branch design. Figures 9-10 illustrate the performance comparison of all ablation variants and the proposed model, emphasizing the significant degradation when individual components are removed.

Figure 8

Comparative performance analysis of ablation models 1 to 4 and the proposed model.

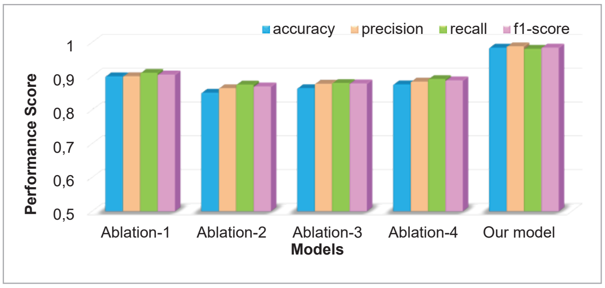


Figure 9

Comparative performance analysis of ablation models 5 to 7 and the proposed model.

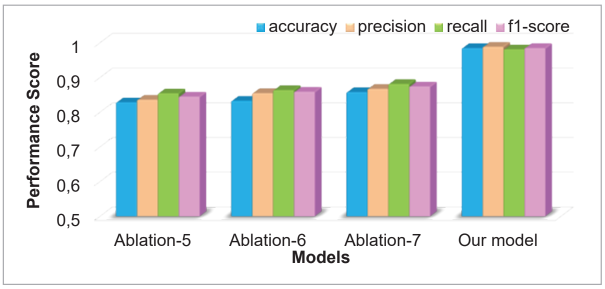


Figure 10

UMAP visualization of (a) the raw input signal space and (b) the learned feature space.

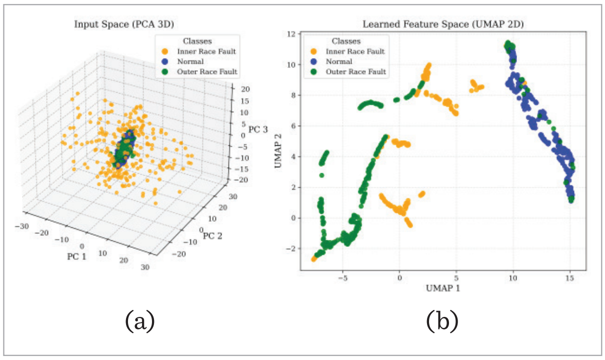
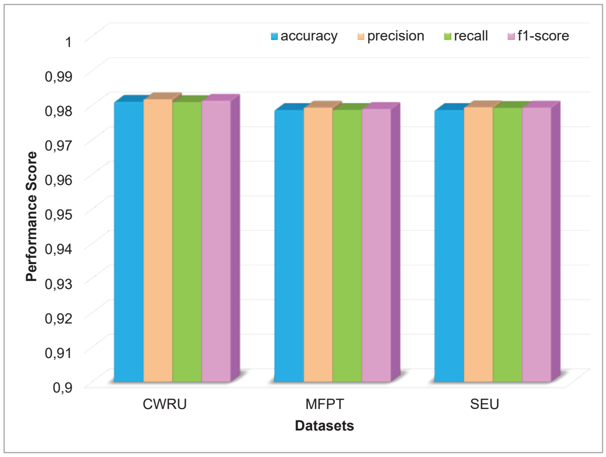


Figure 11

Comparative cross-dataset evaluation of the proposed model.



4.7. Visualization of Learned Feature Spaces

Uniform Manifold Approximation and Projection (UMAP) is used to visualize both the raw input signal space and the learned feature space in two dimensions. In the raw input space, samples from different classes (normal, inner race fault, and outer race fault) are highly entangled with significant overlap, highlighting the difficulty of fault separation from raw vibration signals. In contrast, the learned feature space shows well-separated and compact clusters with minimal overlap, demonstrating the model's ability to extract robust and class-discriminative representations.

This visualization in Figure 10 serves as strong qualitative evidence of the model's capability to distinguish between different fault types. It confirms that the learned representations reduce intra-class variance while maximizing inter-class separation, which is essential for high-accuracy fault diagnosis in real-world rotating machinery systems.

4.8. Cross-Dataset Generalization Analysis

To examine the generalizability of the proposed model, its performance on other similar bearing defect datasets was tested. Besides the SUBF v1.0 dataset, the model was also evaluated on three benchmark bearing fault datasets: CWRU, MFPT, and SEU. Table 8 and Figure 11 show the comparative results obtained, demonstrating that the model is not specific to a particular dataset and can produce highly successful results under different signal characteristics and operating conditions. This indicates that the proposed multi-domain feature extraction and attention-based structures can successfully capture distinctive fault features regardless of the dataset.

CWRU Dataset: This dataset is one of the most frequently used and cited datasets in the literature in the field of bearing fault [5]. The vibration signal obtained at 12 kHz and 48 kHz sampling frequencies using a 2 hp motor in the test setup is found.

MFPT Dataset: It is one of the oldest datasets used in the field of bearing fault diagnosis [44]. The dataset consists of both real-world fault data (from wind turbine bearings) and test setup and artificial data. The dataset has high sampling rates. Data were obtained from different load ranges and 25 Hz frequency.

Table 8

Cross-dataset fault diagnosis performance of the proposed model.

Dataset	Accuracy			Precision			Recall			F1-Score		
	Train	Val	Test	Train	Val	Test	Train	Val	Test	Train	Val	Test
CWRU	0.9778	0.9786	0.9810	0.9778	0.9788	0.9817	0.9780	0.9777	0.9809	0.9779	0.9783	0.9813
MFPT	0.9844	0.9869	0.9786	0.9842	0.9879	0.9793	0.9849	0.9879	0.9786	0.9845	0.9879	0.9789
SEU	0.9811	0.9845	0.9786	0.9809	0.9844	0.9794	0.9814	0.9847	0.9792	0.9811	0.9845	0.9793

SEU Dataset: It includes data on both bearing and gear failures [40]. The data obtained was taken from a simulator. It includes 8-channel multivariate time series data. Vibration was sampled at 12 kHz. Data were obtained from two different speed and load configurations: 20 Hz-0V and 30 Hz-2V.

5. Conclusions and Future Work

This study proposes a novel hybrid architecture that simultaneously exploits time-domain, frequency-domain, and continuous temporal dynamics. The time-domain branch uses convolutional layers and an EGA mechanism to highlight fault-relevant features while suppressing noise. Neural ODEs model the continuous evolution of faults, providing smooth and interpretable temporal representations. The frequency-domain branch employs a FNO network with gated attention to focus on key spectral components. The LSTM layers capture temporal dependencies, and SE blocks recalibrate feature maps adaptively. A multi-scale attention fusion module integrates outputs from all branches, enabling robust and accurate vibration fault diagnosis even in noisy environments.

Future work will build upon the structural characteristics of the proposed model in several concrete technical directions. One promising extension is the development of condition-aware or physics-informed Neural ODE formulations, where the continuous-time dynamics are explicitly conditioned on operating variables such as load, speed, or temperature, enabling more accurate modeling of fault progression and remaining useful life estimation. Another direction involves adaptive cross-domain attention, in which the relative contributions of time-domain and frequency-domain representations are dynamically adjusted according to operating conditions and noise levels, rather than being statically fused. These extensions would further enhance the model's interpretability, robustness, and applicability to complex industrial environments while preserving the multi-domain and continuous-time modeling advantages demonstrated in this study.

Funding Information

This work was not financially supported by any institution or organization.

Conflict of Interest

All authors declare that they have no conflicts of interest.

Data Availability Statement

The dataset analyzed during the current study is the SUBF v1.0 bearing fault vibration dataset, which is publicly available on Kaggle at: https://www.kaggle.com/datasets/sumairaziz/subf-v1-0-dataset-bearing-fault-vibration-data

Supplementary Information

The code used to develop the proposed model and reproduce the experimental results is publicly available at: https://github.com/ctastimurtemiz/Hybrid-FaultDiagnosis.

Appendix – A

Implementation Details and Hyperparameter Settings

A.1 Enhanced Gated Attention (EGA) Module

- Number of attention heads: 1
- Gating function: sigmoid
- Query/Key projection dimension: d_model / 2
- Kernel size: k = 1
- Dropout rate: 0.2

A.2 Neural Ordinary Differential Equation (Neural ODE) Configuration

- ODE solver: Runge–Kutta (RK4)
- Integration time interval: [0, 1]
- Maximum number of function evaluations: 2
- ODE function architecture: 2-layer MLP (hidden size: 12)
- Activation function: Tanh / Swish
- Adjoint sensitivity method: enabled

A.3 Training Stabilization and Optimization Details

- Optimizer: Adam
- Epoch: 30
- Initial learning rate: 1e-4
- Learning rate scheduler: ReduceLROnPlateau
- L2 regularization coefficient: 0.005
- Batch size: 128

References

1. Ayankoso, S., He, Y., Gu, F., Ball, A. Multisensory Data-Based Fault Diagnosis of Induction Motors Using 1D and 2D Convolutional Neural Networks. Proceedings of the International Conference on the Efficiency and Performance Engineering Network, 2023, 1125-1135. https://doi.org/10.1007/978-3-031-49421-5_92
2. An, Y., Zhang, K., Liu, Q., Chai, Y., Huang, X. Rolling Bearing Fault Diagnosis Method Base on Periodic Sparse Attention and LSTM. IEEE Sensors Journal, 2022, 22(12), 12044-12053. https://doi.org/10.1109/JSEN.2022.3173446
3. Antoni, J. The Spectral Kurtosis: A Useful Tool for Characterising Non-Stationary Signals. Mechanical Systems and Signal Processing, 2006, 20(2), 282-307. https://doi.org/10.1016/j.ymssp.2004.09.001
4. Aziz, S., Khan, M. U., Faraz, M., Montes, G. A. Intelligent Bearing Faults Diagnosis Featuring Automated Relative Energy-Based Empirical Mode Decomposition and Novel Cepstral Autoregressive Features. Measurement, 2023, 216, 112871. https://doi.org/10.1016/j.measurement.2023.112871
5. Case Western Reserve University Bearing Data Center. Bearing Data Center. Case School of Engineering, Cleveland, OH. Available at: https://engineering.case.edu/bearingdatacenter (Accessed: 27 January 2026).
6. Che, C., Wang, H., Fu, Q., Ni, X. Deep Transfer Learning for Rolling Bearing Fault Diagnosis under Variable Operating Conditions. Advances in Mechanical Engineering, 2019, 11(12), 1687814019897212. https://doi.org/10.1177/1687814019897212
7. Chen, Y., Zhang, R., Gao, F. Fault Diagnosis of Industrial Process Using Attention Mechanism with 3DCNN-LSTM. Chemical Engineering Science, 2024, 293, 120059. https://doi.org/10.1016/j.ces.2024.120059
8. Chen, R. T. Q., Rubanova, Y., Bettencourt, J., Duvenaud, D. K. Neural Ordinary Differential Equations. Advances in Neural Information Processing Systems, 2018, 31, 6572-6583.
9. Chen, X., Zhang, B., Gao, D. Bearing Fault Diagnosis Based on Multi-Scale CNN and LSTM Model. Journal of Intelligent Manufacturing, 2021, 32(4), 971-987. https://doi.org/10.1007/s10845-020-01600-2
10. Cheng, Y., Tang, B., Hu, Z., Liu, J. Fourier Neural Operator Network for Signal Generation and Prediction. In Proceedings of the 2024 International Conference on Microwave and Millimeter Wave Technology (ICMMT), 2024, 1-3. https://doi.org/10.1109/ICMMT61774.2024.10672398
11. Cui, L., Tian, X., Wei, Q., Liu, Y. A Self-Attention Based Contrastive Learning Method for Bearing Fault Diagnosis. Expert Systems with Applications, 2024, 238, 121645. https://doi.org/10.1016/j.eswa.2023.121645
12. Duruisseaux, V., Kossaifi, J., Anandkumar, A. Fourier Neural Operators Explained: A Practical Perspective. arXiv Preprint arXiv:2512.01421, 2025.
13. Gao, H., Ma, J., Zhang, Z., Cai, C. Bearing Fault Diagnosis Method Based on Attention Mechanism and Multi-Channel Feature Fusion. IEEE Access, 2024, 12, 45011-45025. https://doi.org/10.1109/ACCESS.2024.3381618
14. Gholami, A., Keutzer, K., Biros, G. ANODE: Unconditionally Accurate Memory-Efficient Gradients for Neural ODEs. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019, 730-736. https://doi.org/10.24963/ijcai.2019/103
15. Haidong, S., Hongkai, J., Xingqiu, L., Shuaipeng, W. Intelligent Fault Diagnosis of Rolling Bearing Using Deep Wavelet Auto-Encoder with Extreme Learning Machine. Knowledge-Based Systems, 2018, 140, 1-14. https://doi.org/10.1016/j.knosys.2017.10.024
16. He, K., Zhang, X., Ren, S., Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the European Conference on Computer Vision (ECCV), 2016, 630-645. https://doi.org/10.1007/978-3-319-46493-0_38
17. Hu, J., Shen, L., Sun, G. Squeeze- and -Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745
18. Huang, Y. J., Liao, A. H., Hu, D. Y., Shi, W., Zheng, S. B. Multi-Scale Convolutional Network with Channel Attention Mechanism for Rolling Bearing Fault Diagnosis. Measurement, 2022, 203, 111935. https://doi.org/10.1016/j.measurement.2022.111935
19. Yang, Z. B., Zhang, J. P., Zhao, Z. B., Zhai, Z., Chen, X. F. Interpreting Network Knowledge with Attention Mechanism for Bearing Fault Diagnosis. Applied Soft Computing, 2020, 97, 106829. https://doi.org/10.1016/j.asoc.2020.106829

20. Yin, J., Cen, G. Intelligent Motor Bearing Fault Diagnosis Using Channel Attention-Based CNN. World Electric Vehicle Journal, 2022, 13(11), 208. https://doi.org/10.3390/wevj13110208

21. Ince, T., Kiranyaz, S., Eren, L., Askar, M., Gabbouj, M. Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. IEEE Transactions on Industrial Electronics, 2016, 63(11), 7067-7075. https://doi.org/10.1109/TIE.2016.2582729

22. Janssens, O., Schulz, R., Slavkovikj, V., Stockman, K., Loccufier, M., Van de Walle, R., Van Hoecke, S. Thermal Image-Based Fault Diagnosis for Rotating Machinery. Infrared Physics Technology, 2015, 73, 78-87. https://doi.org/10.1016/j.infrared.2015.09.004

23. Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., Van de Walle, R., Van Hoecke, S. Convolutional Neural Network Based Fault Detection for Rotating Machinery. Journal of Sound and Vibration, 2016, 377, 331-345. https://doi.org/10.1016/j.jsv.2016.05.027

24. Khoshsirat, S., Kambhamettu, C. A Transformer-Based Neural ODE for Dense Prediction. Machine Vision and Applications, 2023, 34, 113. https://doi.org/10.1007/s00138-023-01465-4

25. Kidger, P., Morrill, J., Foster, J., Lyons, T. Neural Controlled Differential Equations for Irregular Time Series. In Advances in Neural Information Processing Systems (NeurIPS 2020), 2020, 33. https://arxiv.org/abs/2005.08926

26. Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., Anandkumar, A. Neural Operator: Learning Maps Between Function Spaces with Applications to PDEs. Journal of Machine Learning Research, 2023, 24(89), 1-97.

27. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A. K. Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. Mechanical Systems and Signal Processing, 2020, 138, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

28. Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S. X. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. IEEE Transactions on Industrial Electronics, 2016, 63(5), 3137-3147. https://doi.org/10.1109/TIE.2016.2519325

29. Lei, Y., Zuo, M. J., He, Z. Condition Monitoring and Fault Diagnosis of Planetary Gearboxes: A Review. Measurement, 2014, 48, 292-305. https://doi.org/10.1016/j.measurement.2013.11.012

30. Li, F., Li, Y., Wang, D. Rolling Bearing Fault Diagnosis via Temporal-Graph Convolutional Fusion. Sensors, 2025, 25(13), 3894. https://doi.org/10.3390/s25133894

31. Li, J., Luo, W., Bai, M. Review of Research on Signal Decomposition and Fault Diagnosis of Rolling Bearing Based on Vibration Signal. Measurement Science and Technology, 2024, 35(9), 092001. https://doi.org/10.1088/1361-6501/ad4eff

32. Li, X., Zhang, C., Li, X., Zhang, W. Federated Transfer Learning in Fault Diagnosis under Data Privacy with Target Self-Adaptation. Journal of Manufacturing Systems, 2023, 68, 523-535. https://doi.org/10.1016/j.jmsy.2023.05.006

33. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A. Fourier Neural Operator for Parametric Partial Differential Equations. arXiv Preprint arXiv:2010.08895, 2020.

34. Liu, Y., Cheng, Y., Zhang, Z., Yang, S. Early Fault Diagnosis of Bearing Faults Using Vibration Signals. In Proceedings of the IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2021, 747-751. https://doi.org/10.1109/ICCASIT53235.2021.9633352

35. Liu, R., Yang, B., Zio, E., Chen, X. Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. Mechanical Systems and Signal Processing, 2018, 108, 33-47. https://doi.org/10.1016/j.ymssp.2018.02.016

36. Pathak J., Subramanian S., Harrington P., Raja S., Chattopadhyay A., Mardani M., Kurth T., Hall D., Li Z., Azizzadenesheli K., Hassanzadeh P., Kashinath K., Anandkumar A. FourCastNet: A Global Data-Driven High-Resolution Weather Model Using Adaptive Fourier Neural Operators. arXiv Preprint arXiv:2202.11214, 2022. https://doi.org/10.48550/arXiv.2202.11214

37. Pontryagin, L. S. Mathematical Theory of Optimal Processes. Routledge, 2018. https://doi.org/10.1201/9780203749319

38. Randall, R. B., Antoni, J. Rolling Element Bearing Diagnostics-A Tutorial. Mechanical Systems and Signal Processing, 2011, 25(2), 485-520. https://doi.org/10.1016/j.ymssp.2010.07.017

39. Randall, R. B., Antoni, J., Smith, W. A. A Survey of the Application of the Cepstrum to Structural Modal Analysis. Mechanical Systems and Signal Processing, 2019, 118, 716-741. https://doi.org/10.1016/j.ymssp.2018.08.059

40. School of Mechanical Engineering, Southeast University, China. SEU Bearing Fault Dataset. Available at: https://github.com/cathysiyu/Mechanical-dataset (Accessed: 27 January 2026)
41. Shao, H., Jiang, H., Zhang, X., Niu, M. Rolling Bearing Fault Diagnosis Using an Optimization Deep Belief Network. Measurement Science and Technology, 2015, 26(11), 115002. https://doi.org/10.1088/0957-0233/26/11/115002
42. Shen, Q., Zhang, Z. Fault Diagnosis Method for Bearing Based on Attention Mechanism and Multi-Scale Convolutional Neural Network. IEEE Access, 2024, 12, 12940-12952. https://doi.org/10.1109/ACCESS.2024.3357113
43. Smith, W. A., Randall, R. B. Rolling Element Bearing Diagnostics Using the Case Western Reserve University Data: A Benchmark Study. Mechanical Systems and Signal Processing, 2015, 64-65, 100-131. https://doi.org/10.1016/j.ymssp.2015.04.021
44. Society for Machinery Failure Prevention Technology (MFPT). MFPT Fault Data Sets. Available at: https://www.mfpt.org/fault-data-sets (Accessed: 27 January 2026).
45. Tandon, N., Choudhury, A. A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings. Tribology International, 1999, 32(8), 469-480. https://doi.org/10.1016/S0301-679X(99)00077-8
46. Wang, B., Lei, Y., Li, N., Li, N., Yan, T. Deep Separable Convolutional Network for Remaining Useful Life Prediction of Machinery. Mechanical Systems and Signal Processing, 2020, 135, 106330. https://doi.org/10.1016/j.ymssp.2019.106330
47. Wang, H., Liu, Z., Peng, D., Cheng, Z. Attention-Guided Joint Learning CNN with Noise Robustness for Bearing Fault Diagnosis and Vibration Signal Denoising. ISA Transactions, 2022, 128, 470-484. https://doi.org/10.1016/j.isatra.2021.11.028
48. Wang, H., Xu, J., Yan, R., Gao, R. X. A New Intelligent Bearing Fault Diagnosis Method Using SDP Representation and SE-CNN. IEEE Transactions On Instrumentation and Measurement, 2019, 69(5), 2377-2389. https://doi.org/10.1109/TIM.2019.2956332
49. Wang, Y., Zhang, H., Lai, C., Hu, X. Transfer Learning Fourier Neural Operator for Solving Parametric Frequency-Domain Wave Equations. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62, 1-11. https://doi.org/10.1109/TGRS.2024.3440199
50. Wang, Z., Xu, X., Song, D., Zheng, Z., Li, W. A Novel Bearing Fault Diagnosis Method Based on Improved Convolutional Neural Network and Multi-Sensor Fusion. Machines, 2025, 13(3), 216. https://doi.org/10.3390/machines13030216
51. Wu, C., Zheng, S. Fault Diagnosis Method of Rolling Bearing Based on MSCNN-LSTM. Computers, Materials Continua, 2024, 79(3), 4395. https://doi.org/10.32604/cmc.2024.049665
52. Wu, Y. Deep Convolutional Neural Network Based on Densely Connected Squeeze- and-Excitation Blocks. AIP Advances, 2019, 9(6), 065306. https://doi.org/10.1063/1.5100577
53. Wu, T., Guo, Y. An End-to-End Fault Diagnosis Method: Dual-Branch Network Feature Fusion Model Combined with MSACNN and Bi-LSTM. In Proceedings of the IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), 2024, 53-58. https://doi.org/10.1109/I2CACIS61270.2024.10649851
54. Wu, Z., Sun, S., Tang, M., Li, P. A Fourier Neural Operator Enhanced Physics-Embedded Iterative Learning Solver for Electromagnetic Scattering Analysis. IEEE Antennas and Wireless Propagation Letters, 2025, 24(7), 1954-1958. https://doi.org/10.1109/LAWP.2025.3552500
55. Xiang, L., Han, Z., Li, X., Hu, A. A Multi-Scale Gated Spatial Attention Network for Rotor Fault Diagnosis under Various Conditions and Equipment. Nondestructive Testing and Evaluation, 2025, 1-21. https://doi.org/10.1080/10589759.2025.2537320
56. Xue, L., Li, X., Zhang, N. L. Not All Attention Is Needed: Gated Attention Network for Sequence Data. In Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4), 6550-6557. https://doi.org/10.1609/aaai.v34i04.6129
57. Zeng, X., Feng, X. Bearing Fault Diagnosis with Small Sample Sizes Based on CNN-SE-LSTM Transfer Learning. In Proceedings of the 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2025, 2617-2622. https://doi.org/10.1109/YAC66630.2025.11149878
58. Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., Si, Y. A Data-Driven Design for Fault Detection of Wind

Turbines Using Random Forests and XGBoost. IEEE Access, 2018, 6, 21020-21031. https://doi.org/10.1109/ACCESS.2018.2818678

59. Zhang, Y., Zhang, S., Zhu, Y., Ke, W. Cross-Domain Bearing Fault Diagnosis Using Dual-Path CNN and Multi-Parallel GCNs. ISA Transactions, 2024, 152, 129-142. https://doi.org/10.1016/j.isatra.2024.06.009

60. Zhao, H., Liu, C., Dang, X., Xu, J., Deng, W. Few-Shot Cross-Domain Fault Diagnosis of Transportation Motor Bearings Using MAML-GA. IEEE Transactions on Transportation Electrification, 2025. https://doi.org/10.1109/TTE.2025.3625779

61. Zhao, M., Kang, M., Tang, B., Pecht, M. Deep Residual Networks with Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes. IEEE Transactions on Industrial Electronics, 2017, 65(5), 4290-4300. https://doi.org/10.1109/TIE.2017.2762639

62. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R. X. Deep Learning and Its Applications to Machine Health Monitoring. Mechanical Systems and Signal Processing, 2019, 115, 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

63. Zhao, Z., Liu, K., Zhao, J. Bearing Fault Diagnosis Based on 1-D DenseNet-LSTM. In Proceedings of the International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), 2022, 590-594. https://doi.org/10.1109/ICCEAI55464.2022.00126

64. Zhou, H., Cheng, L., Teng, L., Sun, H. Bearing Fault Diagnosis Based on RF-PCA-LSTM Model. In Proceedings of the 2nd Information Communication Technologies Conference (ICTC), 2021, 278-282. https://doi.org/10.1109/ICTC51749.2021.9441578