International Journal of Engineering and Technology Innovation, vol. 15, no. 3, 2025, pp. 314-331

Evaluation of Time-Frequency Representations for Deep Learning-Based Rotating Machinery Fault Diagnosis

Delanyo Kwame Bensah Kulevome[1], Man Qiu[2], Feng Cao[2,*], Edward Opoku-Mensah[3]

[1]Department of Computer Science and Information Engineering, Qilu Institute of Technology, Jinan, China

[2]Department of Intelligent Manufacturing and Control Engineering, Qilu Institute of Technology, Jinan, China

[3]Department of Information Technologies, HEC Montréal, Montreal, Canada

Received 21 January 2025; received in revised form 20 May 2025; accepted 22 May 2025

DOI: https://doi.org/10.46604/ijeti.2024.14774

Abstract

This study evaluates and compares five time-frequency representation (TFR) methods for fault diagnosis in rotating machinery, aiming to ensure operational reliability and reduce unexpected downtime. The methods—short-time Fourier transform (STFT), continuous wavelet transform (CWT), modified S-transform (MS-transform), smoothed pseudo Wigner-Ville distribution (SPWVD), and Hilbert-Huang transform (HHT)—are investigated. Vibration signals from benchmark bearing and gearbox datasets are converted into two-dimensional TFR data and classified using a convolutional neural network (CNN). Results show that MS-transform achieves the highest accuracy (up to 99.87%) under ideal conditions. STFT and CWT demonstrate better robustness in noisy environments, maintaining over 99% accuracy at 15 dB signal-to-noise ratio (SNR). SPWVD is computationally intensive with moderate performance, while HHT performs poorly under noise. Renyi entropy, energy conservation, and training time are also used to assess TFR quality. These findings support selecting appropriate TFR methods for industrial fault diagnosis.

Keywords: deep learning, time-frequency representation, fault diagnosis, rotating equipment

1. Introduction

Rotating machinery, including motors, turbines, and pumps, is critical component in industrial processes. As these systems become increasingly prevalent, the demand for accurate and reliable fault diagnosis techniques has increased significantly to ensure operational efficiency. This growing need has driven the evolution of signal analysis methods in fault diagnosis. Traditionally, signals are represented in the time domain $x(t)$ or in the frequency domain $x(f)$, with each representation inherently excluding the other [1]. This duality results in non-localized representations where time information is averaged out in the frequency domain, and frequency information is averaged out in the time domain. While numerous fault diagnosis frameworks utilize end-to-end deep learning (DL) models with raw time-series input, this approach may overlook critical fault characteristics, particularly in non-stationary vibration signals or those obscured by noise.

Time-frequency distributions (TFDs) have emerged as powerful tools for analyzing non-stationary vibration signals, offering a more detailed representation of signal characteristics by bridging the gap between time-domain and frequency-domain analyses. These two-dimensional (2D) representations reveal the time-varying frequency content of one-dimensional (1D) signals, providing a comprehensive view of temporal and spectral characteristics. Transforming vibration

---

* Corresponding author. E-mail address: fengcao@qlit.edu.cn

signals into time-frequency representations (TFR) before inputting them into a learning model presents additional benefits. This process enables convolutional neural networks (CNNs) to learn spatial patterns associated with localized frequency events, which is crucial for identifying transient faults. TFRs can highlight dominant frequency components more effectively in noisy conditions than raw time-domain data, thereby enhancing model robustness.

The field of fault diagnosis has undergone a paradigm shift with the advent of DL methods. These techniques, particularly CNNs, have recently demonstrated remarkable success in fault classification tasks. Additionally, advanced methods that leverage graph convolutional networks [2] and attention mechanisms [3] have been explored for the fault diagnosis of industrial assets. Furthermore, most DL algorithms that process 2D data utilize TFR of vibration signals to capture the intricate details of faults in rotating equipment. However, the effectiveness of these models largely depends on the quality and suitability of the input data, highlighting the critical role of selecting an appropriate TFR. Additionally, these representations are subject to constraints imposed by the uncertainty principle, necessitating careful selection for specific applications.

Given the variety of available TFDs and their significant impact on DL-based fault diagnosis, this study focuses on five widely used time-frequency analysis techniques applied to vibration signals: short-time Fourier transform (STFT), continuous wavelet transform (CWT), S-transform, smoothed pseudo Wigner-Ville distribution (SPWVD), and Hilbert-Huang transform (HHT). Each of these methods offers a unique approach to mapping vibration signals onto the time-frequency plane, with trade-offs in resolution, computational complexity, and interpretability.

Although STFT, CWT, and HHT have been shown to enhance performance, a comprehensive comparison in the context of rotating machinery fault diagnosis remains scarce. Previous studies have been limited in scope or focused on different domains. For instance, Raqib et al. [4] investigated the performance of the fast Fourier transform (FFT), STFT, and CWT with different mother wavelets in monitoring the condition of a DC-link capacitor. It was discovered that the CWT yielded optimal results with fine temporal granularity, although some salient information was lost due to mother wavelet selection. In other domains, the effectiveness of STFT, HHT, CWT, and their variants in extracting information from recorded sounds has been analyzed under varying experimental setups. Recent studies have primarily compared the performance of STFT and wavelet transforms (WT) with other methods [5]. However, despite the proliferation of various TFRs in machinery fault diagnosis, there has been limited comparative analysis to comprehensively evaluate their significance, strengths, and weaknesses.

This paper presents a comprehensive quantitative comparative analysis of five prominent TFR techniques, offering the following key contributions to the field of rotating machinery fault diagnosis:

(1) Comprehensive quantitative analysis: Based on the reviewed literature, this study is the first to provide a detailed quantitative analysis of these TFRs specifically for vibration signals from rotating equipment and their performance in fault diagnosis using a robust CNN with few parameters.
(2) Multifaceted performance assessment: This paper quantitatively discusses crucial performance aspects, including time-frequency concentration and energy conservation (EC) of the TFRs. Importantly, it analyzes how these factors affect the accuracy of fault diagnostic models, offering essential insights.
(3) Extensive experimental validation: The effectiveness of each TFR is validated through comprehensive experiments using vibration datasets from rotating equipment under various fault conditions. The robustness and versatility of the classification baseline model in identifying and classifying faults using each TFR technique are further examined under different noise conditions. This analysis is essential for performance assessment in industrial environments where noise is inevitable.

The rest of the paper is organized as follows: Section 2 presents related works, followed by the problem statement in Section 3. Section 4 details the TFR method and the DL-based architecture used in this paper. Sections 5 and 6 present the experimental studies and discuss the results. Finally, Section 7 summarizes the conclusion, limitations, and future work.

2. Related Work

Research on fault diagnosis using TFR has gained significant momentum in recent years. Various TFR methods have been explored for their ability to analyze non-stationary signals. These techniques enable detailed feature extraction by mapping signal energy across time and frequency domains. This section reviews recent studies that employ these methods.

2.1. STFT-based fault diagnosis

STFT has been a cornerstone technique in machinery fault diagnosis because it provides TFRs of non-stationary signals. Wu et al. [6] proposed a fault diagnosis method for rolling bearings by transforming vibration signals into spectrograms and constructing a neural network model based on AlexNet. Xin et al. [7] proposed a framework for fault diagnosis of wheelset bearings by combining the logarithmic STFT and a modified self-calibrated convolution to improve accuracy and computational efficiency. Similarly, Zhang and Deng [8] proposed an intelligent fault diagnosis method for rolling bearings using a combination of STFT and CNN. Ribeiro Junior et al. [9] also utilized a CNN with STFT to extract detailed information from vibration signals for accurate fault analysis, where the CNN is trained to diagnose faults using STFT data as the sole input.

While STFT has proven effective in multiple fault diagnosis scenarios, challenges remain in optimizing time-frequency resolution trade-offs and handling highly non-stationary signals. The versatility of the STFT has also paved the way for improving the classification performance of DL-based methods using innovative techniques to generate synthetic representations [10]. STFT continues to be a vital tool in machinery fault diagnosis, with ongoing research continually expanding its capabilities and applications.

2.2. WT-based fault diagnosis

WT is a powerful method in rotary machine fault diagnosis, extensively used for fault feature extraction and widely studied for performance enhancement. As data-driven intelligent fault diagnosis becomes increasingly popular, WT is being re-evaluated for its interpretability in intelligent diagnosis models. Fan et al. [11] proposed a new wavelet-based method that decomposes the vibration signals, analyzes their statistical properties, and utilizes a generalized likelihood ratio test for early detection of bearing compound faults. With the recent advancement in DL-based fault diagnosis with multi-sensor data fusion [12], some studies have merged WT and transfer learning for fault diagnosis [13].

Tran et al. [14] proposed a convolutional attention neural network (CANN) that utilizes CWT to convert vibration signals into scalogram feature data, capturing time-frequency information essential for training the CANN model. Yan et al. [15] reviewed the advancement in WT-based fault diagnosis research in the past decade, covering traditional and intelligent fault diagnosis perspectives. Recent advancements have leveraged the flexibility of WT to formulate data augmentation techniques, thereby enhancing the performance of DL-based fault classification models [16].

2.3. S-transform-based fault diagnosis

Several researchers have explored S-transform-based approaches for various fault diagnosis applications. Zheng et al. [17] proposed a fault diagnostics approach that involves training a 2D CNN with signals converted from 1D to 2D TFR using the S-transform. A fault diagnosis method for bearings was proposed by Peng and Ma [18], utilizing parameter optimization and adaptive generalized S-transform (AGST) to enhance fault feature extraction accuracy. Xu et al. [19] proposed a time-extracting S-transform (TEST) to mitigate the uncertainties in parameter selection. Various versions of the S-transform, such as the time-reassigned synchrosqueezing extracting S-transform (TSSEST) [20], adaptive generalized S-synchroextracting transform (AGSSET) [21], and demodulated synchrosqueezing S-transform (DSSST) [22], have been proposed to enhance TFR resolution for improved fault feature extraction. Similarly, Yi et al. [23] proposed a second-order synchrosqueezing modified S-transform (SSMST2) for a more concentrated TFR of vibration signals.

2.4. WVD-based fault diagnosis

The application of the Wigner-Ville distribution (WVD) has shown significant potential in fault diagnosis. Lv and Li [24] proposed combining a graph Wigner-Ville distribution (GWVD) and a CNN-based method for bearing fault diagnosis. The proposed GWVD effectively suppresses cross-interference terms, providing high energy aggregation and resolution that improves CNN's pattern recognition accuracy in noisy environments. Similarly, Li et al. [25] proposed a new method for diagnosing planetary gearbox faults using a CNN to process TFR derived from WVD. The method employs a K-singular value decomposition (K-SVD) technique to enhance the WVD resolution. Since the WVD suffers from cross-term interference, some researchers have used innovative methods to combine Wigner-Ville distribution with generative adversarial networks (WVD-GAN) [26] and sliding mode singular spectrum analysis (SM-SSA) and Wigner-Ville distribution (WVD), combined as SM-SSA-WVD [27], aiming to achieve highly concentrated TFR.

2.5. HHT-based fault diagnosis

The HHT technique has gained significant attention in the scientific community. As a result, there has been growing interest in its application across diverse fields of study. Elbouchikhi et al. [28] proposed a method for detecting rolling element bearing faults in induction machines by analyzing stator currents using HHT. Similarly, Cheng et al. [29] combined HHT with a self-organizing map (SOM) for gear fault detection, enabling accurate fault classification and precise extraction of fault characteristics in vibration signals. Zabin et al. [30] proposed a DL-based model for industrial fault diagnosis using a standard induction motor dataset. Their model employs the Hilbert transform for signal preprocessing, generates texture images, and trains deep convolutional neural networks (DCNNs) for fault classification.

3. Problem Statement

Ensuring the reliable operation of rotating equipment is critical to avoid costly downtime and ensure safety in industrial applications. Fault diagnosis of these machines entails detecting, identifying, and classifying faults from vibration signals, which are often non-stationary. TFRs can analyze these signals by capturing time and frequency information. However, determining the most effective TFR for DL-based fault diagnosis remains challenging. This paper addresses this gap by comparing the performance of different TFRs: STFT, CWT, S-transform, WVD, and HHT.

Given a set of vibration signals from rotating equipment under various fault conditions, the goal is to classify each signal into one of C fault categories. This involves computing the TFR $T_i(t,f)$ for each signal $x_i(t)$ using methods such as STFT, CWT, S-transform, WVD, and HHT. Subsequently, a DL model $\mathcal{M}$ is used to extract the significant features in $T_i(t,f)$ and perform the classification of the various fault categories:

$$\hat{y}_i = \mathcal{M}\left[T_i(t,f)\right] \tag{1}$$

where $\hat{y}_i$ are the predicted values. The objective is to minimize the classification error:

$$\min_{\mathcal{M}} \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(y_i,\hat{y}_i) \tag{2}$$

where $\mathcal{L}$ is the loss function (e.g., cross-entropy loss) and y_i are the accurate fault labels.

4. Materials and Methods

This framework comprehensively compares different TFR methods within a unified diagnostic architecture. The process begins with transforming raw vibration signals into TFRs using various established techniques. The resulting TFR data are fed into a lightweight CNN for fault classification. The network architecture is intentionally kept consistent to ensure that the comparative analysis remains focused on the discriminative power of the input representations rather than differences in model capacity. Furthermore, additive white Gaussian noise is introduced at prescribed signal-to-noise ratio (SNR) levels to simulate realistic industrial environments. This step enables the emulation of different noise intensities encountered in practical settings. Fig. 1 depicts the overall methodological workflow.

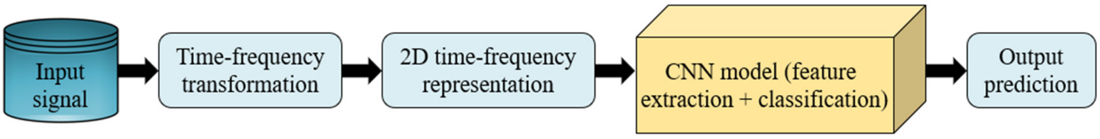


Fig. 1 Methodological workflow

4.1. *Time-frequency analysis and transformation*

Time-frequency analysis represents a signal in both time and frequency domains, capturing the signal's frequency content over time. Fig. 2 illustrates the steps in deriving the TFR used in this paper. Each method processes the input signal through transformations to produce the respective output.

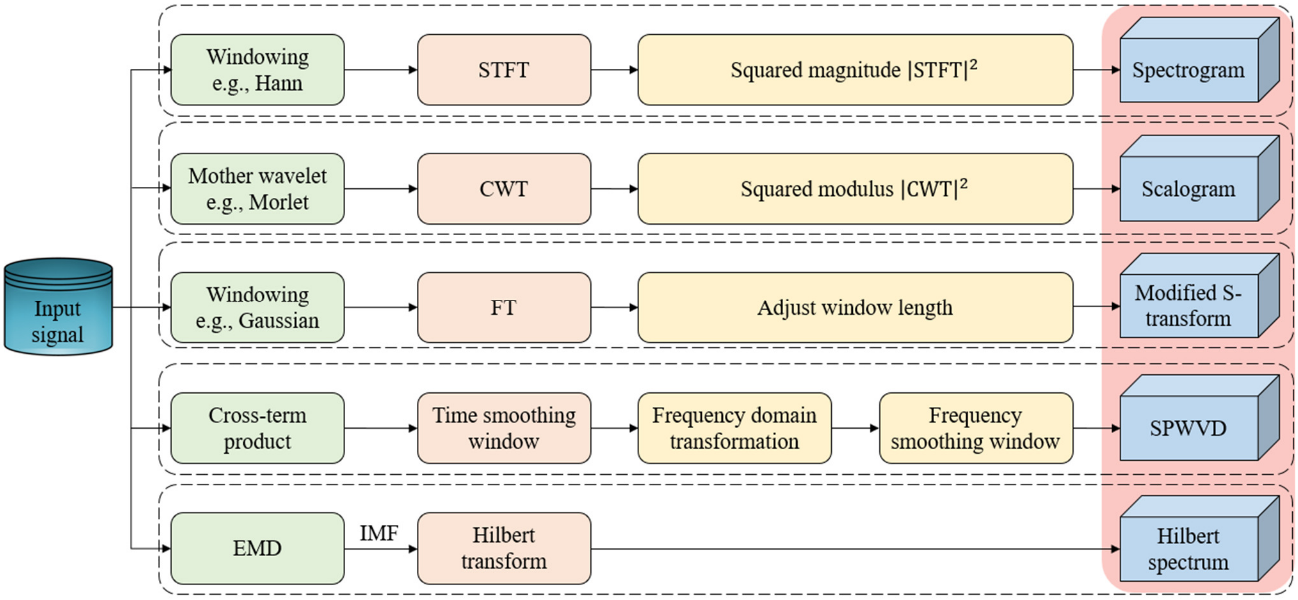


Fig. 2 TFR derivation and generation process

(1) STFT and spectrograms

The STFT divides a signal into short overlapping segments and computes the Fourier transform for each segment, providing time-localized frequency information. The STFT is defined as:

$$X(\tau, f) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-j2\pi f t} dt \tag{3}$$

where $x(t)$ is the signal, $w(t-\tau)$ is a window function centered at time τ, and f is the frequency. The 2D representation, known as the spectrogram, is obtained by computing the squared magnitude of the STFT as:

$$\text{Spectrogram}(\tau, f) = \left| X(\tau, f) \right|^2 \tag{4}$$

The spectrogram is, therefore, energy-preserving and non-negative. This means $\text{Spectrogram}(\tau, f) > 0, \forall \tau, f \in \mathbb{R}^d$.

(2) CWT and scalograms

The CWT decomposes a signal into wavelets localized in time and frequency, enabling multi-resolution analysis. The CWT is defined as:

$$W(b,a)=\frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty}x(t)\psi^{*}\left(\frac{t-b}{a}\right)dt \tag{5}$$

where $x(t)$ is the signal, $\psi(t)$ is the mother wavelet, a and b are the scale and translation parameters, respectively. $W(b,a)$ contains information concerning $x(t)$ at the scale a around the point b. The 2D representation, known as the scalogram, is obtained by computing the squared modulus of the CWT:

$$\text{Scalogram}(b,a)=|W(b,a)|^{2} \tag{6}$$

(3) Modified S-transform (MS-transform)

The S-transform is a specific implementation of the Stockwell transform (ST), combining elements of the STFT and CWT. The S-transform is defined as:

$$S(\tau,f)=\int_{-\infty}^{\infty}x(t)\frac{|f|}{\sqrt{2\pi}}e^{-\frac{(t-\tau)^{2}f^{2}}{2}}e^{-j2\pi ft}dt \tag{7}$$

where $x(t)$ is the signal, τ is the time shift, and f is the frequency. The window function in the S-transform is a scalable Gaussian window that changes with frequency:

$$w(t-\tau,f)=\frac{|f|}{\sqrt{2\pi}}e^{-\frac{(t-\tau)^{2}f^{2}}{2}} \tag{8}$$

However, the MS-transform is adopted herein. The original S-transform places unnecessary restrictions on the window function. This is remedied by parameterizing the generalized S-transform by a width control coefficient γ, as follows:

$$S(\tau,f,\gamma)=\int_{-\infty}^{\infty}x(t)\frac{|f|}{\gamma\sqrt{2\pi}}e^{-\frac{(t-\tau)^{2}f^{2}}{2\gamma^{2}}}e^{-j2\pi ft}dt \tag{9}$$

The 2D representation is obtained by plotting the magnitude squared of the MS-transform:

$$\text{MS}(\tau,f,\gamma)=|S(\tau,f,\gamma)|^{2} \tag{10}$$

(4) Smoothed-pseudo WVD

The WVD is a quadratic TFR that overcomes the artifacts that significantly diminish the usefulness of the Wigner distribution (WD) for real signals by modifying the WD using an analytic signal. The WD is defined as:

$$W_{x}(t,f)=\int_{-\infty}^{\infty}x\left(t+\frac{\tau}{2}\right)x^{*}\left(t-\frac{\tau}{2}\right)e^{-j2\pi ft}d\tau \tag{11}$$

where $x(t)$ is the signal and $x^{*}(t)$ is its complex conjugate. However, due to the bi-linear nature of WVD and the presence of negative values, WVD has confusing results in time-frequency analysis, especially for signals with multi components. Hence, this paper adopts a new TFR, the smoothed pseudo Wigner-Ville distribution (SPWVD), which enables two independent analysis windows, one in the time domain and the other in the frequency domain. The SPWVD of a signal $x(t)$ is computed as:

$$SPWVD_{x}(t,f;g,h)=\int_{-\infty}^{\infty}h(\tau)\int_{-\infty}^{\infty}g(z-t)x\left(z+\frac{\tau}{2}\right)x^{*}\left(z-\frac{\tau}{2}\right)e^{-j2\pi ft}dzd\tau \tag{12}$$

where $z(t)$ is complex and the analytic associate of $x(t)$, h and g are the frequency and time smoothing windows, respectively. Simple geometrical considerations show that such a procedure provides insights into the time-frequency content of the signal.

(5) Hilbert-Huang transform (HHT)

HHT combines a two-step process involving empirical mode decomposition (EMD) followed by the Hilbert spectral analysis to yield a TFR of a signal. The EMD smoothens the non-stationary signals and decomposes them into intrinsic mode functions (IMFs). After the EMD, the target signal $x(t)$ is expressed as the sum of imf_i (i = 1, 2, 3, ..., n) and the residue component r_n as:

$$x(t)=\sum_{i=1}^{n} imf_i + r_n \tag{13}$$

The Hilbert transform is computed as:

$$H\left[imf_i(t)\right]=\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{imf_i(\tau)}{t-\tau}d\tau \tag{14}$$

where P denotes the Cauchy principal value. Subsequently, the analytic signal is derived as:

$$z_i(t)=imf_i(t)+jH\left[imf_i(t)\right]=a_i(t)e^{j\theta_i(t)} \tag{15}$$

where $a_i(t)$ is the instantaneous amplitude and $\theta_i(t)$ is the instantaneous phase. By imposing the Hilbert transform on each IMF component, the Hilbert spectrum of $x(t)$ is obtained as:

$$H(\omega,t)=\sum_{i=1}^{n} a_i(t)\delta\left(\omega-\omega_i(t)\right) \tag{16}$$

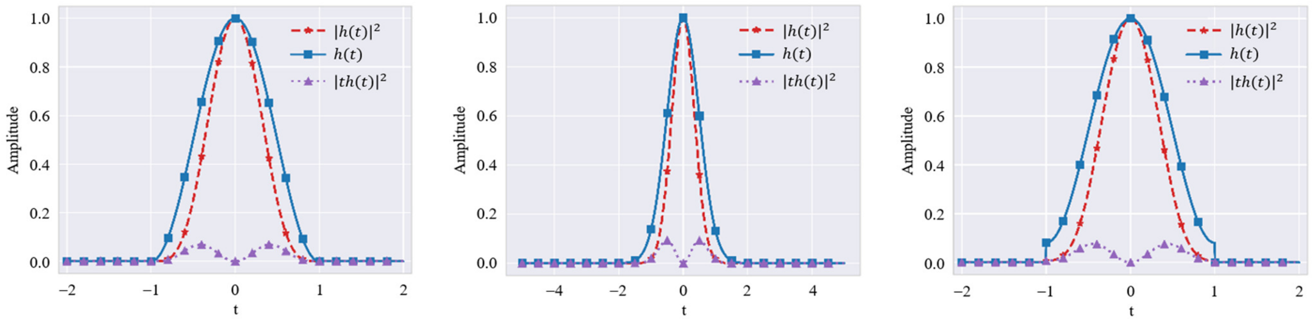
where $\omega_i(t) = d\theta_i(t)/dt$ is the instantaneous frequency.

4.2. *Selecting a window function*

Window functions are critical in certain TFDs, such as STFT, MS-transform, and SPWVD. Careful selection of an appropriate window function is essential, as it can profoundly influence the quality and interpretability of the resulting TFR. The window function must be square integrable to ensure that the resulting TFR is well-defined and has desirable mathematical properties.

Definition 1. If $h(t) \in L^2(\mathbb{R})$, $\|h(t)\|_2 = 0$, and $th(t) \in L^2(\mathbb{R})$, then $h(t)$ is called a window function.

Definition 2. A function $w(t)$ is considered square integrable if the integral of its absolute value squared over its entire domain is finite, expressed as $\int |w(t)|^2 dt < \infty$.



(a) Hann (b) Gaussian (c) Hamming

Fig. 3 Characteristics of three standard window functions

Fig. 3 shows how each window function satisfies the conditions for being a valid window in TFRs, including square integrability and finite time-weighted energy. A narrower window is preferable for bearing faults with impulsive characteristics. In contrast, a window with high frequency resolution could be more suitable for gear faults that often manifest as modulation sidebands. For each window function, three components are plotted:

(1) $h(t)$: The window function itself.

(2) $|h(t)^2|$: The squared magnitude of the window function.

(3) $th(t)^2$: The squared magnitude of the time-weighted window function.

The distinct shapes of these functions highlight their different properties.

4.3. CNN architecture

CNNs are powerful tools for classification, automatically learning features at different abstraction levels through a combination of convolutional, pooling, and fully connected layers. Each convolution kernel is applied across the input data, computing the dot product between the kernel and the input. The resulting k-th feature map, Z_k, after using a nonlinear transformation (σ) is obtained as follows:

$$Z_k = \sigma\left(W_k \otimes x + b_k\right) \tag{17}$$

where W_k and b_k are the k-th convolution kernel and bias term, respectively; x is the input data; the operation $\otimes$ denotes a 2D convolution.

Multiple feature maps are generated by extending the convolution operation across the input data's depth, width, and height. Pooling layers perform downsampling on the resulting feature maps. After multiple convolution blocks, fully connected layers are used for final classification. The CNN model is trained by minimizing cross-entropy loss, which measures the difference between the predicted probabilities and the actual labels. The weights are updated through backpropagation within the network, with gradients computed for the loss function and optimized using stochastic gradient descent. A detailed description of the network architecture used in this study is presented in Table 1.

Table 1 Proposed network architecture

Layer	Output	Description
Input (64, 64)	-	TFR data
Conv2D	(64, 64, 16)	3×3 kernel size, 16 filters, stride 1
Activation	(64, 64, 16)	ReLU
Batch norm	(64, 64, 16)	-
Max pooling	(32, 32, 16)	2×2 pool size
Dropout	(32, 32, 16)	0.2
Conv2D	(16, 16, 32)	3×3 kernel size, 32 filters, stride 2
Activation	(16, 16, 32)	ReLU
Batch norm	(16, 16, 32)	-
Max pooling	(8, 8, 32)	2×2 pool size
Dropout	(8, 8, 32)	0.2
Conv2D	(4, 4, 64)	3×3 kernel size, 64 filters, stride 2
Activation	(4, 4, 64)	ReLU
Batch norm	(4, 4, 64)	-
Max-pooling	(2, 2, 64)	2×2 pool size
Flatten	(256)	-
Dense	(256)	256 units, ReLU activation
Dense	(Number of classes)	Class units, softmax activation

The model is formulated to process 2D input data of size 64×64. The architecture includes three convolutional layers with 16, 32, and 64 filters, respectively. Each convolutional layer is followed by ReLU activation, batch normalization (BN), max-pooling, and dropout layers. The convolutional layers all use a 3×3 kernel size, with strides of 1 and 2, respectively, while the max-pooling layers utilize a 2×2 kernel size. Following the convolutional layers, the network is flattened, leading to two dense layers with 256 units and a final output layer that matches the number of classes. The first dense layer employs ReLU activation, while the output layer employs the softmax function for classification. The model comprises approximately 0.92 million trainable parameters.

5. Experiments

This section presents the experimental setup used to validate the performance of the five TFRs for rotating machinery fault diagnosis. First, the benchmark datasets are described, followed by data preprocessing, implementation details, and the evaluation metrics used in the experiment.

5.1. Dataset description

Two publicly available benchmark datasets are used for the experiments. The Case Western Reserve University (CWRU) bearing dataset consists of vibration signals for normal operation and three fault types across four distinct load levels. The Southeast University (SEU) gearbox dataset includes healthy baseline data and four gear-fault conditions, each recorded under two speed–load settings.

(1) Bearing dataset

The bearing dataset from CWRU is obtained by introducing single-point defects of various fault sizes and load conditions into a test bearing. The data are categorized into normal condition (N), inner race fault (IR), outer race fault (OR), and roller or ball fault (BF). The fault frequencies correspond to multiples of the operating speed, which varies between 1,797 rpm and 1,730 rpm depending on the applied load. Vibration data from a single-row deep groove bearing (SKF 6205-2RS) sampled at 12 kHz at the motor's drive end (DE) were used in the experiments. A detailed description of the dataset is provided in Table 2. For each load, there are ten unique classes: N, OR (0.007), OR (0.014), OR (0.021), IR (0.007), IR (0.014), IR (0.021), BF (0.007), BF (0.014), BF (0.021).

Table 2 Description of the CWRU bearing DE dataset

Load (hp)	Operating speed (rpm)	Fault state	Fault diameter (inches)
0	1,797	N, OR, IR, BF	0, 0.007, 0.014, 0.021
1	1,772	N, OR, IR, BF	0, 0.007, 0.014, 0.021
2	1,750	N, OR, IR, BF	0, 0.007, 0.014, 0.021
3	1,730	N, OR, IR, BF	0, 0.007, 0.014, 0.021

(2) Gearbox dataset

The gearbox fault diagnosis dataset from SEU was acquired from a drivetrain dynamic simulator (DDS) and includes bearing and gear data. For this study, the gear fault sub-dataset was selected. The data were collected under two operating conditions, with system loads corresponding to rotating speeds of "20 Hz – 0V" and "30 Hz – 2V." It comprises five different working conditions: one healthy and four failure states. A detailed description of the gear dataset is illustrated in Table 3.

Table 3 Description of the gear fault sub-dataset

Type	Description	Class ID
Health	Healthy gear	0
Chipped	Crack occurs in the gear feet	1
Miss	Missing one foot in the gear	2
Root	Crack occurs in the root of the gear feet	3
Surface	Wear occurs on the gear surface	4

5.2. Data preprocessing

The time-series vibration signals from the bearing dataset are divided into smaller samples, each containing 400 data points, as illustrated in Fig. 4. It shows a bearing IR vibration signal segmented into ‘n’ equal, non-overlapping segments or samples, each containing a specific number of data points. The number of data points in a sample, N_d, is determined as follows:

$$N_d = 60\frac{Fs}{spd} \tag{18}$$

where Fs and spd are the sampling frequency and operating speed, respectively. Hence, for Fs = 12 kHz and spd = 1,797 rpm, the number of data points per revolution is estimated as:

$$N_d = 60\frac{12000}{1797} \approx 400 \tag{19}$$

Each class in the bearing DE dataset comprises 300 samples with 400 data points per sample.

Each file in the gear dataset contains eight columns of signals. Columns 2, 3, and 4 correspond to the planetary gearbox’s vibrations in the x, y, and z directions, respectively. Column 2 in each file is used in the experimental analysis. The time-series vibration signal, collected under the five working conditions with the rotating speed-system load (RS-SL) set to 20 Hz – 0V or 30 Hz – 2V, is divided into 1,048 segments. Each segment contains 1,000 data points.

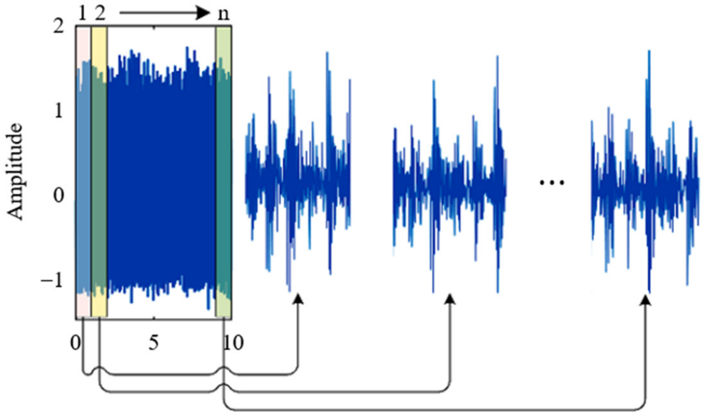


Fig. 4 Segmentation of bearing inner-race fault vibration signal

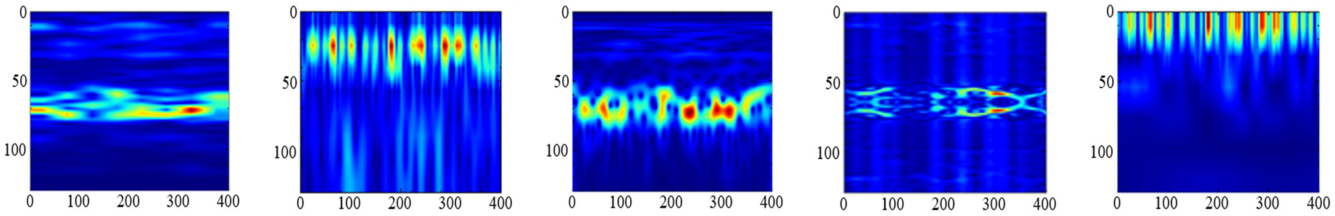
5.3. Implementation details

The time-series samples in each dataset were converted into 2D data representations corresponding to the five TFDs discussed earlier. As detailed in the previous section, the classifier model consists of three convolutional layers and two fully connected layers. Model training is conducted by optimizing the cross-entropy loss between the target and predicted values. The adaptive gradient algorithm (Adagrad) is employed as the optimizer, with a learning rate of 0.001 and a batch size of 16. The training process is set to 50 epochs. The k-fold cross-validation (k = 5) approach was used to avoid overfitting and estimate the model’s generalization performance. The hyperparameters were selected based on experimental studies and their impact on fault diagnosis performance.

The experiments were conducted using Python 3.9.18 with the Keras 2.12.0 API, executed on a Windows system equipped with a 6-core CPU, a 2.69 GHz processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3050 GPU with 4 GB of memory. Table 4 provides the parameter settings for generating each TFR using packages in Python. Samples of the various TFRs generated using the bearing DE dataset are shown in Fig. 5. The intensity level at each TF point in the TFRs reflects signal strength: low-intensity regions indicate weak signal components, while high-intensity regions indicate strong signal components (high-energy content).

Table 4 Parameters for generating TFRs

Method	Essential parameter settings
Spectrogram	Window type: Hann, Overlap: 50%
Scalogram	Mother wavelet: Morlet, Scales: 1-128
MS-transform	Window type: Gaussian, gamma: 3
SPWVD	Time & frequency domain windows: Hamming
HHT	Number of imfs: 7



(a) Spectrogram (b) Scalogram (c) MS-transform (d) SPWVD (e) HHT

Fig. 5 TFR of a bearing DE dataset vibration sample

Further experiments were conducted to assess the fault diagnosis performance under different SNR levels. Gaussian noise was introduced into the vibration signals from both the bearing and gearbox datasets to create realistic and variable industrial noise conditions. This addition of noise allows for the assessment of how various TFRs perform under different levels of signal degradation. A controlled method for adding noise was implemented, targeting specific SNR levels of 0 dB, 5 dB, 10 dB, and 15 dB. A lower SNR signifies more substantial measurement noise, making the fault diagnostics model more challenging. The SNR in dB is defined by the equation:

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{20}$$

where P_{signal} and P_{noise} represent the average powers of the signal and the noise, respectively.

5.4. Evaluation metrics

This section presents the evaluation metrics used to assess the quality and representativeness of the TFRs and the model’s classification performance. The quality of the transformation is measured utilizing Rényi entropy and EC metrics, whereas the model’s performance is evaluated using classification accuracy and the F1 score metrics. Subsequent sections describe these metrics in detail.

(1) Transform evaluation

The concentration measure (CM), EC, and computational efficiency are used to evaluate the quality and representativeness of TFRs. These metrics comprehensively assess the TFR’s performance and suitability for specific tasks in fault diagnostics. The Renyi entropy is used to evaluate energy spread in a TFR. It provides information about how concentrated or dispersed the signal’s energy is in the time-frequency plane. A lower entropy value indicates a higher concentration, whereas a higher entropy value indicates more energy dispersion. For a given TFR $P(t, f)$, the Renyi entropy of order α is defined as:

$$H_{\alpha} = \frac{1}{1-\alpha}\log_2\left(\sum_{t,f} P(t,f)^{\alpha}\right) \tag{21}$$

where $P(t,f) = |TFR(t,f)|^2 / \sum_{t,f}|TFR(t,f)|^2$ is the normalized energy distribution in the time-frequency plane; α is the order of the entropy, typically $\alpha = 2$ is used for time-frequency analysis.

The EC metric evaluates whether the total energy in the TFR is preserved, considering the energy in the time-domain signal. It ensures that no significant information is lost during the transformation. The total energy of the signal in the time domain is evaluated as:

$$E_{\text{signal}} = \sum_{t} \left| x(t) \right|^2 \tag{22}$$

where $x(t)$ is the original time-domain signal. The total energy of the signal in the TFR is:

$$E_{\text{TFR}} = \sum_{t,f} \left| \text{TFR}(t, f) \right|^2 \tag{23}$$

EC is satisfied when:

$$E_{\text{signal}} = E_{\text{TFR}} \tag{24}$$

however, due to numerical approximations or the choice of TFR, some energy may be lost. Hence, the goal is to have $E_{\text{signal}} \approx E_{\text{TFR}}$. This comparison ensures that the TFR does not introduce significant errors or lose critical signal information.

(2) Model evaluation

Accuracy and F1 score are two key evaluation metrics used to assess the performance of the fault diagnostics model. Accuracy measures the overall correctness of the model's predictions across all classes. It is the ratio of correct predictions to the total number of predictions made by the model. The F1 score offers a balance between precision and recall. The definitions of these metrics are:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{25}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{26}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Precision measures the accuracy of positive predictions and is computed as $\text{TP}/\text{TP} + \text{FP}$. It is mainly concerned with minimizing false positives. Recall, $\text{TP}/\text{TP} + \text{FN}$, assesses the model's ability to correctly identify positive instances by focusing on reducing false negatives.

6. Results and Discussion

This section presents a detailed discussion of the experimental results, analyzing the performance of each TFR, including its strengths and limitations. The quality, representativeness, and computational efficiency are thoroughly examined to provide a comprehensive performance evaluation.

6.1. Result analysis

The prediction results presented in Table 5 comprehensively compare the performance of various TFR techniques: spectrogram, scalogram, MS-transform, SPWVD, and HHT. The comparison is performed on two distinct datasets: Bearing and Gear. Quantitatively, the most striking observation is that MS-transform consistently delivers superior performance across both datasets. Regarding the bearing dataset, across all levels (0 hp, 1 hp, 2 hp, and 3 hp), MS-transform achieves near-perfect accuracy, with values ranging from 99.36% to 99.87%, and F1 scores between 0.9936 and 0.9987.

Table 5 Performance comparison of TFR using bearing and gear dataset

Dataset	Spectrogram		Scalogram		MS-transform		SPWVD		HHT	
	Acc (%)	F1 score	Acc (%)	F1 score	Acc (%)	F1 score	Acc (%)	F1 score	Acc (%)	F1 score
0 hp	98.61±1.82	0.9860	92.76±0.66	0.9251	99.70±0.23	0.9970	90.61±0.10	0.9041	82.73±2.27	0.8211
1 hp	99.15±1.02	0.9913	93.76±2.40	0.9357	99.82±0.15	0.9982	91.15±0.12	0.9094	84.97±3.02	0.8391
2 hp	99.72±0.15	0.9972	95.82±1.01	0.9582	99.87±0.20	0.9987	91.19±1.01	0.9101	88.90±1.90	0.8857
3 hp	99.64±0.30	0.9964	99.21±0.17	0.9921	99.36±0.52	0.9936	92.75±0.12	0.9270	89.36±1.11	0.8863
20 Hz – 0V	99.36±0.61	0.9936	99.45±0.20	0.9945	99.51±0.06	0.9951	-	-	84.45±1.26	0.8246
30 Hz – 0V	98.77±0.41	0.9877	99.49±0.20	0.9949	96.58±2.36	0.9658	-	-	77.91±0.92	0.7735

This performance is remarkably consistent, suggesting that the MS-transform is highly robust in handling bearing data with minimal classification errors. For the gear dataset, similar trends emerge, with MS-transform achieving 99.51% accuracy and a 0.9951 F1 score at 20 Hz – 0V, while slightly lower accuracy (96.58%) and F1 score (0.9658) are observed at 30 Hz – 0V. Even at its lowest performance, the MS-transform remains highly effective, outperforming other techniques significantly.

The spectrogram technique is the second-best performer, though with slightly lower results. In the bearing dataset, the spectrogram achieves accuracy between 98.61% and 99.72%, with F1 scores ranging from 0.9860 to 0.9972. Notably, the spectrogram performs better at higher hp levels (2 hp and 3 hp), where its accuracy is above 99%, almost closing the gap with MS-transform. However, at lower hp levels, the spectrogram's performance declines slightly. The spectrogram provides consistent results in the gear dataset, with accuracy ranging from 98.77% to 99.36% and F1 scores from 0.9877 to 0.9936. Despite being competitive with MS-transform, the spectrogram falls short in both datasets, particularly in terms of F1 scores, which indicates that it struggles slightly more with precision or recall in some classification scenarios.

The scalogram technique also performs well but lags behind both the MS-transform and the spectrogram. In the bearing dataset, the scalogram's accuracy ranges from 92.76% at 0 hp to 99.21% at 3 hp, with corresponding F1 scores between 0.9251 and 0.9921. Similar to the spectrogram, the scalogram shows improvement at higher horsepower levels, suggesting that it is better suited to classify data with higher mechanical loads. However, its performance at 0 hp and 1 hp suggests potential limitations when handling lower-power data. In the gear dataset, the scalogram performs comparably to the spectrogram, with accuracy of 99.45% and 99.49% at 20 Hz – 0V and 30 Hz – 0V, respectively. Despite its results, the scalogram consistently falls behind the MS-transform and spectrogram in both datasets, suggesting it may not handle specific signal characteristics as effectively.

The SPWVD technique shows a significant drop in performance compared to the other methods, especially for the bearing dataset, where its accuracy ranges from 90.61% at 0 hp to 92.75% at 3 hp. This result indicates that SPWVD struggles to capture the distinguishing features necessary for effective classification, likely owing to its inherent trade-offs between time and frequency resolution. This may reduce its suitability for the type of non-stationary signals present in the bearing dataset. Notably, SPWVD results are absent from the gear dataset, precluding any direct comparison with the other techniques due to its high computational requirements, as given in Table 5. However, based on its performance in the bearing dataset, it would likely exhibit similar limitations to those in the gear dataset.

Finally, the HHT emerges as the weakest performer overall, particularly in the bearing dataset, where its accuracy is significantly lower, ranging from 82.73% at 0 hp to 89.36% at 3 hp. The corresponding F1 scores (0.8211 to 0.8863) reveal that HHT has considerable difficulty in precision and recall, likely due to its sensitivity to noise and non-stationarity in the signals. In the gear dataset, HHT fares even worse, with accuracies of 84.45% at 20 Hz – 0V and a markedly low 77.91% at 30 Hz – 0V. These results indicate that HHT may not be suitable for tasks requiring high classification accuracy or reliability, as it struggles to distinguish between signal features effectively. Fig. 6 illustrates the accuracy distribution for the four significant TFR methods across the bearing and gear datasets.

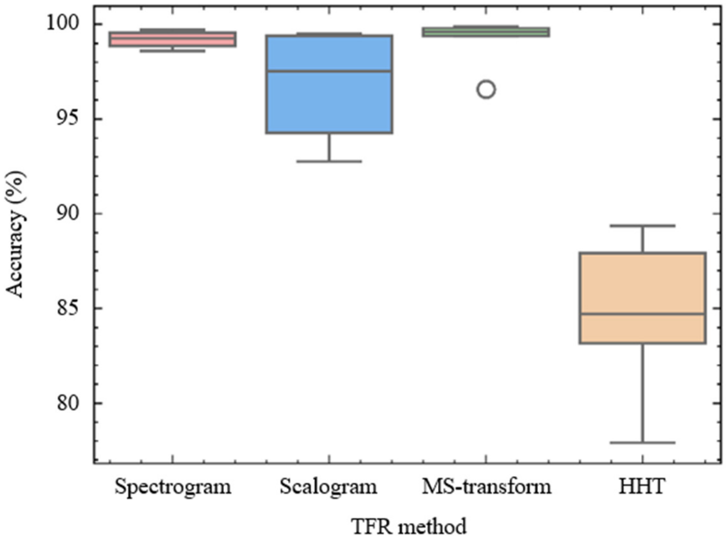


Fig. 6 Box plot of accuracy distributions for four TFRs

Qualitatively, these results highlight the importance of choosing the proper TFR technique based on the nature of the dataset and the specific application. With its consistently high performance across both datasets, the MS-transform demonstrates a clear advantage in robustness and adaptability to different signal characteristics. Its near-perfect accuracy and F1 scores suggest it is particularly adept at capturing essential features for classification, likely due to its ability to represent signals in both time and frequency domains with minimal loss of information. This performance is consistent with MS-transform's theoretical advantages regarding resolution and flexibility.

While not as robust as MS-transform, the spectrogram and scalogram techniques still offer reliable performance, especially in high hp or frequency conditions. These methods are more familiar in practical applications, particularly in industries where signal interpretation in the frequency domain is paramount. However, their slight underperformance compared to MS-transform suggests that they may not capture some finer details in the signal, possibly due to their fixed resolution grids, which could lead to oversimplification of the data.

In contrast, the poor performance of SPWVD and HHT raises concerns about their applicability in practical settings for this classification task. SPWVD's inherent trade-offs in TFR likely explain its lower accuracy, while HHT's instability in handling noisy or non-stationary signals is evident from its subpar results. These findings suggest that these techniques may be more suited for exploratory analysis or theoretical applications than real-world classification tasks where accuracy is critical.

While this study focuses on bearings and gearboxes, transforming vibration signals into TFRs and classifying them with DL-based models such as CNNs is inherently adaptable to a wide range of rotating machinery, including turbines, compressors, and industrial pumps. These systems, like bearings and gearboxes, display non-stationary vibration characteristics that benefit from the detailed temporal and spectral insights provided by TFR analysis. However, extending the approach to different types of machinery may require tailored adjustments. Machinery-specific fault patterns, structural dynamics, and operating conditions will likely influence the optimal choice of TFR parameters. In these scenarios, tuning time-frequency parameters can capture the relevant fault features and maintain high classification performance.

6.2. Quality, representativeness, and computational efficiency analysis

Evaluating the quality and representativeness of the different TFRs offers critical insights into how each TFR method performs in terms of accuracy, energy concentration, and computational cost associated with each approach. This evaluation reveals necessary trade-offs between performance, computational efficiency, and the representational quality of each method. The results are shown in Table 6.

Table 6 Quality and representativeness evaluation results of various TFRs

Data transform (TFR)	Average training time (sec) per fold	CM (Renyi entropy)	EC (Esignal/ETFR)
Spectrogram	83.8063	9.16	114.99/129
Scalogram	6,449.19	15.46	114.99/201
MS-transform	2,739.23	14.31	114.99/127
SPWVD	>10,000	14.92	114.99/400
HHT	5,209.74	10.59	114.99/6.76

The spectrogram, with a CM value of 9.16, has a moderate level of signal concentration, leaning more toward a focused energy distribution. This focused concentration likely contributes to its good accuracy, especially given that it retains a significant proportion of the original signal's energy as indicated by its EC ratio of 114.99/129, meaning most of the signal's energy is preserved in the transformed domain. Additionally, its low average training time of 83.8063 seconds per fold makes the spectrogram a highly efficient method regarding computational cost, balancing accuracy, and speed.

On the other hand, the scalogram exhibits the highest CM value of 15.46, indicating that its energy is the most spread out among the methods. While this broad spread theoretically allows for capturing a wide range of signal features, the EC ratio of 114.99/201 shows that it loses more energy than the spectrogram and MS-transform. This lower energy preservation might contribute to the scalogram's slightly lower accuracy. Moreover, the 6,449.19 seconds per fold training time makes it one of the most computationally expensive methods, indicating a trade-off between signal capture breadth and practicality regarding computational resources.

The MS-transform, which demonstrated one of the best accuracies in previous analyses, achieves a CM value of 14.31, reflecting a well-distributed energy profile, slightly less spread out than the scalogram but more than the spectrogram. The EC ratio of 114.99/127 shows that it conserves a significant portion of the signal's energy, supporting its superior classification performance. With a training time of 2,739.23 seconds, the MS-transform is computationally more expensive than the spectrogram but more efficient than the scalogram. This middle ground between energy spread, EC, and computational time explains why the MS-transform consistently delivers high accuracy while remaining relatively computationally efficient.

The SPWVD, with a CM value of 14.92, exhibits a relatively spread-out energy distribution similar to the MS-transform and scalogram. However, its EC ratio of 114.99/400 reveals a considerable signal energy loss during transformation, likely contributing to its underperformance in classification accuracy. The excessively high training time (exceeding 10,000 seconds) further reduces its practicality, indicating that the trade-off between energy spread and EC is unfavorable for this method in real-world applications.

Finally, the HHT presents an interesting case with a CM value of 10.59, indicating a more concentrated energy distribution than the scalogram, MS-transform, and SPWVD. However, its EC ratio of 114.99/6.76 is the lowest by a significant margin, indicating that HHT loses the most signal energy during transformation. This substantial energy loss, coupled with a training time of 5,209.74 seconds, explains its weak classification performance, as much of the signal's essential information is lost during the transformation process.

The time complexity and accuracy trade-off analysis provides a critical perspective on the efficiency and performance of the TFR techniques. This analysis reveals the inherent compromises between computational cost and classification performance, as depicted in Fig. 7. Practical deployment of fault diagnosis models often requires real-time performance. In this context, the feasibility of applying computationally intensive TFR methods becomes a critical concern. Although these methods yield high classification accuracy, they can introduce significant processing latency, which, without further optimization, makes them less suitable for direct use in embedded or real-time environments. More efficient TFRs that balance accuracy and computational efficiency should be favored for time-sensitive applications.

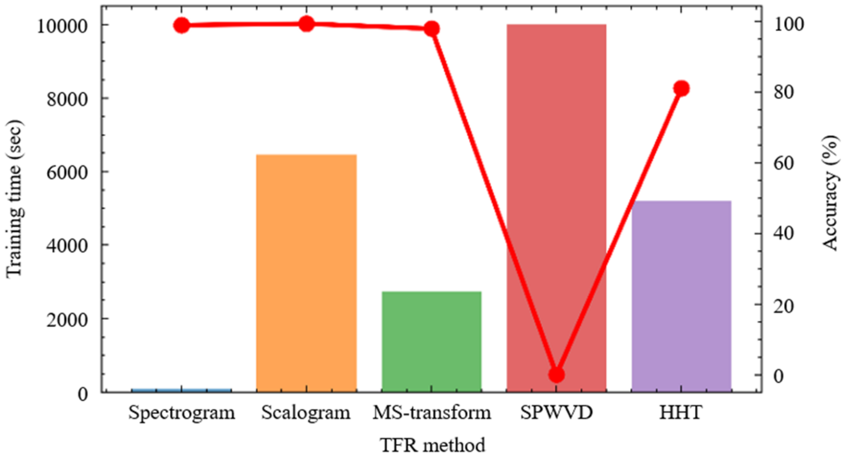


Fig. 7 Time complexity versus accuracy trade-off visualization

6.3. *Noise sensitivity analysis*

This section analyzes the sensitivity of fault diagnosis performance to various degrees of noise. Fig. 8 presents the accuracy of four TFR methods (spectrogram, scalogram, MS-transform, and HHT) across different SNRs of 0, 5, 10, and 15 dB, respectively. The spectrogram consistently performs well, starting at 75.09% accuracy at an SNR of 0 dB and increasing steadily to 99.07% at an SNR of 15 dB, indicating its robustness in noisy environments. The scalogram shows a similar trend but with slightly lower performance at lower SNR values, starting at 56.99% at an SNR of 0 dB before matching and slightly surpassing the spectrogram at higher noise levels, reaching 99.34% at an SNR of 15 dB.

The MS-transform exhibits the poorest performance, particularly at lower SNRs, where its accuracy is only 26.84% at an SNR of 0 dB, improving to 69.33% at an SNR of 15 dB but still significantly lagging behind the other methods. The HHT performs better than the MS-transform, starting at 40.48% accuracy at an SNR of 0 dB and showing a significant improvement as the SNR increases, reaching 83.61% at an SNR of 15 dB. Overall, the spectrogram and scalogram emerge as the most resilient and effective methods for handling noise, whereas the MS-transform struggles with noisy data, and HHT shows moderate performance in comparison.

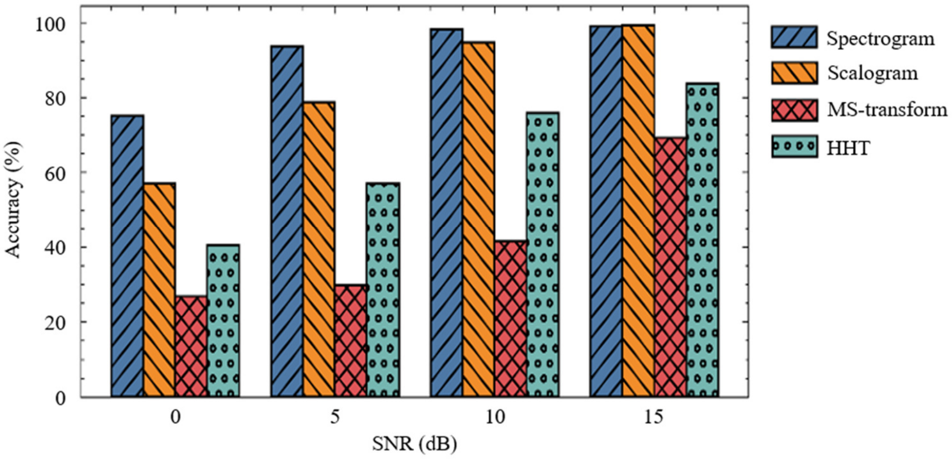


Fig. 8 Comparative accuracy of four TFRs across varying SNR levels

7. Conclusions

This paper presents a comprehensive comparative study of five widely used TFRs, spectrogram, scalogram, MS-transform, SPWVD, and HHT, for DL-based fault diagnosis in rotating machinery under various operating conditions. The classification accuracy, TF energy concentration, and computational cost of each TFR are quantitatively evaluated using datasets from bearings and gearboxes. The main findings are summarized as follows:

(1) The MS-transform achieved the highest classification accuracy of 99.87% under ideal conditions, demonstrating superior feature representation. However, its accuracy dropped significantly to 69.33% under noisy conditions (SNR = 15 dB). In contrast, the spectrogram and scalogram showed better noise robustness, maintaining accuracy above 99% at higher SNRs, thereby making them suitable for noisy environments. SPWVD and HHT exhibited comparatively lower accuracy and computational efficiency.

(2) The analyses indicate that the MS-transform is ideal for tasks requiring high accuracy and reliability under noise-free conditions. In contrast, the spectrogram and scalogram provide competitive alternatives across different noise levels. Conversely, SPWVD has limitations and should be applied cautiously in classification tasks.

(3) The Renyi entropy and EC metrics explain why certain TFR methods excel in classification. The spectrogram stands out due to its computational efficiency, balanced energy distribution, and high energy conservation, which support its robust performance. These findings underscore the importance of energy retention for accurate classification and provide valuable guidance for signal processing and fault diagnosis applications. They also highlight the necessity to tailor TFR methods to the specific characteristics of each dataset.

Although this study offers a comprehensive comparative analysis demonstrating the effectiveness of five TFR methods under various conditions and noise levels, it is subject to certain limitations. Only default parameter settings were used for the TFRs, and a lightweight CNN model was employed consistently, while alternative architectures may yield improved results. Furthermore, only Gaussian noise was considered in robustness tests, leaving the impact of other noise types unexplored. Future work will focus on exploring optimal parameter combinations for each TFR, fine-tuning these parameters, evaluating performance across different DL models and other industrial datasets, and investigating the effects of various noise types on model robustness.

Acknowledgment

This research was supported by the Qilu Institute of Technology Research Grant No. (QIT24TP029).

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] I. D. P. Karyatanti, A. Noersena, F. R. Purnomo, R. S. Zulkifli, and A. Wijayanto, “Analysis of Outer Race Bearing Damage by Calculation of Sound Signal Frequency Based on the FFT Method,” International Journal of Engineering and Technology Innovation, vol. 13, no. 1, pp. 28-39, 2023.

[2] X. Zhang, L. Jiang, L. Wang, T. Zhang, and F. Zhang, “A Pruned-Optimized Weighted Graph Convolutional Network for Axial Flow Pump Fault Diagnosis with Hydrophone Signals,” Advanced Engineering Informatics, vol. 60, article no. 102365, 2024.

[3] X. Zhang, J. Liu, X. Zhang, and Y. Lu, “Self-Supervised Graph Feature Enhancement and Scale Attention for Mechanical Signal Node-Level Representation and Diagnosis,” Advanced Engineering Informatics, vol. 65, part A, article no. 103197, 2025.

[4] A. Raqib, H. Jahanger, M. Sumner, T. Cox, T. Klumpner, I. Agha, et al., “Comparison of Frequency Transform Methods for Condition Monitoring of a DC-Link Capacitor,” Proceedings of the 10th International Conference on Power Electronics, Machines and Drives, vol. 2020, no. 7, pp. 621-626, 2020.

[5] D. K. B. Kulevome, H. Wang, and X. Wang, “Deep Neural Network Based Classification of Rolling Element Bearings and Health Degradation through Comprehensive Vibration Signal Analysis,” Journal of Systems Engineering and Electronics, vol. 33, no. 1, pp. 233-246, 2022.

[6] G. Wu, G. Wang, X. Liu, and R. Duan, “Bearing Fault Diagnosis Method Based on STFT Image and AlexNet Network,” International Conference on the Efficiency and Performance Engineering Network, pp. 1056-1068, 2022.

[7] G. Xin, Z. Li, L. Jia, Q. Zhong, H. Dong, N. Hamzaoui, et al., “Fault Diagnosis of Wheelset Bearings in High-Speed Trains Using Logarithmic Short-Time Fourier Transform and Modified Self-Calibrated Residual Network,” IEEE Transactions on Industrial Informatics, vol. 18, no. 10, pp. 7285-7295, 2022.

[8] Q. Zhang and L. Deng, “An Intelligent Fault Diagnosis Method of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network,” Journal of Failure Analysis and Prevention, vol. 23, no. 2, pp. 795-811, 2023.

[9] R. F. Ribeiro Junior, I. A. dos Santos Areias, M. M. Campos, C. E. Teixeira, L. E. B. da Silva, and G. F. Gomes, “Fault Detection and Diagnosis in Electric Motors Using Convolution Neural Network and Short-Time Fourier Transform,” Journal of Vibration Engineering & Technologies, vol. 10, no. 7, pp. 2531-2542, 2022.

[10] D. K. B. Kulevome, H. Wang, and X. Wang, “Rolling Bearing Fault Diagnostics Based on Improved Data Augmentation and ConvNet,” Journal of Systems Engineering and Electronics, vol. 34, no. 4, pp. 1074-1084, 2023.

[11] W. Fan, Q. Zhou, J. Li, and Z. Zhu, “A Wavelet-Based Statistical Approach for Monitoring and Diagnosis of Compound Faults with Application to Rolling Bearings,” IEEE Transactions on Automation Science and Engineering, vol. 15, no. 4, pp. 1563-1572, 2018.

[12] X. Zhang, X. Zhang, J. Liu, B. Wu, and Y. Hu, “Graph Features Dynamic Fusion Learning Driven by Multi-Head Attention for Large Rotating Machinery Fault Diagnosis with Multi-Sensor Data,” Engineering Applications of Artificial Intelligence, vol. 125, article no. 106601, 2023.

[13] H. K. R. Pillai, M. Nanappan, M. V. Prabhakaran, and S. P. Sathyabhama, “A Robust Technique for Detection, Diagnosis, and Localization of Switching Faults in Electric Drives Using Discrete Wavelet Transform,” International Journal of Engineering and Technology Innovation, vol. 13, no. 1, pp. 14-27, 2023.

[14] M. Q. Tran, M. K. Liu, Q. V. Tran, and T. K. Nguyen, “Effective Fault Diagnosis Based on Wavelet and Convolutional Attention Neural Network for Induction Motors,” IEEE Transactions on Instrumentation and Measurement, vol. 71, article no. 3501613, 2022.

[15] R. Yan, Z. Shang, H. Xu, J. Wen, Z. Zhao, X. Chen, et al., “Wavelet Transform for Rotary Machine Fault Diagnosis: 10 Years Revisited,” Mechanical Systems and Signal Processing, vol. 200, article no. 110545, 2023.

[16] D. K. B. Kulevome, H. Wang, B. M. Cobbinah, E. S. Mawuli, and R. Kumar, “Effective Time-Series Data Augmentation with Analytic Wavelets for Bearing Fault Diagnosis,” Expert Systems with Applications, vol. 249, part A, article no. 123536, 2024.

[17] J. Zheng, J. Wang, H. Wang, J. Ding, and C. Yi, “Diagnosis and Classification of Gear Composite Faults Based on S-Transform and Improved 2D Convolutional Neural Network,” International Journal of Dynamics and Control, vol. 12, no. 6, pp. 1659-1670, 2024.

[18] Y. Peng and X. Ma, “A Fault Diagnosis Method of Rolling Bearings Based on Parameter Optimization and Adaptive Generalized S-Transform,” Machines, vol. 10, no. 3, article no. 207, 2022.

[19] Y. Xu, L. Wang, A. Hu, and G. Yu, “Time-Extracting S-Transform Algorithm and Its Application in Rolling Bearing Fault Diagnosis,” Science China Technological Sciences, vol. 65, no. 4, pp. 932-942, 2022.

[20] Z. Cui, G. Yu, and W. Tian, “S-Transform Based Time–Frequency Analysis Tool with Application to Detection of Bearing Fault,” Journal of Physics: Conference Series, vol. 2352, no. 1, article no. 012004, 2022.

[21] H. Wang, Z. Fang, H. Wang, Y. Li, Y. Geng, L Chen, et al., “A Novel Time-Frequency Analysis Method for Fault Diagnosis Based on Generalized S-Transform and Synchroextracting Transform,” Measurement Science and Technology, vol. 35, no. 3, article no. 036101, 2023.

[22] W. Liu, Y. Liu, S. Li, and Z. Zhai, “Demodulated Synchrosqueezing S-Transform and Its Application to Machine-Fault Diagnosis,” Measurement Science and Technology, vol. 34, no. 6, article no. 065004, 2023.

[23] C. Yi, J. Qin, H. Xiao, and T. Zhou, “Second-Order Synchrosqueezing Modified S Transform for Wind Turbine Fault Diagnosis,” Applied Acoustics, vol. 189, article no. 108614, 2022.

[24] X. Lv and H. Li, “Rolling Bearing Fault Diagnosis Based on GWVD and Convolutional Neural Network,” 19th International Conference on Advanced Intelligent Computing Technology and Applications, pp. 514-523, 2023.

[25] H. Li, Q. Zhang, X. Qin, and Y. Sun, “K-SVD-Based WVD Enhancement Algorithm for Planetary Gearbox Fault Diagnosis under a CNN Framework,” Measurement Science and Technology, vol. 31, no. 2, article no. 025003, 2019.

[26] D. Quan, F. Ren, X. Wang, M. Xing, N. Jin, and D. Zhang, “WVD-GAN: A Wigner-Ville Distribution Enhancement Method Based on Generative Adversarial Network,” IET Radar, Sonar & Navigation, vol. 18, no. 6, pp. 849-865, 2024.

[27] Z. Li, S. Sun, Y. Wang, W. Dai, J. Qiu, and Z. Liu, “Time-Frequency Analysis of Non-stationary Signal Based on Sliding Mode Singular Spectrum Analysis and Wigner-Ville Distribution,” 3rd International Conference on Information Science and Education, pp. 218-222, 2022.

[28] E. Elbouchikhi, V. Choqueuse, Y. Amirat, M. E. H. Benbouzid, and S. Turri, “An Efficient Hilbert–Huang Transform-Based Bearing Faults Detection in Induction Machines,” IEEE Transactions on Energy Conversion, vol. 32, no. 2, pp. 401-413, 2017.

[29] G. Cheng, Y. L. Cheng, L. H. Shen, J. B. Qiu, and S. Zhang, “Gear Fault Identification Based on Hilbert–Huang Transform and SOM Neural Network,” Measurement, vol. 46, no. 3, pp. 1137-1146, 2013.

[30] M. Zabin, H. J. Choi, J. Uddin, M. H. Furhad, and A. B. Ullah, “Industrial Fault Diagnosis Using Hilbert Transform and Texture Features,” IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 121-128, 2021.