

Article

A Hybrid Deep Learning Approach for Bearing Fault Diagnosis Using Continuous Wavelet Transform and Attention-Enhanced Spatiotemporal Feature Extraction

Muhammad Farooq Siddique ¹, Faisal Saleem ¹, Muhammad Umar ¹, Cheol Hong Kim ² and Jong-Myon Kim ^{1,*}

¹ Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan 44610, Republic of Korea; mfarooq229@mail.ulsan.ac.kr (M.F.S.); faisal1999@mail.ulsan.ac.kr (F.S.); muhammadumar@mail.ulsan.ac.kr (M.U.)

² School of Computer Science and Engineering, Soongsil University, Seoul 06978, Republic of Korea; cheolhong@ssu.ac.kr

* Correspondence: jmkim07@ulsan.ac.kr; Tel.: +82-52-259-2217

Abstract: This study presents a hybrid deep learning approach for bearing fault diagnosis that integrates continuous wavelet transform (CWT) with an attention-enhanced spatiotemporal feature extraction framework. The model combines time-frequency domain analysis using CWT with a classification architecture comprising multi-head self-attention (MHSA), bidirectional long short-term memory (BiLSTM), and a 1D convolutional residual network (1D conv ResNet). This architecture effectively captures both spatial and temporal dependencies, enhances noise resilience, and extracts discriminative features from non-stationary and nonlinear vibration signals. The model is initially trained on a controlled laboratory bearing dataset and further validated on real and artificial subsets of the Paderborn bearing dataset, demonstrating strong generalization across diverse fault conditions. t-SNE visualizations confirm clear separability between fault categories, supporting the model's capability for precise and reliable feature learning and strong potential for real-time predictive maintenance in complex industrial environments.



Academic Editor: Fanglai Zhu

Received: 31 March 2025

Revised: 23 April 2025

Accepted: 24 April 2025

Published: 25 April 2025

Citation: Siddique, M.F.; Saleem, F.; Umar, M.; Kim, C.H.; Kim, J.-M. A Hybrid Deep Learning Approach for Bearing Fault Diagnosis Using Continuous Wavelet Transform and Attention-Enhanced Spatiotemporal Feature Extraction. *Sensors* **2025**, *25*, 2712. <https://doi.org/10.3390/s25092712>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rotary machines are fundamental to modern industry, playing an essential role in sectors such as manufacturing, aerospace, transportation, and power generation [1,2]. Ensuring their reliable operation is essential to prevent unexpected failures, reduce downtime, and avoid substantial economic losses. Among the key components of rotary machines, rolling bearings are particularly critical as they support rotational motion and maintain mechanical stability [3,4]. However, bearings are also among the most failure-prone elements, responsible for nearly 50% of mechanical breakdowns [5]. Common causes of bearing faults include electrical stress, unbalanced loads, material fatigue, and overloading, which can result in secondary damage and unplanned shutdowns [6,7]. As such, bearing fault diagnosis (BFD) is vital for maintaining operational safety and extending machinery lifespan. A core aspect of BFD is fault feature extraction, which involves analyzing vibration signals collected by sensors through effective signal processing methods to detect fault-related patterns [8]. Although recent advancements in intelligent diagnostic techniques have significantly improved fault detection accuracy, challenges remain, particularly in

noisy industrial environments where signal clarity is compromised [9]. Therefore, the development of robust and noise-resilient fault diagnosis methods continues to be a crucial area of research in industrial health monitoring [10–12].

Traditional approaches to bearing fault diagnosis rely on model-based, empirical, or machine learning (ML)-driven methods. Model-based techniques use mathematical equations to simulate and predict bearing behavior, whereas empirical methods depend on domain expertise and manual analysis of vibrational or acoustic emission signals [13,14]. However, the growing complexity of modern machinery has made it increasingly difficult to develop accurate models and rely solely on human expertise. Additionally, these traditional methods often struggle to adapt to dynamic operating conditions, varying fault severities, and noise-prone environments. They also require extensive feature engineering, which is time-consuming and prone to human bias. To overcome these limitations, data-driven methods, particularly those based on deep learning (DL) have gained significant traction in fault detection and classification due to their ability to automatically learn complex patterns directly from raw sensor data [15–17].

In vibration signal-based fault diagnosis, various machine learning algorithms, such as support vector machine (SVM), random forest, and K-nearest neighbors (KNN), have been widely used [18–20]. These methods involve multiple steps, including signal pre-processing, feature extraction, and classification. However, traditional ML-based approaches often depend heavily on manually crafted features, limiting their generalizability and efficiency in real-world scenarios. To address these challenges, deep learning methods, particularly convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, have emerged as powerful alternatives. CNNs have demonstrated remarkable ability in extracting spatial patterns from vibration signals, while BiLSTM models excel at capturing temporal dependencies in time-series data [21].

Deeper DL architectures further enhance performance by capturing complex patterns; however, they often encounter gradient-related issues like vanishing or exploding gradients that hinder effective training and reduce diagnostic accuracy. To mitigate these problems, He et al. [22] introduced the residual network (ResNet), which employs shortcut connections to maintain gradient flow during backpropagation. Building on this, Zhang et al. [23] proposed an attention-augmented ResNet for fault diagnosis in wind-turbine gearboxes, which effectively captured time-frequency features and enhanced classification performance. Similarly, Liang et al. [24] used wavelet transformation in combination with an improved ResNet for bearing fault detection, while Zhao et al. [25] introduced a deep residual shrinkage network that further improved diagnostic precision. These innovations demonstrate the advantages of residual learning frameworks like ResNet over traditional CNN architectures in fault diagnosis tasks. Despite the notable progress in deep learning-based fault diagnosis, current models still face challenges in effectively capturing both spatial and temporal features, especially in nonstationary and noisy industrial environments. Standard CNNs often struggle to focus on the most relevant regions of complex signals, while LSTM-based models alone may not fully exploit local feature hierarchies or suppress irrelevant noise [21]. Additionally, many existing approaches lack the adaptability needed to generalize across varying fault conditions and datasets [26]. These limitations highlight the need for a more comprehensive and robust diagnostic framework capable of extracting discriminative features, handling temporal dependencies, and enhancing attention to critical signal components under real-world conditions [27].

This study builds upon these advancements by proposing a novel hybrid deep learning approach for bearing fault diagnosis. The proposed model integrates CWT with a deep learning framework combining MHSA, BiLSTM, and 1D conv ResNet. This integration enhances feature extraction, suppresses noise, and improves the model's ability to

learn temporal dependencies from nonstationary and nonlinear vibrational signals. Unlike conventional CNNs, the inclusion of MHSA allows the model to focus on critical signal regions, while BiLSTM captures long-term dependencies essential for accurate fault identification. The effectiveness of the proposed model is evaluated using both a controlled laboratory bearing dataset and the publicly available Paderborn bearing dataset to assess its generalizability across different fault conditions. The model achieves superior classification performance, demonstrating high precision, recall, and robustness in distinguishing fault types. The results confirm the model's suitability for real-time predictive maintenance, offering an efficient and reliable solution for intelligent fault detection in industrial applications. The key innovations and contributions of this study are as follows:

- (1) Hybrid Deep Learning Architecture: A novel deep learning framework is introduced, combining 1D conv ResNet, MHSA, and BiLSTM. This hybrid approach enhances feature extraction, suppresses noise, and captures long-range dependencies in nonstationary and nonlinear signals.
- (2) Improved Fault Classification Performance: Unlike conventional backpropagation-based models, the proposed architecture efficiently processes time-series data with enhanced feature learning and temporal modelling, leading to superior classification accuracy and robustness.
- (3) Comprehensive Validation on Multiple Datasets: The model was validated on both a bearing testbed dataset and the Paderborn dataset, demonstrating its generalizability and effectiveness in diverse fault conditions.
- (4) Enhanced Interpretability and Reliability: The experimental results confirm the model's ability to accurately distinguish fault types, with t-SNE visualizations showcasing clear separability between fault classes, making it a viable solution for real-time predictive maintenance in industrial applications.

The remainder of this paper is structured as follows:

- Section 2 details the proposed methodology, including data preprocessing, feature extraction, and model architecture;
- Section 3 presents the experimental setup, results, and performance analysis;
- Section 4 provides the conclusion and potential future directions.

2. Proposed Method for Fault Diagnosis in Bearings

The proposed methodology integrates CWT for time-frequency analysis with a hybrid deep learning model combining MHSA, BiLSTM, and 1D conv ResNet in the following four steps as shown in Figure 1.

Step 1—Data Acquisition:

Vibration signals are collected from two setups: the UIAI Lab experimental platform and the Paderborn bearing test rig. These signals, representing different bearing fault conditions, are acquired using accelerometers and AE sensors to ensure high-quality data for analysis.

Step 2—Signal Preprocessing:

The raw signals are first transformed into time-frequency representations using CWT, producing scalograms. These scalograms are then converted to grayscale and subjected to data augmentation techniques (e.g., flipping, rotation, affine transformation, color jitter, and Gaussian blur) to enhance generalization and reduce class imbalance.

Step 3—Feature Extraction:

The preprocessed scalograms are fed into a hybrid deep learning framework. A 1D residual network extracts spatial features, which are refined using a MHSA mechanism to focus on critical fault regions. The BiLSTM layer then captures temporal dependencies, ensuring comprehensive feature representation.

Step 4—Classification:

The refined features are passed through FC layers with softmax activation to classify the signal into one of four categories: inner race fault, outer race fault, roller fault, or normal condition (i.e., IRF, ORF, RF, NC respectively). Results are visualized using metrics such as confusion matrices, t-SNE plots and ROC curves to confirm classification performance and fault separability.

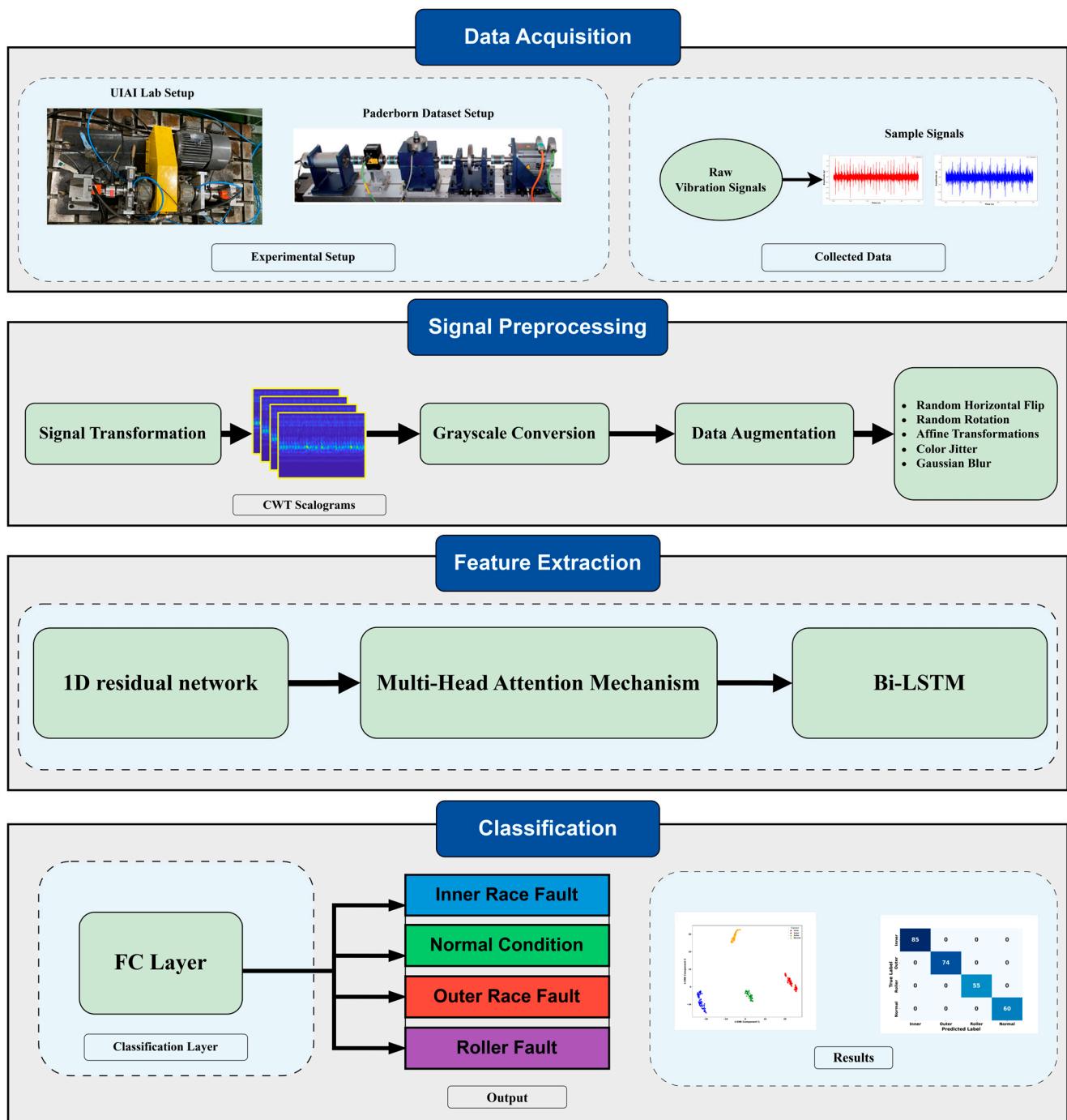


Figure 1. Workflow of the proposed hybrid CWT-based deep learning model integrating MHSA, BiLSTM, and 1D conv ResNet for bearing fault diagnosis.

2.1. Continuous Wavelet Transform

CWT is a fundamental technique in analyzing bearing vibration signals, especially for detecting and classifying faults in rolling element bearings. Defects in bearing components generate transient and non-stationary vibrations, which challenge traditional techniques like Fourier transform and short-time Fourier transform due to their assumptions of stationarity and fixed window sizes [28]. CWT overcomes these by offering a time-frequency representation, allowing precise fault signature identification [29]. Mathematically, the CWT is defined as a convolution of the signal $x(t)$ with a scaled and translated wavelet function $\psi(t)$, given by.

$$C(a, b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

where a represents the scale factor, which determines the frequency resolution, and b denotes the translation, which controls the time localization of the transformation. The mother wavelet $\psi(t)$ serves as a basis function, affecting the sensitivity of CWT to fault-related features. A larger a highlights low-frequency components, while a smaller a captures high-frequency transients. This multi-resolution nature is ideal for detecting localized impulses from defects [30]. The scalogram defined as,

$$S(a, b) = |C(a, b)|^2 \quad (2)$$

Equation (2) provides a visual representation of how energy is distributed across time and frequency. The presence of localized high-energy regions in the scalogram often indicates the occurrence of periodic fault impacts. Adaptive filtering may suppress fault-related high frequencies [31]. EMD suffers from mode mixing, embedding noise in fault features [32]. CWT provides an effective solution by distributing noise energy across multiple scales, reducing its impact on fault detection while preserving transient fault signatures. Another key aspect of CWT in bearing fault diagnosis is its ability to detect fault frequencies associated with defects in different bearing components can be mathematically expressed as,

$$f_{bpf} = \frac{n}{2} f_r (1 + \frac{D}{d} \cos \theta) \quad (3)$$

$$f_{bpfo} = \frac{n}{2} f_r (1 - \frac{D}{d} \cos \theta) \quad (4)$$

where f_{bpf} and f_{bpfo} correspond to the fault frequencies of the inner and outer races, respectively, n is the number of rolling elements, f_r is the shaft rotational frequency, D is the pitch diameter, d is the rolling element diameter, and θ is the contact angle. These frequencies appear as periodic bursts in scalograms with unique energy patterns. Wavelet choice influences CWT's performance. The Morlet wavelet, resembling a sinusoid modulated by a Gaussian, is ideal for frequency analysis. Selecting wavelets involves balancing time and frequency resolution to capture transients and minimize noise. A key feature extraction method enabled by CWT is the calculation of wavelet energy entropy, which quantifies the complexity of the time-frequency representation and can serve as an indicator of fault severity. The wavelet energy entropy is defined as

$$E = - \sum_{i=1}^n p_i \log p_i \quad (5)$$

where p_i is normalized wavelet energy in each band. Higher entropy implies complex, multi-frequency signals; lower entropy indicates regular patterns, possibly from well-defined faults. This helps distinguish fault types and track damage severity. CWT transforms vibration signals into a domain where fault characteristics are clear. Its scalograms

carry rich diagnostic information for precise fault localization and severity estimation. By capturing transient and stationary components, CWT offers better interpretability over conventional time and frequency-domain methods [33].

2.2. 1D Residual Network

A 1D ResNet is an advanced deep learning architecture designed for sequential one-dimensional data, making it well-suited for vibration signals, time-series analysis, and acoustic emissions. In the proposed model, it extracts meaningful features from raw vibration signals before passing them to attention and BiLSTM layers. Its residual learning framework enables deep feature extraction while avoiding the vanishing gradient problem. The core unit of a 1D ResNet is the residual block, comprising convolutional layers, batch normalization, and activation functions [34]. Rather than learning a new representation, the residual block models the difference between input and output:

$$H(x) = F(x) + x \quad (6)$$

where x is the input signal, $F(x)$ represents the learned transformation through a series of convolutions, and $H(x)$ is the final output. The presence of the identity mapping x ensures that essential signal characteristics are preserved, preventing excessive modifications that could affect fault-related features in vibration signals. Each block applies multiple 1D convolutions sliding over time to extract local and global patterns. The convolution operation is defined as:

$$y_i = \sum_{k=1}^K w_k \cdot x_{i+k} + b \quad (7)$$

where y_i is at the position i , w_k are learnable weights, K is kernel size, and b is bias. After each convolution operation, batch normalization is applied to normalize the feature distribution and stabilize training. This normalization is computed as:

$$\hat{x} = \frac{x - u}{\sigma} \quad (8)$$

where u is the batch mean and σ is the batch standard deviation. This ensures that the network maintains stable feature distributions across layers, accelerating convergence and improving generalization. The network then applies a ReLU (rectified linear unit) activation function, which introduces nonlinearity into the model, allowing it to learn complex representations. ReLU is defined as:

$$f(x) = \max(0, x) \quad (9)$$

Which ensures that only positive activations are passed forward, preventing issues such as vanishing gradients. The residual connection then adds the original input signal x back to the transformed features $f(x)$, ensuring that important signal characteristics are not lost during deep feature extraction. As the vibration signal passes through the ResNet, early layers extract statistical features while deeper layers extract high-level representations as shown in Figure 2. These features enable precise classification of bearing faults. Mathematically, in a residual network with L layers, the transformation applied to the input signal can be expressed as:

$$H_L(x) = x + \sum_{i=1}^L F_i(x) \quad (10)$$

where $F_i(x)$ represents the feature transformation learned at the i^{th} residual block. The presence of the original input x in the final output ensures that even if some layers contribute only minor refinements, the core structure of the signal remains intact. The 1D ResNet has three convolutional layers, each with batch normalization and ReLU, ensuring multi-scale

signal variation capture while avoiding overfitting. The residual output is passed to a multi-head attention mechanism, which enhances relevant features and suppresses noise. This refinement is essential because vibration signals often contain background interference. Integrating 1D ResNet ensures deep hierarchical feature learning without performance degradation. Its ability to extract time- and frequency-domain features suits real-world, non-stationary vibration signals.

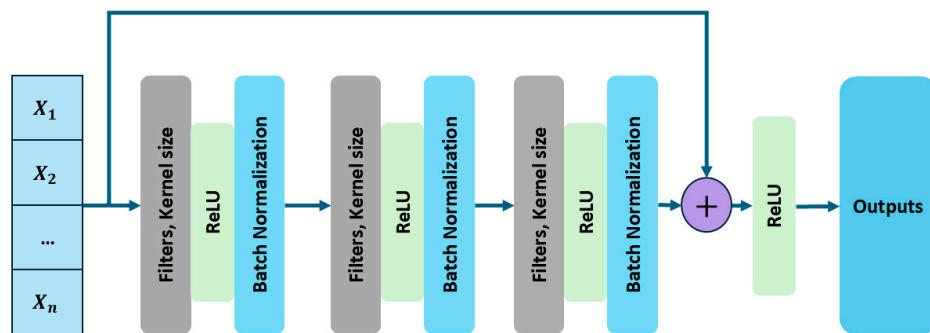


Figure 2. Architecture of the 1D convolutional residual network used for extracting deep temporal-spatial features from CWT-based vibration data.

2.3. Multi-Head Self-Attention Mechanism

The MHSA mechanism is an important component in modern deep learning architectures, especially for sequential data like vibration signals, NLP, and fault diagnosis [35]. In the proposed model, MHSA enhances feature representations from the 1D ResNet by selectively focusing on different parts of the extracted feature maps. It captures long-range dependencies in time-series data, essential for analyzing extended fault-induced patterns. At the core of MHSA is self-attention, which assigns dynamic importance to different positions in the input. Unlike convolutional operations with local receptive fields, attention mechanisms compute global dependencies across all time steps. This is useful in bearing fault diagnosis, where defects create periodic or irregular patterns not always captured by convolution alone. Mathematically, self-attention is computed using three key transformations: the query (Q), key (K), and value (V) matrices. Given an input sequence represented as a matrix X , these matrices are obtained through learned weight matrices:

$$Q = XW^Q, K = XW^K, V = XW^V \quad (11)$$

where W^Q , W^K , and W^V are trainable weight matrices that project the input sequence into different sub-spaces. The attention scores are computed using the scaled dot-product attention formula

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

where d_k is the dimension of the key vectors, and the softmax function ensures that the attention weights sum to one, allowing the model to focus more on important regions. The division by $\sqrt{d_k}$ prevents the dot product values from becoming excessively large. The multi-head attention mechanism extends this concept by performing multiple attention operations in parallel, each with a different set of learned weights. The input is projected into multiple lower-dimensional subspaces, each producing its own attention output:

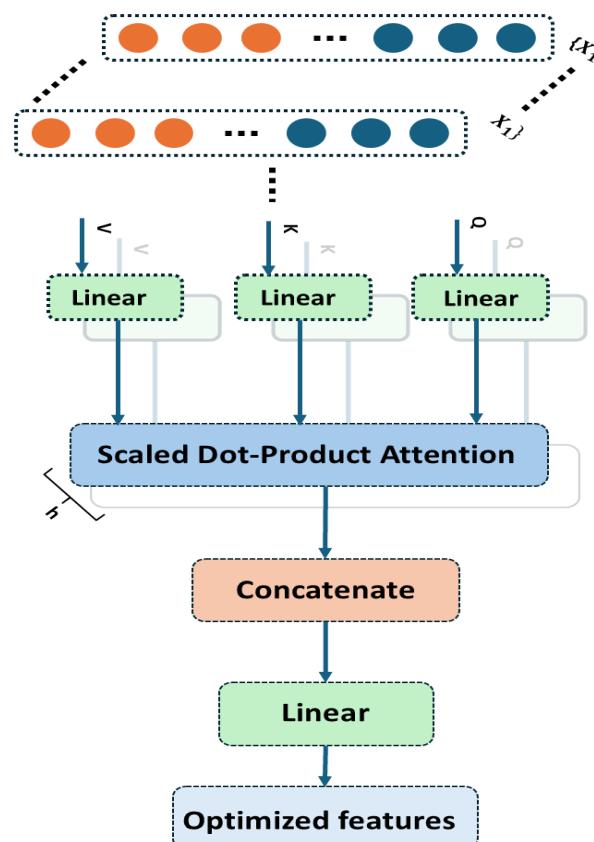
$$\text{MultiHead}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (13)$$

where each attention head is computed independently as

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \quad (14)$$

The final output is obtained by concatenated and linearly transformed by W^O . This allows analysis from multiple perspectives: some heads may attend to high-frequency transients, others to low-frequency trends, creating a richer representation. In bearing fault diagnosis, MHSA refines 1D ResNet features. While ResNet captures local dependencies, it lacks global correlation. Attention is compensated by reweighing feature maps, emphasizing fault-relevant patterns [36].

Figure 3 illustrates the inner workings of the MHSA mechanism. The input features are transformed into Q , K , and V vectors through learned linear projections. Each attention head computes scaled dot-product attention independently, capturing different contextual dependencies. These parallel outputs are concatenated and passed through a final linear transformation, enabling the network to focus on multiple fault-relevant regions in the signal concurrently. By integrating MHSA, the model becomes more robust to variations in load, speed, and operating conditions. Each head captures a distinct representation, improving generalization across fault types. The attention-enhanced features are then passed to the BiLSTM layer, where temporal dependencies are further processed before classification.



classification. Traditional RNNs suffer from vanishing gradients, limiting their ability to learn long sequences. LSTM overcomes this using gating mechanisms that control information flow, retaining relevant features while discarding noise [38].

The core component of an LSTM is the memory cell, which maintains long-term dependencies across time steps. This memory cell is controlled by three main gates: the forget gate f_t , which decides how much of the previous cell state C_{t-1} should be retained or forgotten; the input gate i_t , which determines how much new information \tilde{C}_t should be added to the memory; and the output gate o_t , which controls how much of the updated memory should be passed forward as the hidden state. These gates are computed as follows:

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (15)$$

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \quad (16)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, X_t] + b_C) \quad (17)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (18)$$

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \quad (19)$$

$$h_t = o_t \odot \tanh(C_t) \quad (20)$$

where σ represents the sigmoid activation function, which ensures that values are between 0 and 1, and \tanh introduces nonlinearity to help model complex sequential relationships. The element-wise multiplication \odot allows for selective updates to the memory and hidden states, ensuring that only the most relevant information is retained. Unlike a traditional LSTM, which only processes the input sequence in one direction (from past to future), a BiLSTM extends this by introducing two LSTM networks, one moving forward and the other moving backward through time. This bidirectional structure allows the model to capture dependencies in both directions, making it particularly useful in tasks where context from both past and future time steps is essential. The forward LSTM computes

$$\vec{h}_t = \text{LSTM}\left(X_t, \vec{h}_{t-1}\right) \quad (21)$$

while the backward LSTM processes the sequence in reverse:

$$\overleftarrow{h}_t = \text{LSTM}\left(X_t, \overleftarrow{h}_{t+1}\right) \quad (22)$$

The final hidden state at each time is obtained by concatenating the outputs of both directions, such as

$$h_t = \begin{bmatrix} \vec{h}_t; \overleftarrow{h}_t \end{bmatrix} \quad (23)$$

This bidirectional process significantly improves the model's ability to capture temporal dependencies in vibration signals, particularly in bearing fault detection, where fault-induced patterns may be spread over time rather than occurring at fixed intervals. In the proposed model, BiLSTM follows the 1D ResNet and MHSA mechanism. ResNet extracts hierarchical features, MHSA enhances them, and BiLSTM models short- and long-term dependencies. This improves discrimination between normal and faulty states. BiLSTM effectively handles varying fault classes. As defects evolve, they alter vibration signal patterns—BiLSTM detect these transitions using both past and future context [39]. It also handles noisy data by learning directly from raw signals, filtering out irrelevant noise and focusing on fault signatures. The output is a sequence of representations encoding

both local and long-range fault features. This ensures that the model utilizes both spatial and temporal information, optimizing fault detection accuracy [40]. Mathematically, the FC layer applies a linear transformation followed by a nonlinear activation function. Given an input feature vector h of size d , the fully connected layer computes the output y using the equation,

$$y = W_h + b \quad (24)$$

where W represents the weight matrix, which contains the learnable parameters defining how inputs are mapped to outputs, while b is the bias vector, providing additional flexibility to the transformation. Following this transformation, a nonlinear activation function is applied to enable the network to learn complex mappings. For classification tasks, the final FC layer typically employs the softmax activation function, which converts the raw output scores (logits) into a probability distribution over the classes. The softmax function is defined as

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (25)$$

where z_i represents the raw score for class i , N denotes the total number of classes, and $P(y_i)$ represents the probability of the input belonging to class i . To prevent overfitting and enhance generalization, dropout regularization is often applied within FC layers. Dropout functions by randomly deactivating a fraction of neurons during training, ensuring that the network does not become overly reliant on specific neurons and instead learns more generalized representations. Mathematically, dropout is implemented as

$$h' = D \odot h \quad (26)$$

where D represents a binary mask with values drawn from a Bernoulli distribution, h is the input feature vector before dropout, and h' is the modified feature vector after dropout [41]. The number of fully connected layers in a network varies based on the complexity of the problem. In most classification models, one or two fully connected layers are followed by a final softmax layer, which produces probability distribution over the output classes. The final classification decision is determined by selecting the class with the highest probability, computed as

$$\hat{y} = \text{argmax}P(y_i) \quad (27)$$

where \hat{y} represents the predicted fault class. This structured decision-making process ensures that the model outputs a robust and reliable prediction based on the learned representations as depicted in Figure 4.

Overall, the FC layers in the proposed method function as the final stage of the diagnosis pipeline, transforming deep hierarchical representations into a structured probability distribution over fault classes. It is worth mentioning that the three FC layers serve different roles—the first (1024 units) expands the feature space for deep representation learning, the second (512 units) compresses features to reduce complexity and overfitting, and the final (4 units) maps features to fault classes using softmax activation for classification. The proposed model architecture summary and training parameter details are summarized in the following Tables 1 and 2, respectively.

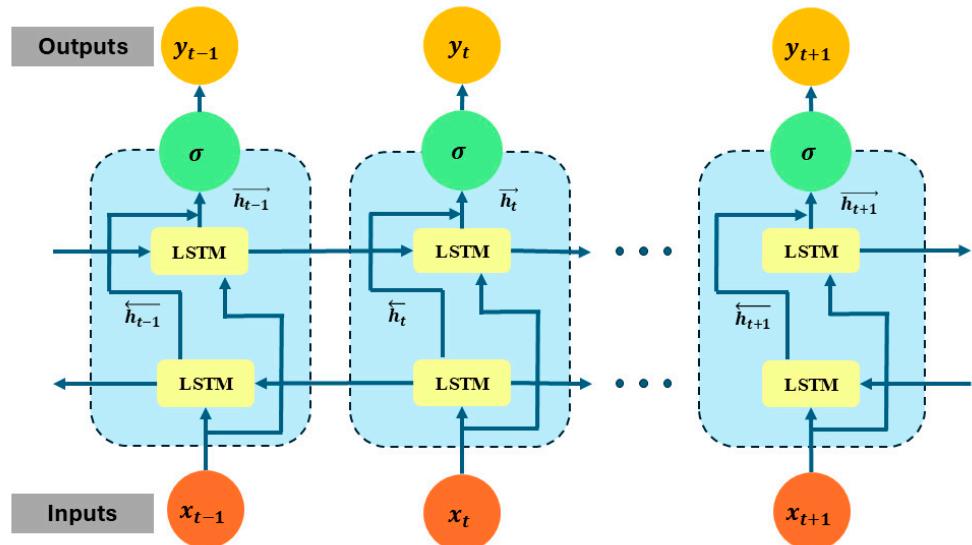


Figure 4. Structure of the bidirectional LSTM module for learning forward and backward temporal patterns in CWT-transform vibration signals.

Table 1. Layer-wise architecture of the proposed hybrid model.

Layer Type	Parameters	Activation	Output Shape
Conv2D	(1, 32, 3 × 3)	ReLU	(32, 128, 128)
BatchNorm2D	32	-	(32, 128, 128)
Conv2D	(32, 64, 3 × 3)	ReLU	(64, 128, 128)
BatchNorm2D	64	-	(64, 128, 128)
AdaptiveAvgPool2D	(16, 16)	-	(64, 16, 16)
Flatten		-	(256, 64)
Multi-head attention	4 heads (64 dim)	-	(256, 64)
BiLSTM (2 layers)	Hidden size: 128	-	(256, 256)
Fully connected	1024	ReLU	1024
Fully connected	512	ReLU	512
Fully connected	4	softmax	4

Table 2. Key training parameters used for optimizing the proposed deep learning model.

Parameter	Value
Optimizer	Adam
Learning rate	0.0005
Scheduler	StepLR (step_size = 10, gamma = 0.5)
Loss function	CrossEntropyLoss
Batch size	32
Epochs	40

3. Results and Performance Evaluation

The effectiveness of the proposed method is assessed using VS data obtained from actual bearing testbeds. Since the primary aim of this method is to detect and diagnose faults in bearings, this section begins by comparing the fault detection capability of the proposed method.

3.1. Experimental Setup and Data Acquisition

Figure 5 depicts the structure of the bearing testbed developed by the Ulsan Industrial Artificial Intelligence (UIAI) Laboratory at Ulsan University, Ulsan, Republic of Korea, which was utilized for this analysis. The collected data were categorized into four distinct bearing conditions: normal, outer race damage, inner race damage, and roller damage. During the experiment, a three-phase motor operated the testbed at a constant speed of 1800 rpm. The rotational motion was transferred from the rotor shaft to the main shaft via a belt system installed on both sides of the test bearings. The bearings used in this study were FAG NJ206-E-TVP2 cylindrical roller bearings made of chrome steel (AISI 52100), subjected to a constant vertical load of 100 kgf. Artificial defects were introduced with dimensions of approximately $3\text{ mm} \times 0.3\text{ mm} \times 1\text{ mm}$ to simulate real fault conditions. To ensure precise data acquisition, the maximum vibration signal was recorded from the left side of the target bearing using both a vibration sensor and an acoustic emission (AE) accelerometer. A schematic representation of the complete experimental setup is provided in Figure 6, offering a detailed overview of the testbed configuration and data acquisition process.

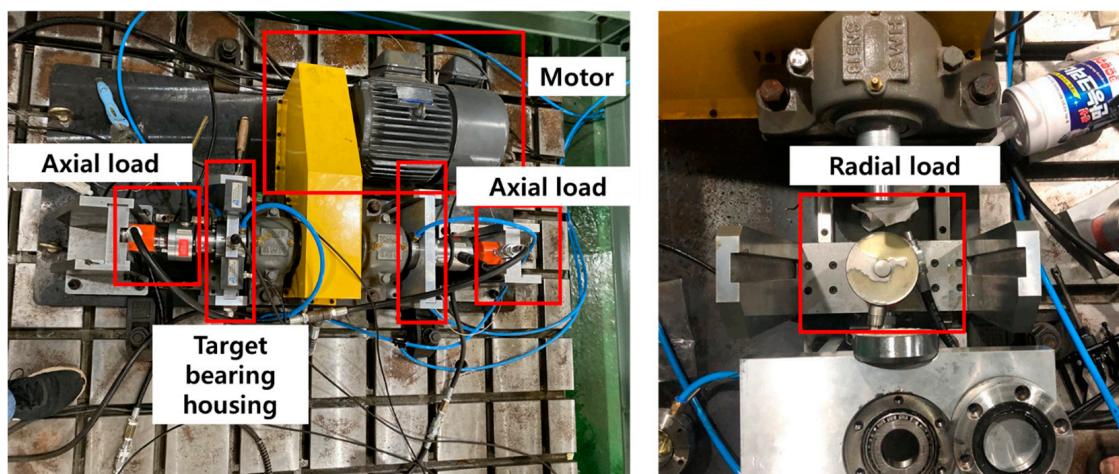


Figure 5. Experimental setup for acquiring vibration signals under various fault conditions in a controlled laboratory environment.

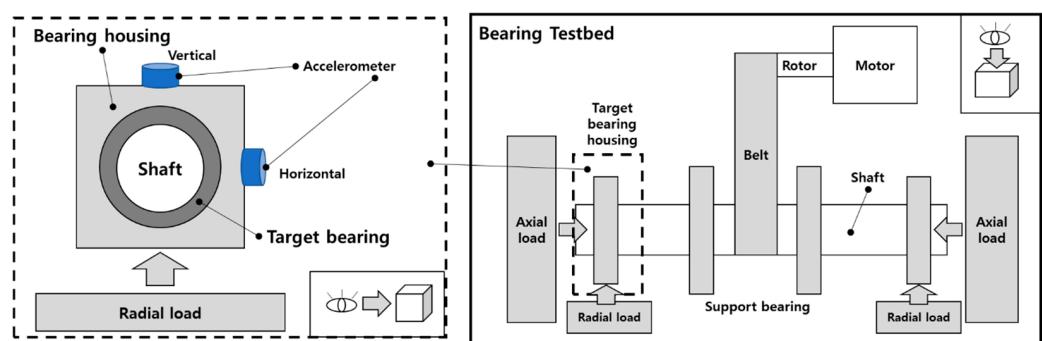


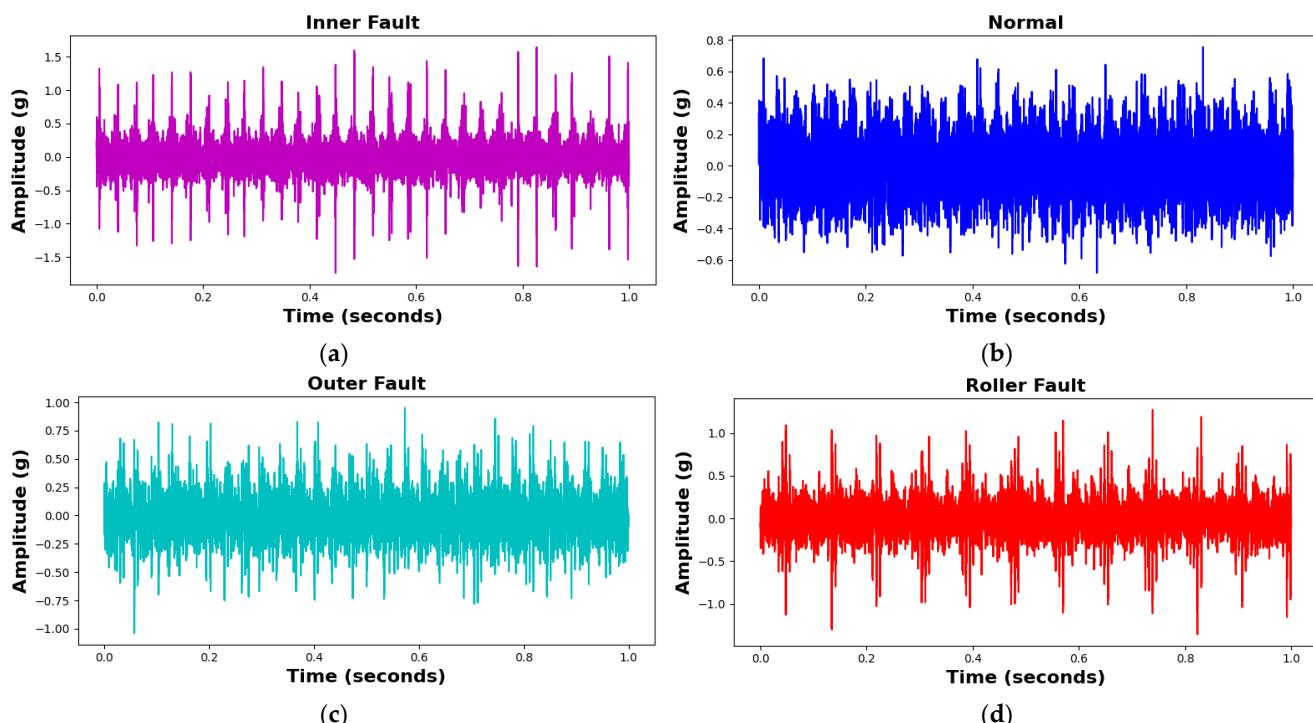
Figure 6. Schematic diagram of experimental setup for bearing fault diagnosis.

The data acquisition system, detailed in Table 3, utilizes a FAG NJ206-3-TVP2 bearing, a cylindrical roller type. For capturing VS, an accelerometer (model PCB-622B01) was employed, while AE signals were detected using R15I-AST AE sensors. Both sensors were interfaced with an NI-9234 data acquisition (DAQ) device to ensure precise data collection from the integrated electronics piezoelectric (IEPE) sensors.

Table 3. Specifications of the data acquisition system.

Device	Specification	Value
Vibration sensor (PCB-622B01)	Measurement range	$\pm 490 \text{ m/s}^2$
	Frequency	0.2–15,000 Hz
	Sensitivity	100 mV/g
AE sensor (R151-AST)	Operating range	50–400 kHz
	Resonant frequency	150 kHz
	Peak sensitivity	–22 dB
DAQ (NI 9234)	Dynamic range	102 dB
	Resolution	24-bit
	Operating temperature	–40 °C–70 °C

Vibration data were acquired at a sampling rate of 25 kHz, with five minutes of continuous data collection conducted for all bearing conditions. Subsequently, the data were divided into 1 s segments, with each 1 s segment containing 309–390 data samples for different fault types. The test procedure can be repeated for different fault types by replacing the test bearing in the same testbed setup. The samples of each class of the bearing fault as well as its vibration data are shown in the following Figures 7 and 8, followed by the dataset details in Table 4.

**Figure 7.** Vibration time domain signals for different bearing fault conditions: (a) IRF, (b) NC, (c) ORF, and (d) RF.**Table 4.** Data acquisition summary.

Testing Condition	Samples Count	Sampling Rate (KHz)	Time (min)
IRF	370	25	5
NC	344	25	5
ORF	347	25	5
RF	309	25	5

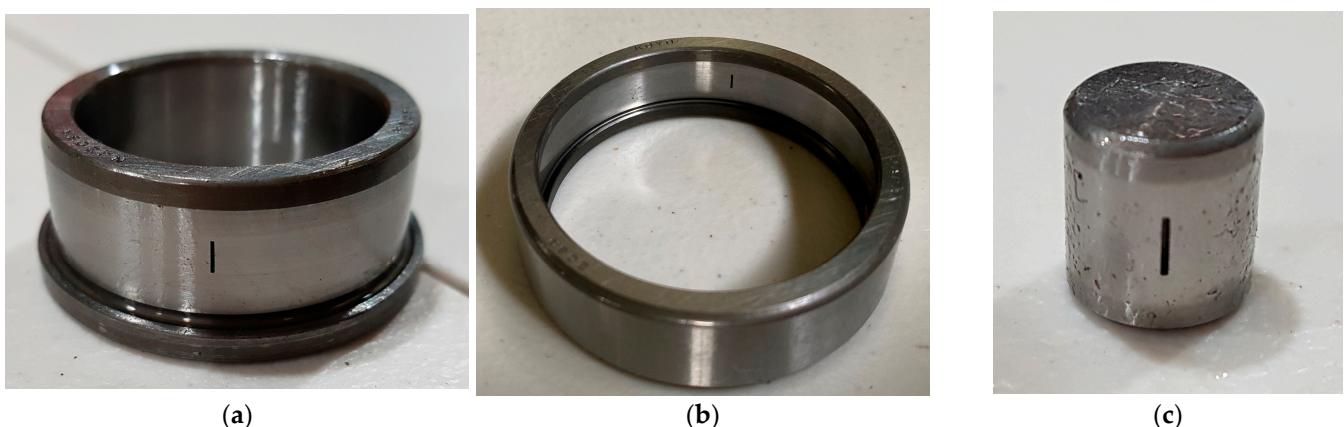


Figure 8. Fault components used during the experiment: (a) IRF, (b) ORF, and (c) RF.

3.2. Paderborn University Dataset

The experimental test rig developed by Lessmeier et al. [42] for vibration data acquisition consisted of several key components, as illustrated in Figure 9. Designed as a versatile platform, it enabled the testing of bearings under various fault conditions, allowing for precise fault characterization and analysis. At the core of the setup, the bearing test module facilitated controlled testing by applying a constant radial load to the test bearings, adjustable up to 10 kN before each experiment. This ensured consistency and repeatability across different testing conditions. The system was powered by a 425 W permanent magnet synchronous motor (PMSM) with a nominal torque of 1.35 Nm, a rotational speed of 3000 rpm, a nominal current of 2.3 A, and four pole pairs. Manufactured by Hanning Elektro-Werke GmbH & Co. KG (Model: SD4CDu8S009), the motor was controlled using an industrial inverter (KEB Combivert 07F5E 1D-2B0A, Oerlinghausen, Germany) with a 16 kHz switching frequency, effectively replicating real industrial environments, including PWM-induced noise typically found in such systems. To capture vibration signals, a piezoelectric accelerometer (Model 336C04, PCB Piezotronics, Inc., Depew, NY, USA) was securely mounted on the bearing housing adapter. The signals were amplified using a Type 5015A charge amplifier (Kistler Group, Winterthur, Switzerland) and filtered through a 30 kHz low-pass filter, which removed high-frequency noise. The filtered signals were then digitized at a 64 kHz sampling rate, ensuring high-resolution data acquisition for precise fault analysis. To mimic real-world operational conditions, a flywheel and load motor were incorporated into the setup, simulating dynamic load variations and system inertia. The experiments were conducted under four distinct operating conditions, varying in rotational speed, radial force, and load torque, as outlined in Table 5.

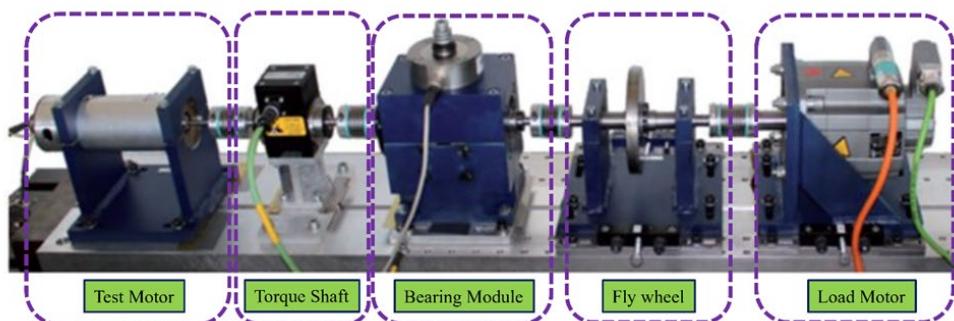


Figure 9. Experimental setup of the Paderborn bearing dataset used for validating the model on real and artificially induced fault scenarios.

Table 5. Operating conditions.

S. No	Rotational Speed (rpm)	Load Torque (Nm)	Radial Force (N)
0	1500	0.7	1000
1	900	0.7	1000
2	1500	0.1	1000
3	1500	0.7	400

The study utilized 32 bearings, categorized into three groups: 12 bearings with artificially induced faults, 14 bearings subjected to accelerated lifetime testing to develop real-time faults, and 6 healthy bearings serving as baseline references.

3.2.1. Case 1: Artificial Induced Faults

Lessmeier et al. [42] induced artificial damage in bearings using three distinct methods, each designed to replicate common fault characteristics observed in industrial environments. The first method, electric discharge machining (EDM), was employed to create precise trenches approximately 0.25 mm in length along the rolling direction, with depths ranging from 1 mm to 2 mm. This technique ensures highly controlled and repeatable damage patterns, making it ideal for maintaining consistent experimental conditions. The second method involved drilling, which was used to introduce faults of varying diameters, simulating damage caused by localized stress concentrations or surface abrasions. This approach helps replicate real-world wear and tear commonly found in industrial applications. The third method utilized electric engraving to create faults with damage lengths ranging from 1 mm to 4 mm. This technique produces irregular damage profiles, mimicking defects that arise due to material fatigue or improper handling. By employing these three damage-inducing techniques, the dataset captures a diverse range of fault characteristics, ensuring a comprehensive evaluation of the proposed fault detection framework. The bearing codes used in this study for Case 1, which are publicly available, are detailed in Table 6.

Table 6. Details of bearing codes used in Case 1.

Bearing State	Bearing Code
Normal	K001, K002, K003, K004, K005, K006
Artificial IR fault	KI01, KI03, KI05, KI07, KI08
Artificial OR fault	KA01, KA03, KA05, KA06, KA07, KA09

3.2.2. Case 2: Real Bearing Faults

Lessmeier et al. [42] generated real damaged ball bearings using an accelerated lifetime test rig. The test bearings were subjected to a radial load applied through a spring-screw mechanism. To accelerate the formation of fatigue damage, the load was deliberately set higher than standard operational conditions while remaining below the bearings' static load capacity to prevent immediate failure. Additionally, low-viscosity oil was introduced to simulate poor lubrication conditions, further promoting the gradual development of bearing damage. The accelerated testing process produced several damaged bearings, with approximately 70% exhibiting fatigue-induced pitting on both the inner race (IR) and outer race (OR). Among the remaining bearings, most displayed plastic deformations in the form of indentations caused by debris, primarily manifesting as OR faults. One bearing experienced a complete fracture, while no damage was observed to the rolling elements. This controlled approach to accelerated lifetime testing ensured the generation of realistic and diverse fault scenarios, allowing the collected data to closely reflect real-world

failure mechanisms. This makes the dataset valuable for robust fault detection research. The bearing codes used in this study for Case 2, which are publicly available, are listed in Table 7.

Table 7. Details of bearing codes used in Case 2.

Bearing State	Bearing Code
Normal	K001, K002, K003, K004, K005, K006
Real IR fault	KI04, KI14, KI16, KI17, KI18, KI21
Real OR fault	KA04, KA15, KA16, KA22, KA30

3.3. Performance Metrics for Comparisons

The evaluation of the proposed method was conducted using key performance metrics, including accuracy, precision, recall, and F1-score, as outlined in Equations (28)–(31). These metrics were essential in measuring the model's ability to accurately classify different fault conditions in the milling machine. By leveraging this comprehensive evaluation approach, the proposed model was thoroughly assessed for its capability in real-world fault detection and diagnosis, ensuring its suitability for industrial applications. The proposed model was validated using stratified k-fold cross validation ($k = 5$) to avoid overfitting. This thorough evaluation confirmed the model's reliability and effectiveness in practical fault classification tasks.

$$\text{Accuracy} = \frac{(TN + TP)}{(TP + TN + FP + FN)} \times 100\% \quad (28)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (29)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (30)$$

$$\text{F1 - Score} = \frac{2TP}{2TP + FP + FN} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (31)$$

True positives (TPs) indicate instances where the model correctly identifies faulty conditions, while true negatives (TNs) refer to non-faulty cases that are accurately recognized. False positives (FPs) occur when the model incorrectly labels non-faulty instances as faulty, whereas false negatives (FNs) represent faulty cases that are misclassified as non-faulty. Together, these metrics provide a comprehensive evaluation of the model's classification performance, offering valuable insights into its robustness, accuracy, and reliability in detecting and diagnosing bearing faults.

3.4. Comparative Analysis of Fault Diagnosis Methods

To comprehensively evaluate the performance and generalization capability of the proposed model, it was tested on the UIAI Lab dataset as well as both the real and artificial subsets of the Paderborn dataset. Each of the datasets was divided into 80% for training and 20% for testing purposes. Furthermore, the UIAI Lab dataset was also used to validate the recent model proposed by Guanghua Fu et al. [43] for a direct performance comparison, and an ablation study of the proposed model was carried out. The proposed hybrid deep learning model combines CWT with an advanced classification architecture comprising 1D conv ResNet, MHSA, and BiLSTM. Initially, vibration signals are transformed into time-frequency representations using CWT, effectively capturing nonstationary and nonlinear signal characteristics across multiple scales. These CWT scalograms are then processed through the hybrid deep learning network. The proposed hybrid deep learning model com-

bines CWT with a powerful classification framework that integrates MHSA, BiLSTM, and 1D conv ResNet. Initially, vibration signals are transformed into time-frequency representations using CWT, effectively capturing nonstationary and nonlinear signal characteristics across multiple scales. These CWT scalograms are then processed through the hybrid architecture, where each component contributes to robust and interpretable feature learning: 1D conv ResNet extracts localized features while ensuring efficient gradient propagation, BiLSTM captures long-term temporal dependencies important for fault progression tracking, and MHSA enhances the model's focus on the most informative segments of the signal. This enables the model to remain resilient in noisy industrial environments and effectively distinguish between fault types. The model was trained on a bearing dataset encompassing four classes—IRF, ORF, RF, and NC. Training and validation results demonstrated the model's rapid convergence and stability. Both accuracy curves quickly approached higher accuracy after a dozen epochs, while training and validation losses showed a sharp decline and stabilize near zero—indicating highly effective learning without overfitting, as shown in Figure 10. On internal testing, the model achieved perfect classification accuracy across all four fault categories, as reflected in the confusion matrix in Figure 11: 85/85 for IRF, 74/74 for ORF, 55/55 for RF, and 60/60 for NC. Precision, recall, and F1-scores for all classes reached 1.00, further emphasizing the model's diagnostic precision, as shown in Table 8. The t-SNE visualization revealed clearly clustered and non-overlapping distributions for each class, showcasing the model's ability to learn well-separated, discriminative features as evident from Figure 12. Additionally, ROC curves for each class yielded an AUC of 1.00, confirming flawless class separation and zero false positives or false negatives. As shown in Table 9, the class-wise metrics further validate these findings, with all fault types achieving a true positive rate (TPR) of 1.00 and minimal or zero values for false positive rate (FPR), false negative rate (FNR), and false discovery rate (FDR), except for a slight drop in normal class due to minor misclassifications—demonstrating the model's reliability and precision across fault categories.

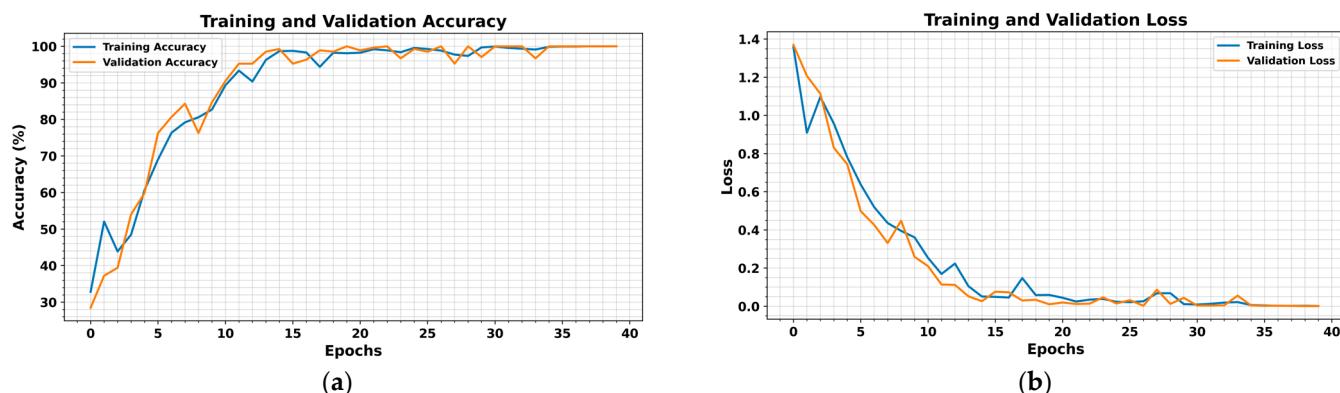


Figure 10. Proposed model training and validation. (a) Accuracies vs number of epochs and (b) losses vs number of epochs.

These confidence interval (CI) values reflect a high-performing model with minimal variance, making them statistically realistic and better for bearing fault diagnosis [43]. To evaluate generalization and real-world applicability, the model was further tested on the Paderborn bearing dataset, which includes both real-world and artificially induced fault conditions. On the real dataset, the model achieved an overall accuracy of $97.98 \pm 0.18\%$, with precision, recall, and F1-score all at $98.33 \pm 0.16\%$, as mentioned in Table 9. Class-wise performance remained high, with perfect classification for normal and OR real samples, and minor misclassifications between IR real and OR real classes. The confusion matrix confirmed 739/774 correct predictions for IR real, 395/395 for normal, and 615/616 for

OR real, as shown in Figure 11. As shown in Figure 12, the t-SNE visualizations indicated clear class separation in all three datasets, with minor overlaps observed in the Paderborn real set due to naturally developed faults. Similarly, the ROC curves in Figure 13 reflect strong classification capability, with high AUC values supporting the model's effectiveness in distinguishing fault types across diverse conditions.

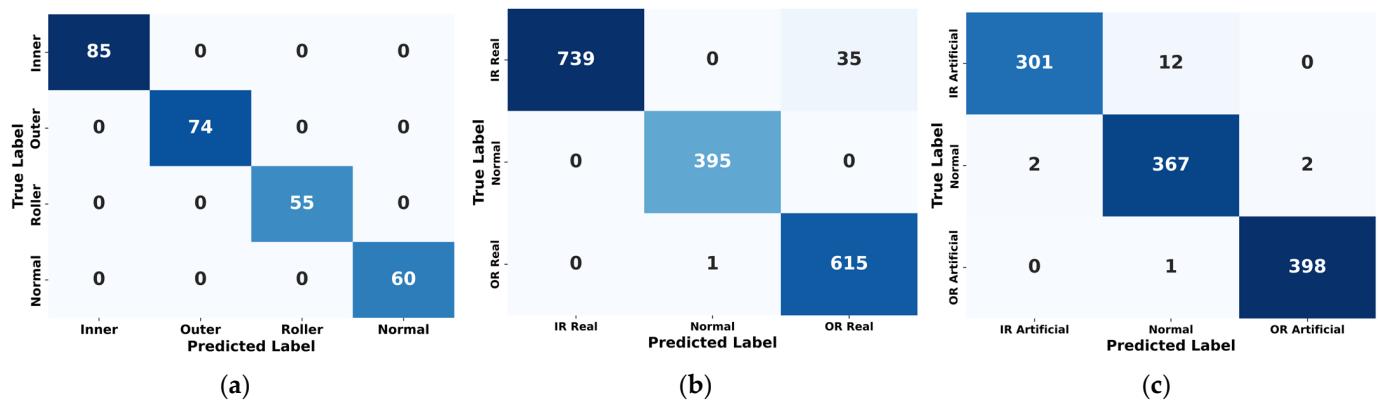


Figure 11. Confusion matrices of the proposed model on (a) UIAI, (b) Paderborn real, and (c) Paderborn artificial datasets, showing classification accuracy across fault classes.

Table 8. Proposed model evaluation of performance metrics on diverse datasets and comparison with Fu et al. and ablation studies.

Performance Metrics		Proposed		Fu et al [44].		Ablation Study (UIAI Dataset)		
Datasets	UIAI	Real Paderborn	Artificial Paderborn	UIAI	No BiLSTM	No MHSA	No Conv1D	Conv1D Only
Accuracy (%)	99.27 ± 0.12	97.98 ± 0.18	98.71 ± 0.15	92.02 ± 0.20	82.12 ± 0.22	86.50 ± 0.19	95.05 ± 0.14	63.68 ± 0.16
Precision (%)	99.25 ± 0.11	98.33 ± 0.16	98.67 ± 0.14	92.50 ± 0.21	83.60 ± 0.21	87.05 ± 0.17	96.96 ± 0.13	62.55 ± 0.18
F1 Score (%)	99.25 ± 0.10	98.33 ± 0.15	98.67 ± 0.13	92.50 ± 0.22	83.09 ± 0.22	87.07 ± 0.18	96.91 ± 0.13	62.43 ± 0.20
Recall (%)	99.25 ± 0.13	98.33 ± 0.17	98.67 ± 0.12	92.50 ± 0.23	82.59 ± 0.20	87.01 ± 0.19	96.87 ± 0.15	62.62 ± 0.17

Table 9. Class-wise TPR, FPR, FNR, and FDR showing the model's accuracy and error distribution across fault types.

Class	True Positive Rate (TPR)	False Positive Rate (FPR)	False Negative Rate (FNR)	False Discovery Rate (FDR)
Inner	1.0000	0.0099	0.0000	0.0274
Outer	1.0000	0.0000	0.0000	0.0000
Roller	1.0000	0.0000	0.0000	0.0000
Normal	0.9730	0.0000	0.0270	0.0000

On the artificial Paderborn dataset, the model achieved an even higher accuracy of $98.71 \pm 0.15\%$, with precision, recall, and F1-score values all at $98.67 \pm \sim 0.13\%$. Only a few misclassifications were observed between IR artificial and normal samples, while OR artificial was classified with 100% accuracy. The ROC curves for all artificial classes once again demonstrated an AUC of 1.00, and t-SNE visualizations in Figures 12 and 13, respectively, revealed distinct, non-overlapping class clusters, further validating the model's robustness and reliability.

An ablation study was conducted to assess the individual contributions of core components in the proposed model using the UIAI lab dataset. As shown in Table 9,

four configurations were tested: no BiLSTM (CWT + Conv1D + MHSA), no MHSA (CWT + Conv1D + BiLSTM), no Conv1D (CWT + MHSA + BiLSTM), and Conv1D only (CWT + Conv1D). The results indicate that excluding BiLSTM led to an average accuracy of $82.12 \pm 0.22\%$, precision of $83.60 \pm 0.21\%$, recall of $82.59 \pm 0.20\%$, and F1-score of $83.09 \pm 0.22\%$. Removing MHSA yielded improved results with $86.50 \pm 0.19\%$ accuracy, $87.05 \pm 0.17\%$ precision, $87.01 \pm 0.19\%$ recall, and $87.07 \pm 0.18\%$ F1-score. The best performance among ablation variants was achieved when Conv1D was removed, resulting in $95.05 \pm 0.14\%$ accuracy, $96.96 \pm 0.13\%$ precision, $96.87 \pm 0.15\%$ recall, and $96.91 \pm 0.13\%$ F1-score. In contrast, the Conv1D Only configuration produced the poorest performance with $63.68 \pm 0.16\%$ accuracy, $62.55 \pm 0.18\%$ precision, $62.62 \pm 0.17\%$ recall, and $62.43 \pm 0.20\%$ F1-score, highlighting the critical contributions of both temporal modeling and attention mechanisms in the proposed framework.

In comparison to recent state-of-the-art models, the proposed method demonstrates notable improvements in both accuracy and robustness. For instance, the model developed by Guanghua Fu et al. [44], which integrates a CNN with BiLSTM and a residual module, employs a dual-path feature extraction strategy. In this framework, BiLSTM captures temporal characteristics from vibration signals, while CNN processes spatial features from time-frequency representations. These complementary features are fused and further refined by a residual module to enhance robustness, particularly under noisy conditions. When this model was applied to the same UIAI Lab dataset, it achieved an overall accuracy of $92.02 \pm 0.20\%$, with precision, recall, and F1-score values of $92.50 \pm \sim 0.22\%$, respectively, as shown in Table 9. While it outperformed basic CNN and BiLSTM models individually, it still exhibited several misclassifications across fault classes. In contrast, the proposed model not only surpassed Fu et al.'s architecture in all evaluation metrics but also maintained perfect or near-perfect class separation, illustrating its superior capability in handling complex, real-world vibration signals.

Overall, these results highlight the strength of the proposed framework in delivering high-precision and reliable bearing fault diagnosis. Its consistent performance across both controlled and real-world datasets confirms its strong generalization ability and practical relevance for industrial scenarios. The ablation analysis further validates the critical contribution of each architectural component, demonstrating a well-balanced and optimized design. To assess real-time suitability, the model's latency and complexity were evaluated. It achieved an average inference time (note that the reported inference time refers only to the model's forward pass and does not include preprocessing steps such as CWT computation) of 0.19 ms per sample measured over 100 runs on an NVIDIA RTX 3060 GPU and contained only 262,308 trainable parameters, with a model size under 1 MB. These results confirm its feasibility for real-time deployment on edge or embedded industrial systems. Moreover, the model effectively tackled key challenges such as nonstationary signal characteristics, class imbalance, and noise interference, making it a robust and scalable candidate for deployment in real-time predictive maintenance and intelligent health monitoring systems.

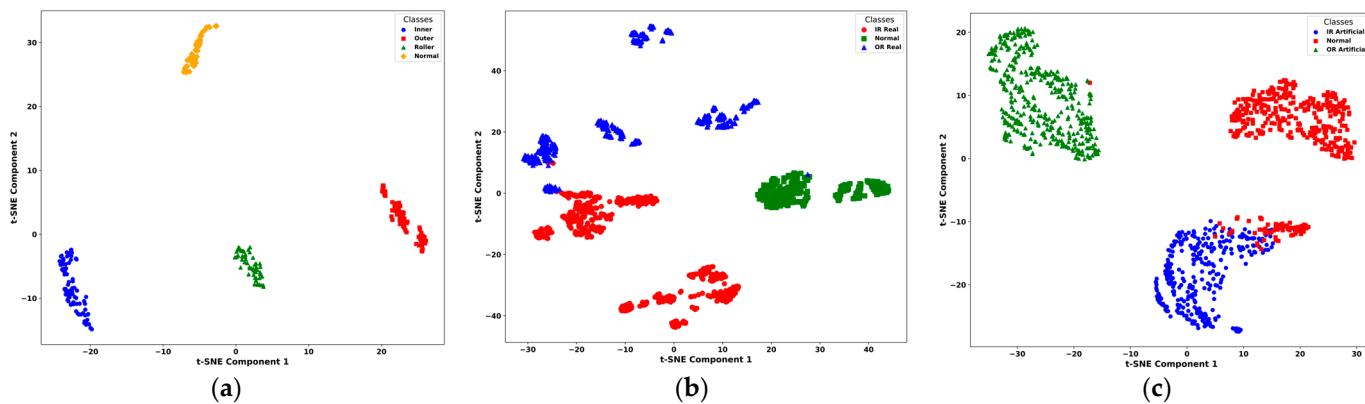


Figure 12. t-SNE visualizations of learned features on (a) UIAI, (b) Paderborn real, and (c) Paderborn artificial datasets, showing clear fault class separation.

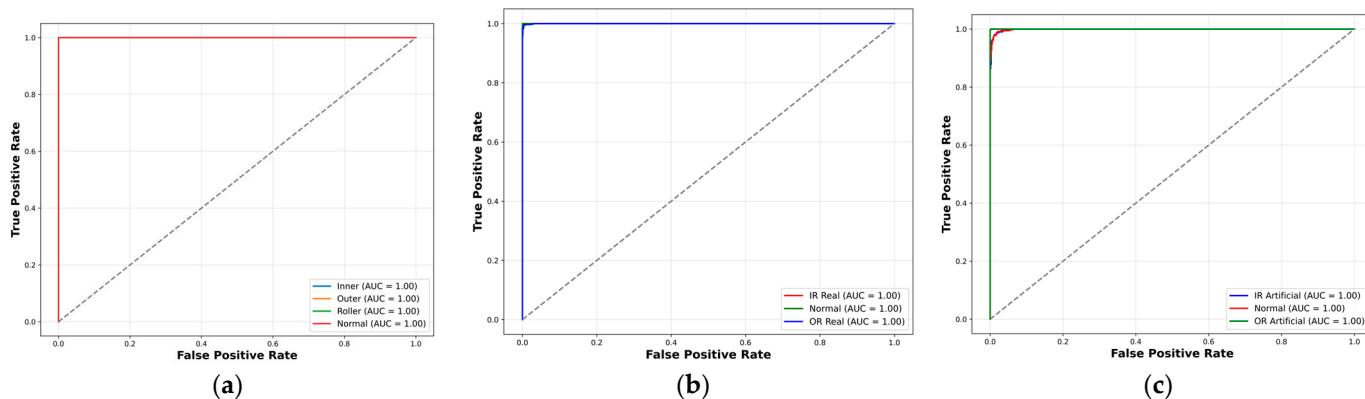


Figure 13. ROC curves of the proposed model on (a) UIAI, (b) Paderborn real, and (c) Paderborn artificial datasets, indicating strong class-wise discrimination performance.

4. Conclusions

This study presents a hybrid deep learning approach for bearing fault diagnosis that integrates continuous wavelet transform (CWT) with an attention-enhanced spatiotemporal feature extraction framework. The model combines multi-head self-attention (MHSA), bidirectional long-short-term memory (BiLSTM), and a 1D convolutional residual network (1D conv ResNet) to effectively capture both spatial and temporal features from nonstationary and nonlinear vibration signals. By leveraging CWT-based time-frequency representations and the complementary strengths of attention and sequential modeling, the approach enabled robust and highly accurate fault classification. Evaluated on the UIAI Lab dataset as well as the real and artificial subsets of the Paderborn dataset, the model consistently achieved near-perfect performance, with precision, recall, and F1-scores close to 1.00. It also demonstrated excellent fault separability in both t-SNE and ROC analyses. The model's strong generalization capability and resilience to noise affirm its practical suitability for real-time predictive maintenance in industrial environments.

Despite its high accuracy, the model shows slight limitations in distinguishing overlapping fault types. Future work could focus on enhancing class separability using advanced attention mechanisms or multi-scale feature fusion. Additionally, deploying models in real-time environments via edge computing could further improve its practical applicability in dynamic industrial settings.

Author Contributions: Conceptualization, M.F.S., F.S., M.U. and J.-M.K.; methodology, M.F.S., F.S., M.U. and J.-M.K.; validation, M.F.S., F.S. and J.-M.K.; formal analysis, F.S., C.H.K. and J.-M.K.; resources, J.-M.K.; writing—original draft preparation, M.F.S., F.S., M.U. and F.S.; writing—review and editing, J.-M.K.; visualization, M.F.S., F.S., M.U., C.H.K. and J.-M.K.; project administration, J.-M.K.; funding acquisition, J.-M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by 2025 Research Fund of University of Ulsan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

Nomenclature

BFD	bearing fault diagnosis
CWT	continuous wavelet transform
CNN	convolutional neural network
BiLSTM	bidirectional long short-term memory
ReLU	rectified linear unit
TPR	true positive rate
ANN	artificial neural network
UIAI	Ulsan Industrial Artificial Intelligence

References

1. Sohaib, M.; Kim, J.-M. Fault Diagnosis of Rotary Machine Bearings Under Inconsistent Working Conditions. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 3334–3347. [[CrossRef](#)]
2. Rai, A.; Upadhyay, S.H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribol. Int.* **2016**, *96*, 289–306. [[CrossRef](#)]
3. Cui, M.; Wang, Y.; Lin, X.; Zhong, M. Fault Diagnosis of Rolling Bearings Based on an Improved Stack Autoencoder and Support Vector Machine. *IEEE Sens. J.* **2021**, *21*, 4927–4937. [[CrossRef](#)]
4. Sun, W.; Cao, X. Curvature enhanced bearing fault diagnosis method using 2D vibration signal. *J. Mech. Sci. Technol.* **2020**, *34*, 2257–2266. [[CrossRef](#)]
5. Siddique, M.F.; Zaman, W.; Ullah, S.; Umar, M.; Saleem, F.; Shon, D.; Yoon, T.H.; Yoo, D.-S.; Kim, J.-M. Advanced Bearing-Fault Diagnosis and Classification Using Mel-Scalograms and FOX-Optimized ANN. *Sensors* **2024**, *24*, 7303. [[CrossRef](#)]
6. Samanta, B.; Al-Balushi, K.R. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328. [[CrossRef](#)]
7. Liu, R.; Yang, B.; Zhang, X.; Wang, S.; Chen, X. Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *75*, 345–370. [[CrossRef](#)]
8. Verstraete, D.; Ferrada, A.; Drogue, E.L.; Meruane, V.; Modarres, M. Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings. *Shock. Vib.* **2017**, *2017*, 5067651. [[CrossRef](#)]
9. Kumar, P.; Raouf, I.; Song, J.; Prince; Kim, H.S. Multi-size wide kernel convolutional neural network for bearing fault diagnosis. *Adv. Eng. Softw.* **2024**, *198*, 103799. [[CrossRef](#)]
10. Vashishtha, G.; Chauhan, S.; Sehri, M.; Hebda-Sobkowicz, J.; Zimroz, R.; Dumond, P.; Kumar, R. Advancing machine fault diagnosis: A detailed examination of convolutional neural networks. *Meas. Sci. Technol.* **2025**, *36*, 022001. [[CrossRef](#)]
11. Jiang, Z.; Zhang, K.; Xiang, L.; Yu, G.; Xu, Y. A time-frequency spectral amplitude modulation method and its applications in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2023**, *185*, 109832. [[CrossRef](#)]
12. Civera, M.; Surace, C. An Application of Instantaneous Spectral Entropy for the Condition Monitoring of Wind Turbines. *Appl. Sci.* **2022**, *12*, 1059. [[CrossRef](#)]
13. Zaman, W.; Siddique, M.F.; Kim, J.-M. Centrifugal Pump Fault Detection with Hybrid Feature Pool and Deep Learning. In Proceedings of the 2023 20th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Murree, Pakistan, 22–25 August 2023; pp. 1–6. [[CrossRef](#)]

14. Saleem, F.; Ahmad, Z.; Siddique, M.F.; Umar, M.; Kim, J.-M. Acoustic Emission-Based Pipeline Leak Detection and Size Identification Using a Customized One-Dimensional DenseNet. *Sensors* **2025**, *25*, 1112. [[CrossRef](#)]
15. He, M.; He, D. Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065. [[CrossRef](#)]
16. Xue, Y.; Wen, C.; Wang, Z.; Liu, W.; Chen, G. A novel framework for motor bearing fault diagnosis based on multi-transformation domain and multi-source data. *Knowl. Based Syst.* **2024**, *283*, 111205. [[CrossRef](#)]
17. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [[CrossRef](#)]
18. Gao, X.; Wei, H.; Li, T.; Yang, G. A rolling bearing fault diagnosis method based on LSSVM. *Adv. Mech. Eng.* **2020**, *12*, 1687814019899561. [[CrossRef](#)]
19. Zhang, Q.; Deng, L. An Intelligent Fault Diagnosis Method of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network. *J. Fail. Anal. Prev.* **2023**, *23*, 795–811. [[CrossRef](#)]
20. Zhao, B.; Zhang, X.; Zhan, Z.; Pang, S. Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable working conditions and domains. *Neurocomputing* **2020**, *407*, 24–38. [[CrossRef](#)]
21. Dao, F.; Zeng, Y.; Qian, J. Fault diagnosis of hydro-turbine via the incorporation of bayesian algorithm optimized CNN-LSTM neural network. *Energy* **2024**, *290*, 130326. [[CrossRef](#)]
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
23. Zhang, K.; Tang, B.; Deng, L.; Liu, X. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox. *Measurement* **2021**, *179*, 109491. [[CrossRef](#)]
24. Liang, P.; Wang, W.; Yuan, X.; Liu, S.; Zhang, L.; Cheng, Y. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105269. [[CrossRef](#)]
25. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Pecht, M. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4681–4690. [[CrossRef](#)]
26. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)] [[PubMed](#)]
27. Wang, J.; Liao, N.; Du, X.; Chen, Q.; Wei, B. A semi-supervised approach for the integration of multi-omics data based on transformer multi-head self-attention mechanism and graph convolutional networks. *BMC Genom.* **2024**, *25*, 86. [[CrossRef](#)]
28. Manhertz, G.; Bereczky, A. STFT spectrogram based hybrid evaluation method for rotating machine transient vibration analysis. *Mech. Syst. Signal Process.* **2021**, *154*, 107583. [[CrossRef](#)]
29. Gou, L.; Li, H.; Zheng, H.; Li, H.; Pei, X. Aeroengine Control System Sensor Fault Diagnosis Based on CWT and CNN. *Math. Probl. Eng.* **2020**, *2020*, 5357146. [[CrossRef](#)]
30. Siddique, M.F.; Ahmad, Z.; Ullah, N.; Ullah, S.; Kim, J.-M. Pipeline Leak Detection: A Comprehensive Deep Learning Model Using CWT Image Analysis and an Optimized DBN-GA-LSSVM Framework. *Sensors* **2024**, *24*, 4009. [[CrossRef](#)]
31. Xu, B.; Wang, X.; Zhang, J.; Guo, Y.; Razzaqi, A.A. A Novel Adaptive Filtering for Cooperative Localization Under Compass Failure and Non-Gaussian Noise. *IEEE Trans. Veh. Technol.* **2022**, *71*, 3737–3749. [[CrossRef](#)]
32. Alabied, S.; Haba, U.; Daraz, A.; Gu, F.; Ball, A.D. Empirical Mode Decomposition of Motor Current Signatures for Centrifugal Pump Diagnostics. In Proceedings of the 2018 24th International Conference on Automation and Computing (ICAC), Newcastle Upon Tyne, UK, 6–7 September 2018; pp. 1–6. [[CrossRef](#)]
33. ALTobi, M.A.S.; Bevan, G.; Wallace, P.; Harrison, D.; Ramachandran, K.P. Fault diagnosis of a centrifugal pump using MLP-GABP and SVM with CWT. *Eng. Sci. Technol. Int. J.* **2019**, *22*, 854–861. [[CrossRef](#)]
34. Zaman, W.; Siddique, M.F.; Ullah, S.; Saleem, F.; Kim, J.-M. Hybrid Deep Learning Model for Fault Diagnosis in Centrifugal Pumps: A Comparative Study of VGG16, ResNet50, and Wavelet Coherence Analysis. *Machines* **2024**, *12*, 905. [[CrossRef](#)]
35. Yao, Y.; Gui, G.; Yang, S.; Zhang, S. A recursive multi-head self-attention learning for acoustic-based gear fault diagnosis in real-industrial noise condition. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108240. [[CrossRef](#)]
36. Raouf, I.; Kumar, P.; Kim, H.S. Deep learning-based fault diagnosis of servo motor bearing using the attention-guided feature aggregation network. *Expert. Syst. Appl.* **2024**, *258*, 125137. [[CrossRef](#)]
37. Xiao, B.; Miao, S.; Xia, D.; Huang, H.; Zhang, J. Detecting the backfill pipeline blockage and leakage through an LSTM-based deep learning model. *Int. J. Miner. Metall. Mater.* **2023**, *30*, 1573–1583. [[CrossRef](#)]
38. Wang, B.; Guo, Y.; Wang, D.; Zhang, Y.; He, R.; Chen, J. Prediction model of natural gas pipeline crack evolution based on optimized DCNN-LSTM. *Mech. Syst. Signal Process.* **2022**, *181*, 109557. [[CrossRef](#)]
39. Al Hamoud, A.; Hoenig, A.; Roy, K. Sentence subjectivity analysis of a political and ideological debate dataset using LSTM and BiLSTM with attention and GRU models. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 7974–7987. [[CrossRef](#)]

40. Ullah, S.; Ullah, N.; Siddique, M.F.; Ahmad, Z.; Kim, J.-M. Spatio-Temporal Feature Extraction for Pipeline Leak Detection in Smart Cities Using Acoustic Emission Signals: A One-Dimensional Hybrid Convolutional Neural Network–Long Short-Term Memory Approach. *Appl. Sci.* **2024**, *14*, 10339. [[CrossRef](#)]
41. Korkmaz, M.Ç.; Leiva, V.; Martin-Barreiro, C. The Continuous Bernoulli Distribution: Mathematical Characterization, Fractile Regression, Computational Simulations, and Applications. *Fractal Fract.* **2023**, *7*, 386. [[CrossRef](#)]
42. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification. In Proceedings of the PHM Society European Conference, Bilbao, Spain, 5–8 July 2016; Volume 3. [[CrossRef](#)]
43. Patil, C.; Theotokatos, G.; Wu, Y.; Lyons, T. Investigation of logarithmic signatures for feature extraction and application to marine engine fault diagnosis. *Eng. Appl. Artif. Intell.* **2024**, *138*, 109299. [[CrossRef](#)]
44. Fu, G.; Wei, Q.; Yang, Y.; Li, C. Bearing fault diagnosis based on CNN-BiLSTM and residual module. *Meas. Sci. Technol.* **2023**, *34*, 125050. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.