A Review of Time–Frequency Analysis Methods for Feature Engineering

Pierre Lague

[1] Inria, Rennes (France)

June 5, 2026

1 Introduction

Vibration signal analysis has established itself as a cornerstone methodology in rotating machinery condition monitoring and fault diagnosis. The diagnostic objective centers on early detection of mechanical degradation before catastrophic failure occurs, enabling predictive maintenance strategies that reduce downtime and operational costs. Traditional approaches to vibration analysis have historically operated within single domains: either analyzing signals exclusively in the time domain (examining amplitude variations and statistical features) or in the frequency domain (using Fourier analysis to identify characteristic frequencies associated with specific fault types). However, machinery faults manifest as transient, nonstationary phenomena where fault signatures exhibit both temporal localization and frequency content that evolve over time. This temporal-spectral dynamics render single-domain approaches inadequate, particularly during early fault stages where signatures remain weak and diffuse.

Over the past decae, time-frequency analysis has emerged as the dominant framework for vibration signal characterization in machinery diagnostics. Rather than selecting between time and frequency information, time-frequency methods simultaneously represent how frequency content evolves through time, revealing fault transients that remain obscured in purely temporal or spectral representations. The evolution of this field has been documented across multiple comprehensive reviews: Bagri et al. (2024) [1] synthesized 270 published articles on rotating machinery diagnosis, establishing that signal decomposition combined with time-frequency analysis constitutes essential methodology for accurate feature extraction. Similarly, systematic reviews by Li et al. (2024) [2] , Kannan et al. (2024) [3], and recent analysis in EIII and MDPI journals consistently identify time-frequency representation as the critical preprocessing stage between raw signal acquisition and machine learning-based inference.

Before 2014, review such as the ones proposed by Neild et al. (2003) [4] and Feng et al. (2013) [5] presented major methods, advantages and disadvantages for fault diagnosis of machinery or structural vibration analysis, that are still of use to this day or have been derived over the last decade. The transformation from 2014 to 2025 reveals a fundamental methodological shift. Early work (2014-2017) treat time-frequency analysis as an exploratory tool, with researchers extracting hand-crafted numerical features (spectral moments, kurtosis, entropy measures) from time-frequency plots. By 2018-2021, deep learning methodologies catalyzed a new framework [6]: time-frequency representations converted to image format served as direct inputs to convolutional neural networks, eliminating manual feature engineering, but loosing in interpretability. Contemporary approaches (2022-2025) further advance this trajectory through multi-source time-frequency fusion [7](combining complementary transform methods), physics-informed architectures embedding time-frequency transforms within neural networks [8], and self-supervised learning frameworks [9] that extract features from time-frequency representations with minimal labeled data.

This chapter reviews the current landscape of time-frequency analysis methods employed for vibration signal feature engineering. The presentation synthesizes recent advances documented in systematic reviews and state-of-the-art publications while emphasizing the methodological foundations that enable effective fault diagnosis

under realistic operating conditions (variable speed, load, temperature). The figures proposed were inspired by those presented by Feng et al. (2013) [5], we reach similar results for our use of the various transforms.

2 Feature Engineering Based on Time-Frequency Analysis Methods

Traditional signal analysis operates within fundamental constraints. Time-domain analysis (direct examination of signal amplitude versus time) provides excellent temporal localization but reveals little about frequency content. Frequency-domain analysis (Fourier transform) decomposes signals into sinusoidal components, identifying which frequencies are present but completely losing temporal information—the Fourier transform cannot indicate when specific frequencies occur. For stationary signals with constant frequency content, this limitation proves unproblematic. However, machinery vibration signals are inherently nonstationary: fault signatures emerge at specific times (when a bearing race spall contacts a rolling element) with specific frequency content (determined by bearing geometry), then persist for fractions of a second before attenuating.

Time-frequency representations address this fundamental tension by providing simultaneous local information in both time and frequency dimensions. Mathematically, this requires accepting the Heisenberg uncertainty principle's constraint: perfect localization in both time and frequency simultaneously is impossible. Different time-frequency methods resolve this tradeoff differently, trading temporal resolution against frequency resolution and vice versa [10]. This fundamental asymmetry becomes crucial in feature engineering: selecting appropriate time-frequency methods requires understanding signal characteristics and diagnostic objectives.

The utility of time-frequency representation in machinery diagnostics rests on a straightforward principle: different fault types produce characteristic signatures visible only when time and frequency information are simultaneously analyzed. A vehicle's chassis, for instance, produces impacts occurring at it's natural frequency (determined by the beams it is made of and their material), generating energy concentrated in narrow frequency bands that appear as repeated impulses in the time domain. These impulses become clearly visible in time-frequency representations where one observes periodic bursts of energy at specific frequencies, yet remain obscured when examining either time or frequency domains separately.

2.1 Core Time-Frequency Transforms for Vibration Analysis

2.1.1 Continuous Wavelet Transform

The continuous wavelet transform (CWT) employs wavelets, instead of sinusoidal functions, as the basis. It adds a scale variable in addition to the time variable in the inner product transform. Hence, it is effective for time–frequency localization, and is suited to transient signal analysis [11, 12]:

For any enerhy limited signal $x(t) \in L^2(R)$, the wavelet transform can be defined as

$$WT_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(\tau)\Psi^*(\frac{\tau - t}{a})d\tau \tag{1}$$

where wavelet $\Psi^*(\frac{\tau-t}{a})$ is derived by dilating and translating the wavelet basis, or "mother" wavelet, $\Psi(t)$, a is the scale parameter ($a > 0$), t is the time shift, and $1/\sqrt{a}$ is a normalization factor to maintain energy conservation [5].

Scales corresponding to lower values of a (high frequencies) automatically obtain narrow temporal support, while scales corresponding to higher values of a (low frequencies) obtain broader temporal support. This automatic frequency-dependent resolution proves highly advantageous for machinery signals: system impacts (containing high-frequency components) are localized to precise time instants, while modulation effects (containing low-frequency components) appear with broader temporal extent, matching the physical characteristics of machinery faults.

CWT's flexibility in wavelet basis selection—Morlet wavelet, Mexican hat, Daubechies wavelets, reverse biorthogonal wavelets—permits tailoring to specific applications. The CWT output forms a 2D time-scale matrix (or equivalently, time-frequency matrix after scale-frequency conversion) called a scalogram, where each cell's magnitude represents the coefficient magnitude at that time-frequency location.

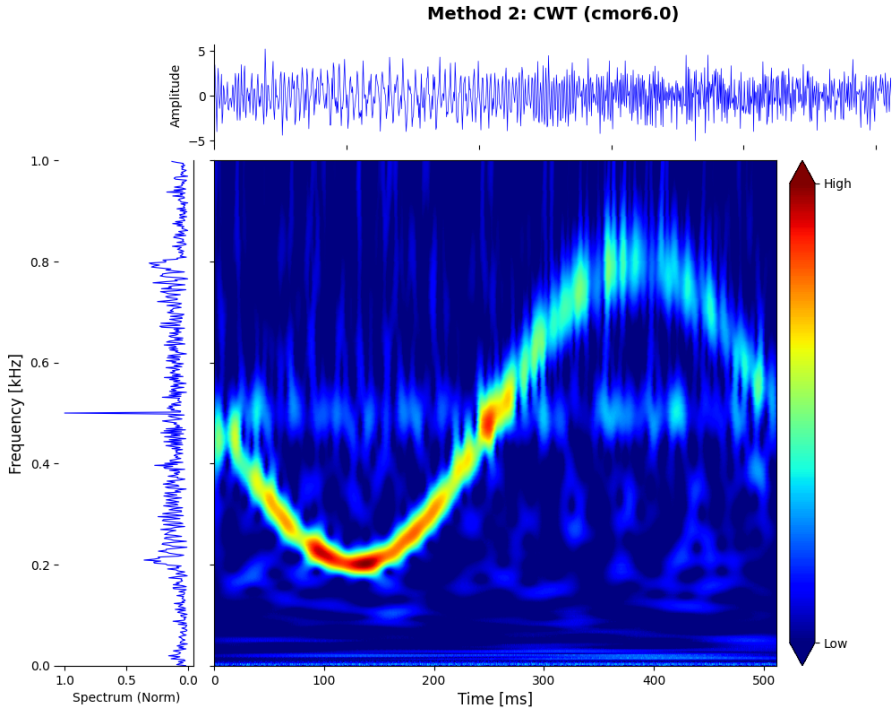

Figure 1: Continuous Wavelet Transform using the Morlet Wavelet.

Fig. 1 shows the continuous wavelet transform scalogram of our synthetic signal 1. This signal contains a pure 500 Hz tone and a frequency-modulated component (500 Hz carrier with 1.95 Hz modulation). White noise is added to test robustness. Signal parameters are fs = 2000 Hz and N = 1024 samples. The Continuous Wavelet Transform uses a complex Morlet wavelet (cmor6.0-1.0) with 500 frequency scales spanning 1 Hz to 1000 Hz for high-resolution time-frequency decomposition. The signal is given by this equation:

$$x(t) = \sin(2\pi f_{sine} t) + 2\cos(2\pi f_{carrier} t + 153.6\cos(2\pi f_{FM} * t)) + n(t) \tag{2}$$

We can see that the CWT identifies the two components, but at a low resolution due to the Heisenberg uncertainty principle. In addition, for the higher frequency part of the sinusoidal FM component (in the interval from 350 ms to 450 ms), the time localization is fine but the frequency resolution is low. Although the signal has only two true components, it is adequate to mimic the harmonic and time-varying vibration components often encountered in applications. In particular, the instantaneous frequency of the sinusoidal FM component is highly nonlinear with time, and intersects with that of the sinusoidal wave. Hence, the signal is complex enough in terms of time–frequency structure to illustrate the performance of multiple time–frequency analysis method. On the subsequent figures, the signal waveform is on the top sub-plot, the corresponding power spectrum on the left, the time–frequency analysis result in the middle, and a color bar indicating the magnitude levels on the right. This presentation scheme will be used for all the time–frequency analysis figures hereafter in this paper

Contemporary applications extend beyond simple scalogram visualization. Makrouf et al. (2025) [13] demonstrated that CWT scalograms, when treated as images and fed to convolutional neural networks, enable automated feature extraction superior to hand-designed features. Multi-axis applications have further expanded CWT utility: Wang et al. (2025) [14] showed that generating separate CWT scalograms for triaxial accelerometer data (X, Y, Z axes) and fusing them as RGB image channels produces multi-dimensional feature representations capturing fault signatures across spatial dimensions. This multi-dimensional approach achieved superior cross-condition generalization, suggesting that combining time-frequency information from multiple measurement axes provides more robust fault representation than single-axis analysis.

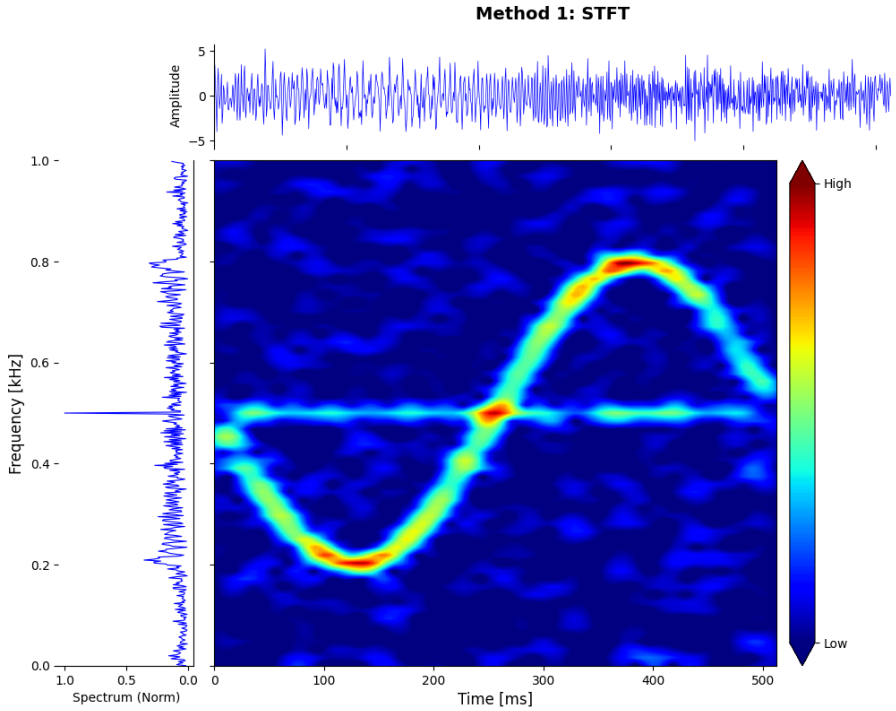

Figure 2: Short time Fourier Transform spectrogram.

2.1.2 Short-Time Fourier Transform (STFT)

The short-time Fourier transform applies Fourier analysis within sequential narrow time windows, computing the FFT independently within each window. This approach maintains the familiar frequency interpretation of Fourier analysis while introducing temporal information through windowing [15]. With this method, a short time window moves along time to slice a signal. In a short duration, the segmented signal usually does not change too much and hence can be assumed to be stationary. By applying the Fourier transform, the local Fourier spectrum of each segment around the time location of the short time window can be obtained and the time variation of the signal can be revealed according to the local spectrum [16].

For any temporal signal $x(t)$, the STFT is defined as follows:

$$STFT_x(t,f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)exp(-j2\pi f_\tau)d\tau \tag{3}$$

where $w(\tau - t)$ represents a time window centered at time t. The window length determines the time-frequency resolution trade-off: narrow windows provide excellent temporal resolution but degrade frequency resolution (few frequency bins), while broad windows provide fine frequency resolution but poor temporal localization. Unlike CWT's frequency-dependent resolution, STFT maintains constant time-frequency resolution across all frequencies, because the same window length applies at all scales.

To illustrate, we use signal 1, generated earlier. The Short-Time Fourier Transform is computed using a Hann window (nperseg=128) with 93.75% overlap to analyze the time-varying spectral content. Fig. 2 shows that though the STFT can resolve the time-frequency structure of the two signal components, the resolution is low.

Despite this fundamental limitation compared to CWT, STFT offers computational advantages critical for real-time embedded systems. STFT computation requires $O(N \log N)$ operations (via FFT), while CWT computation scales as $O(N \times M)$ where M is the number of scales analyzed. Wang et al. (2025) [14] leveraged this computational efficiency, demonstrating that STFT-based time-frequency images enable semi-supervised fault diagnosis achieving over 99% accuracy with minimal labeled samples (50 per fault type). This efficiency makes STFT preferable for resource-constrained platforms and real-time diagnostic systems.

STFT's fixed time-frequency resolution also provides practical advantages: the rectangular structure of spec-

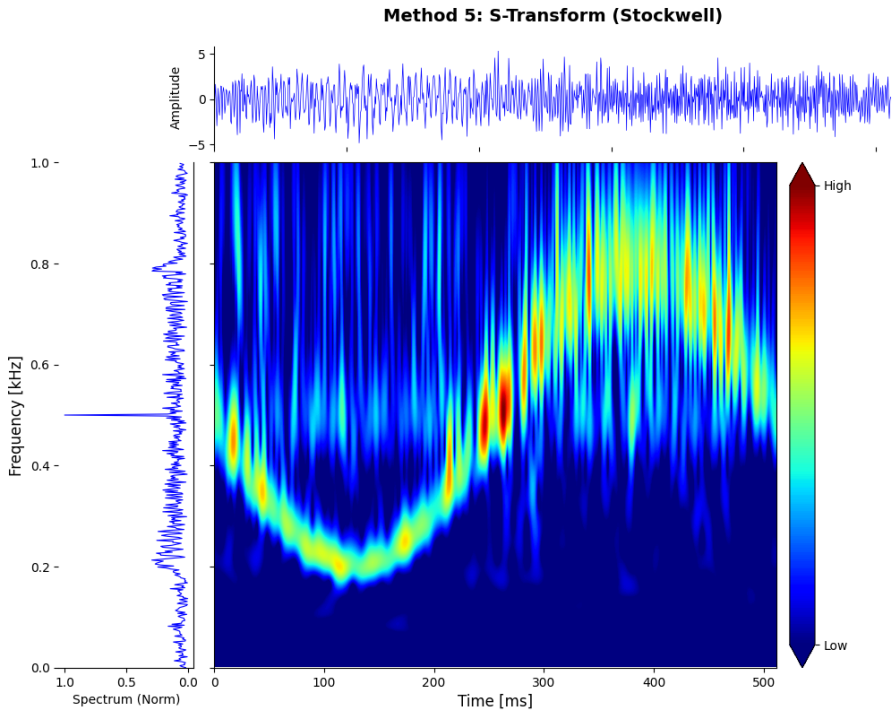

Figure 3: Stockwell transform spectrogram of signal 1.

trograms aligns naturally with standard image sizes (e.g. 256×256, 512×512) required by CNN architectures, requiring minimal interpolation. Methods by Alikhani et al. (2025) [17] or Zhang et al. (2023) [18] increasingly employ STFT for this practical convenience, accepting the fundamental resolution tradeoff to gain computational efficiency and implementation simplicity.

2.1.3 Stockwell Transform (S-Transform)

The Stockwell transform represents a hybrid approach combining STFT's computational characteristics with CWT's adaptive resolution [19]. The S-transform generates a time-frequency representation with frequency-dependent resolution: high frequencies receive narrow time windows (excellent temporal localization) while low frequencies receive broad time windows (excellent frequency localization). This frequency-dependent resolution naturally aligns with machinery fault characteristics where impact transients contain high-frequency components requiring precise temporal localization, while load-modulation effects contain low-frequency components requiring frequency precision.

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{(-(\frac{(\tau-t)^2 f^2}{2})} e^{-j2\pi f t} dt \tag{4}$$

The S-Transform analysis on Fig. 3 reveals a distinctive parabolic energy distribution in the lower frequency band (150-400 Hz) during the first 300 ms, forming a characteristic "U-shaped" pattern. This structure reflects the frequency-modulated component's instantaneous frequency trajectory, where the large modulation index (β = 153.6) produces significant frequency deviation around the 500 Hz carrier. A strong energy concentration appears at 500 Hz throughout the entire signal duration, corresponding to the superposition of both the pure sinusoidal tone and the FM carrier frequency. The time-frequency representation exhibits periodic vertical striations after approximately 300 ms, consistent with the 1.95 Hz modulation frequency (period ≈ 513 ms). The S-Transform's frequency-dependent resolution is evident: lower frequencies display enhanced frequency resolution revealing the detailed FM structure, while higher frequencies show improved temporal localization. The adaptive windowing successfully captures both the rapid frequency variations of the modulated component and the stationary spectral content at 500 Hz, demonstrating the method's effectiveness for analyzing non-stationary signals with complex time-frequency characteristics.

The Gaussian window's width adjusts inversely with frequency, automatically providing optimal resolution distribution for machinery signals. Zaman et al. (2023) [20] applied S-transform scalograms with edge enhancement (Sobel filtering) to highlight fault feature boundaries in time-frequency representations. Edge-enhanced scalograms revealed fault signatures more distinctly, improving subsequent CNN feature extraction. Li et al. (2019) [21] propose a novel sensor data-driven fault diagnosis method is proposed by fusing S-transform (ST) algorithm and CNN, namely ST-CNN. They design a layer based on the ST algorithm and feed it 2D time-frequency matrix. Experimental results show that it is more robust and outperforms existing methods for fault diagnosis based on CNN models.

These approaches demonstrated that applying computer vision techniques (edge detection, texture analysis, convolution) to time-frequency images further enhances feature clarity, suggesting valuable synergies between signal processing and image analysis domains.

2.1.4 Hilbert-Huang Transform (HHT)

The Hilbert-Huang transform employs empirical mode decomposition (EMD) to decompose signals adaptively into intrinsic mode functions (IMFs), then applies Hilbert transform to extract instantaneous frequency and amplitude [22]. Unlike predetermined basis functions employed in CWT, STFT, and S-transform, EMD provides completely data-driven decomposition adapting to signal characteristics.

EMD operates through iterative sifting: repeatedly identifying local maxima and minima, interpolating through them to generate upper and lower envelopes, then subtracting their mean from the signal to obtain the first IMF. This process repeats on the residual until obtaining a monotonic trend component. Subsequent IMFs are extracted from successive residuals. This adaptive decomposition accommodates highly nonlinear signal components that fixed-basis methods may poorly represent.

Following decomposition, the Hilbert transform applied to each IMF yields instantaneous frequency and amplitude (the analytic signal). Olalere et al. (2023) [23] demonstrated HHT feature extraction generating interpretable quantities (instantaneous energy, frequency, amplitude at multiple scales) that could be further optimized via genetic algorithms to identify most discriminative features. This interpretability—extracting quantities with clear physical meaning—distinguishes HHT from purely data-driven deep learning approaches.

However, EMD exhibits well-documented limitations. Mode mixing (where a single IMF contains multiple frequency components) degrades decomposition quality. Ensemble EMD (EEMD) addresses mode mixing by decomposing multiple noise-added signal copies and ensemble-averaging results, improving mode separation at the cost of substantially increased computation [24]. Contemporary variants (CEEMDAN—Ensemble EMD with Adaptive Noise) achieve improved mode separation with fewer decomposition levels [25]. These variants have become standard preprocessing steps in recent machinery diagnostics papers, typically preceding time-frequency analysis of individual IMFs.

To illustrate the performance of the Hilbert–Huang transform, we generate new synthetic signals 2 and 3 according to the following equation:

$$x(t) = \sin(2\pi f_{sine} t) + 2\cos(2\pi f_{carrier} t + 153.6\cos(2\pi f_{FM} t)) \tag{5}$$

For synthetic signal 2, the frequency of the sinusoidal wave and the carrier frequency of the sinusoidal FM component are set to $f_{sine} = f_{carrier} = 125$ Hz and the frequency of the FM component is $f_{FM} = 1.95$ Hz. These frequency parameters are lower than those for synthetic signal 1, because the EMD uses cubic spline interpolation to fit signals and thus needs a higher sampling frequency. In addition, noise is neglected because the EMD is susceptible to noise interferences. Fig. 4 shows the Hilbert energy spectrum of synthetic signal 2. even though the sampling frequency is high enough (2000Hz) for synthetic signal 2 and there is no noise interference, the Hilbert-Huang transform fails to identify the true time-frequency structure of the two components. This is mainly due to the fact that the instantaneous frequency trajectories of the two components have crossings on the time-frequency plane, resulting to mode mixing.

For synthetic signal 3, we set $f_{sine} = 25$Hz, $f_{carrier} = 125$Hz, and $f_{FM} = 1.95$Hz. Here we reduce the frequency of the sinusoidal wave so that its instantaneous frequency trajectory does not intersect with that of the sinusoidal

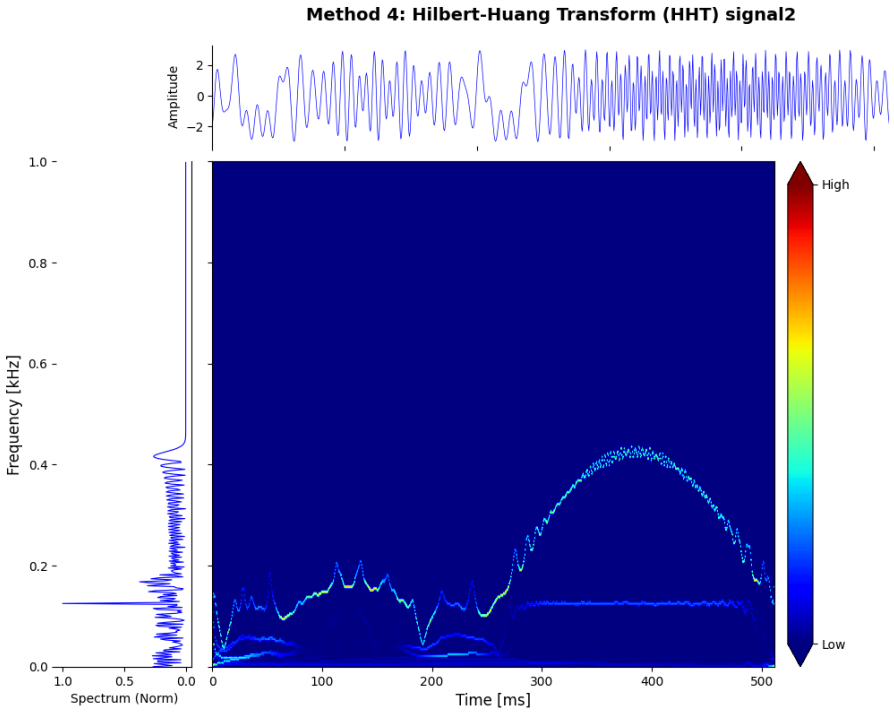

Figure 4: Hilbert energy spectrum of signal 2.

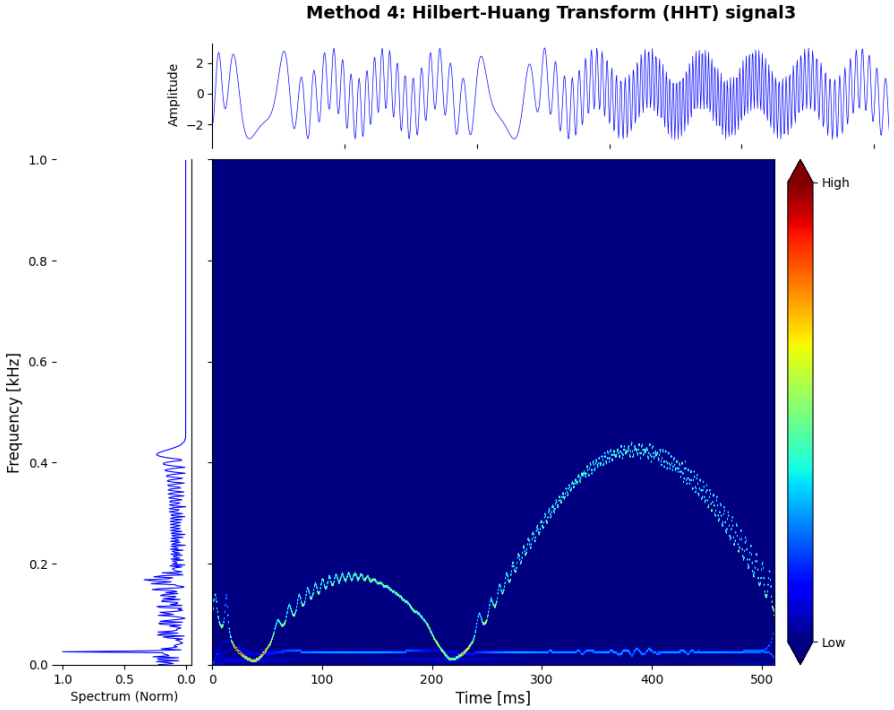

Figure 5: Hilbert energy spectrum of signal 3.

FM component. In that case, two components are clearly identified and separated. Fig. 5 shows the Hilbert energy spectrum of synthetic signal 3. Both the sinusoidal FM component with highly nonlinear instantaneous frequency and the sinusoidal wave of a very low frequency are resolved with fine details. Their instantaneous frequency trajectories exactly follow the theoretical ones.

2.1.5 Wigner-Ville Distribution

For the past 20 years, alternative time-frequency representations have been investigated. The conclusion of these investigations is that the Wigner-Ville distribution (WVD) is the most powerful and fundamental time-frequency representation [26, 27].

The Wigner–Ville distribution is the basis of almost all the bilinear time–frequency distributions. It has the highest time–frequency resolution. However, for multi-component signals, it suffers from the inevitable cross-term interferences, thus is not suitable for many real applications. In order to maintain a higher time–frequency resolution, to obtain a nonnegative distribution, and to suppress the cross-term interferences in the Wigner–Ville distribution, improved versions such as Cohen class distributions, affine class distributions, adaptive optimal kernel method and reassignment method have been proposed [10, 28, 29].

For any signal $x(t)$ the WVD is defined as:

$$WVD_x(t,f) = \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x*(t-\frac{\tau}{2})e^{-2j\pi f\tau}d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_x(\tau,v)e^{-2j\pi(vt+f\tau)}dyd\tau \tag{6}$$

where $A_x(\tau, v)$ is the ambiguity function given by:

$$A_x(\tau,v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x*(t-\frac{\tau}{2})e^{-j2\pi vt}dt \tag{7}$$

As the Wigner–Ville distribution is not linear, the Wigner–Ville distribution of a sum of multiple signal components is not equal to the sum of the Wigner–Ville distributions of these signal components. For example, the Wigner–Ville distribution of signal $x(t) = x_1(t) + x_2(t)$ is

$$WVD_x(t,f) = WVD_{x_1}(t,f) + WVD_{x_2}(t,f) + 2\mathbf{Re}[WVD_{x_1,x_2}(t,f)] \tag{8}$$

where

$$WVD_{x_1,x_2}(t,f) = \int_{-\infty}^{\infty} x_1(t+\frac{\tau}{2})x_2*(t-\frac{\tau}{2})e^{-2j\pi f\tau}d\tau \tag{9}$$

In Equation 8, the first two are auto-terms, and the third one is the cross-term. The cross-term results from the interference between time and frequency due to the nonlinear quadratic transform. It is real, and exists between the auto-terms. Moreover, the cross-term is oscillating and its amplitude is twice that of auto-terms. Each pair of auto-terms results in one cross-term. For multi-component signals in real applications, cross-terms produce a pseudo distribution and may overlap with the auto-terms on the time– frequency plane. In this case, the interference is even more complicated leading to blurred time–frequency features and making it more difficult to interpret the physical meaning of the Wigner–Ville distribution hence the development of higher-order Cohen classes to smooth the interference.

The Wigner–Ville distribution does not involve any window function and thus is free from the interference between time localization and frequency resolution. The product of time interval and frequency bandwidth reaches the lower bound of the Heisenberg uncertainty principle. Consequently it has the highest time–frequency resolution among all the time–frequency distributions. However, for multi-component signals in mechanical systems, its applications are inevitably hindered by the cross-term interferences. Fig. 6 illustrates the Wigner-Ville distribution of synthetic signal 3. Compared to Fig. 5, the definition of the time-frequency components is much more granular, but the presence of interference caused by the cross-term is significant. Even so, the fine

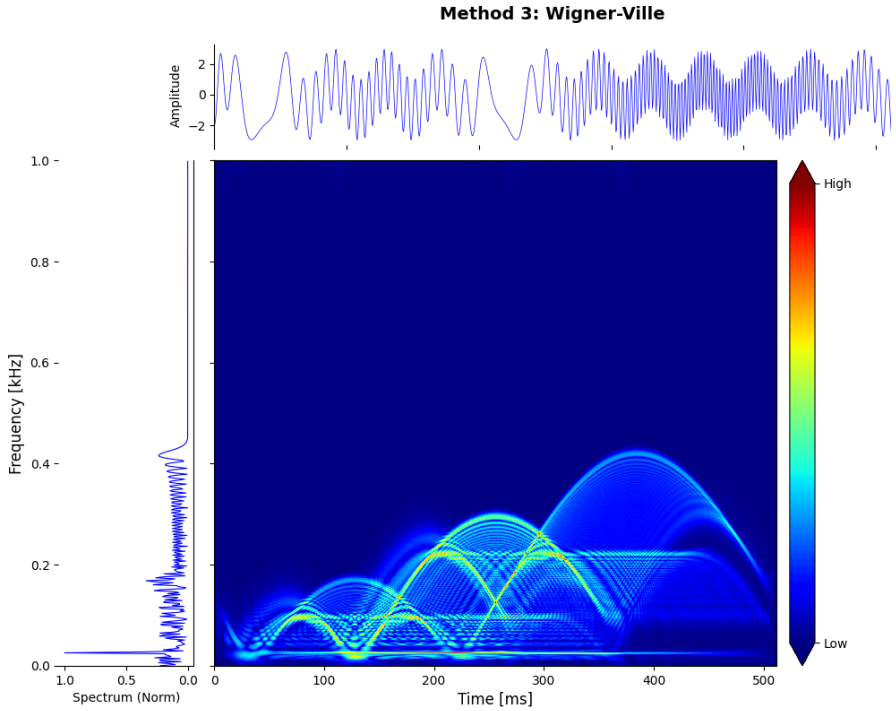

Figure 6: Wigner-Ville distribution of signal 3.

time–frequency resolution of the Wigner–Ville distribution is still very attractive. Furthermore, as the cross-terms also contain information about signal structures, the question of how to effectively exploit the properties of cross-terms to reveal useful information could be an interesting topic.

2.2 Signal Decomposition as Feature Engineering Foundation

Recent practice recognizes that raw vibration signals contain multiple concurrent components (bearing defect impacts, load modulation, noise) with overlapping frequency ranges. Applying time-frequency analysis directly to undecomposed raw signals produces scalograms where multiple components interact through cross-term interference and frequency overlap, complicating fault feature extraction. This recognition has elevated signal decomposition to primary importance in contemporary feature engineering pipelines [30].

2.2.1 Variational Mode Decomposition (VMD)

Variational mode decomposition represents a modern decomposition approach addressing EMD's limitations [31]. VMD formulates decomposition as a constrained optimization problem: decompose the signal into K adaptive components (variational modes) whose center frequencies are determined by optimization, minimizing the sum of mode bandwidth constraints. Unlike EMD's greedy sequential decomposition, VMD determines all modes simultaneously through iterative optimization. It is formulated as follows:

$$\min_{\{u_k\},\{\omega_k\}}\{\sum_k ||\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-j\omega_k t}||_2^2\} \tag{10}$$

subject to $\sum_k u_k = f$. Where u_k and ω_k are shorthand notations for the set of all modes and their center frequencies, respectively. Equally, $\sum_k := \sum_{k=1}^{K}$ is understood as the summation of all modes.

Li et al. (2017) [32] proposed an independence-oriented VMD method via correlation analysis to adaptively extract weak and compound fault feature of wheel set bearing. To overcome the information loss problem, the appropriate mode number is determined by the criterion of approximate complete reconstruction. Then the similar modes are combined according to the similarity of their envelopes to solve the over decomposition problem. This method showed more effective than the original VMD, EMD, and EEMD methods on wheel bearing fault diagnosis.

Liu et al. (2018) [33] propose a novel feature extraction and fault diagnosis method for planetary gears based on variational mode decomposition (VMD), singular value decomposition (SVD), and convolutional neural network (CNN). The authors use VMD to decompose the signal, and each mode had its features extracted using SVD building singular value vector matrices corresponding to the current fault state. These matrices where then fed into a CNN model for gear fault identification and classification. The VMD-based partition extraction method is better than ensemble empirical mode decomposition (EEMD), resulting in a higher CNN total recognition rate of 100% with fewer training times (14 times).

An et al. (2021) [34] introduced parameter-optimized VMD employing multi-objective evaluation functions to guide mode selection. Rather than accepting all decomposed modes, the optimization process identified which modes contained fault information (based on mode distinctiveness, energy distribution, and noise suppression criteria). Following mode selection, envelope spectrum analysis (Hilbert transform followed by FFT) was applied to selected modes, generating features capturing pure fault signatures without noise contamination. This synergistic decomposition + time-frequency approach achieved superior feature quality compared to time-frequency analysis of undecomposed raw signals.

The practical implication extends to feature interpretability: decomposition explicitly separates signal components, enabling examination of fault signatures in isolation. A systems' inner characteristic frequency signature becomes visible in specific modes rather than buried within the complete signal, facilitating both interpretation and feature extraction.

2.3 Ensemble Empirical Mode Decomposition (EEMD) and Variants

Ensemble approaches to EMD emerged to address mode mixing through ensemble averaging [24]. EEMD adds Gaussian noise to the signal at controlled amplitude, decomposes each noise-added copy independently, then ensemble-averages corresponding IMFs. The ensemble averaging process suppresses noise while reducing mode mixing compared to single-pass EMD.

Complete ensemble EMD with adaptive noise (CEEMDAN) further improved the procedure by using adaptive noise strategies where noise is added differently to each stage of decomposition [25]. These improvements reduced the number of required ensemble realizations (typically 50–100 for standard EEMD down to 10–20 for CEEMDAN), significantly accelerating computation while maintaining decomposition quality.

Contemporary machinery diagnostics papers almost universally employ EEMD or CEEMDAN as preprocessing steps. Recent work like Wang et al. (2020) [35] proposed a new criterion to quantify the fault-related degree of a vibration signal, in which the ratio of periodic modulation components caused by fault to the generalized interferences is defined. A reweighted version en CEEMDAN is applied with good de-noising performances. Recent methods typically follow this trend [14, 36] to decompose raw vibration signals, then select a certain of number of dominant modes for analysis (eliminating noise-dominated high-frequency residuals), then applies time-frequency analysis to selected modes. This two-stage approach—adaptive decomposition followed by time-frequency analysis—has become standard practice, reflecting consensus that decomposition fundamentally improves feature quality.

2.4 Multi-Source Time-Frequency Fusion

Contemporary feature engineering increasingly recognizes that different time-frequency methods capture complementary fault information. Rather than selecting a single "optimal" method, advanced systems deliberately employ multiple time-frequency transforms, then fuse their outputs into unified feature representations.

Feng et al. (2025) [7] systematically fused four complementary time-frequency representations: continuous wavelet transform, short-time Fourier transform, Hilbert-Huang transform, and Wigner-Ville distribution. CWT provides automatic scale selection with flexible wavelet basis; STFT offers computational efficiency with fixed resolution; HHT delivers completely adaptive decomposition capturing nonlinear components; Wigner-Ville distribution supplies highest time-frequency resolution (quadratic time-frequency representation). By integrating features from all methods, the approach captured fault signatures in multiple complementary representations.

Gong et al. (2024) [37] proposed a novel hierarchical vision transformer (NHVT) and wavelet time–frequency

architecture combined with a multi-source information fusion (MSIF) strategy to boost stable performance by extracting and integrating rich features. The goal was to improve the end-to-end fault diagnosis performance of mechanical components. First, multi-source signals are transformed into two-dimensional time and frequency diagrams. Then, a novel hierarchical vision transformer is introduced to improve the nonlinear representation of feature maps to enrich fault features. Next, multi-source information diagrams are fused into the proposed NHVT to produce more comprehensive presentations. Finally, the proposed model was compared to state-of-the-art models on multi-source datasets and experimental results show it is capable of extracting useful features.

Yu et al. (2025) [38] introduced MSIFT, an end-to-end mechanical fault disgnosis framework under limited and imbalanced data using multi-source information fusion. The method first applies time–frequency preprocessing to capture rich signal representations, then extracts multi-scale, information-invariant features from multiple sources to reduce domain shift. These features are fused using a transformer-based network with cross-attention, followed by global pooling to obtain representative global features. Finally, a dual-stream predictor combining binary and multi-class diagnosis integrates results via a reweighting mechanism, achieving superior diagnostic performance across multiple multi-source datasets.

Generally speaking, the multi-source fusion framework operates as follows: (1) apply each time-frequency method independently to the same vibration signal; (2) normalize resulting representations to uniform scale and range; (3) stack normalized representations as multi-channel inputs (analogous to RGB color images); (4) feed multi-channel time-frequency images to CNN architectures with channel-attention mechanisms; (5) allow learned attention layers to adaptively weight each time-frequency method's contribution for different fault types.

Experimental validation demonstrated consistent performance improvements: multi-source fusion typically achieved 1–3% accuracy improvements compared to best single-method approach. More importantly, cross-condition generalization improved substantially: models trained under specific operating conditions maintained performance better when transferring to different conditions, suggesting that multi-source fusion captures more fundamental fault signatures less dependent on operating condition specifics.

The theoretical justification rests on signal processing fundamentals: each time-frequency method's underlying mathematics emphasizes different signal characteristics. CWT's scale-dependent resolution emphasizes transient localization; STFT's fixed resolution preserves frequency identity; HHT's adaptive decomposition separates nonlinear components; Wigner-Ville's quadratic formulation maximizes time-frequency concentration. Complementary emphasis on different characteristics naturally produces richer fault feature representations when combined.

2.5 Time-Frequency Representations as Image-Based Features

A fundamental paradigm shift emerged approximately 2017–2018, driven by deep learning's ascendancy [39]. Rather than extracting hand-crafted numerical features (spectral moments, kurtosis, entropy) from time-frequency scalograms through manual engineering, researchers began converting time-frequency representations directly to image format and feeding them to convolutional neural networks for automated feature learning.

2.5.1 Direct Scalogram-to-CNN Processing

Verstraete et al. (2017) [6] pioneered this approach, demonstrating that time-frequency scalograms (2D arrays of time-frequency coefficients) serve effectively as direct CNN inputs. The CNN architecture automatically learns fault-discriminative features through its convolutional layers, eliminating manual feature selection. This "featureless" or representation-learning paradigm offers multiple advantages. The first being the automatic feature discovery rather than designing features based on domain knowledge, the CNN identifies fault-discriminative patterns within time-frequency images adaptively. Another advantage is the generalization ability that the model has to accommodate to new fault types without modification. Finaly, the learned features adapt to noise variations, speed changes, load variations, better than hand-crafted features, likely because CNN learns complementary patterns representing the same fault under varying conditions.

Contemporary implementations typically standardize scalogram size, normalize magnitude values to 0–255 range for grayscale images, then feed to standard CNN architectures (ResNet, VGG, DenseNet, or custom architec-

tures). The time-frequency domain expertise shifts from feature engineering to hyperparameter selection: preprocessing parameters (sampling frequency, signal length, frequency band), time-frequency method selection, scalogram normalization strategy, and CNN architecture design.

2.5.2 Time-Frequency Image Enhancement

While directly feeding scalograms to CNNs proves effective, recent work demonstrates that preprocessing time-frequency images before CNN input further improves feature extraction. Ding et al. (2019) [40] introduced time-frequency manifold image synthesis (TFMIS), enhancing transient fault signatures through a discrete wavelet decomposition, manifold learning capturing fault feature variations, reconstructing enhanced time-frequency images emphasizing fault regions, finally processing at multiple wavelet decomposition levels.

This enhancement process reveals fault transients more clearly in the resulting images. Experimental validation reported 2–5% classification accuracy improvements from enhanced compared to unenhanced scalograms, suggesting that image preprocessing provides value beyond raw scaling information.

Similarly, Zaman et al. (2023) [20] applied Sobel edge detection to S-transform scalograms, highlighting fault feature boundaries. Edge enhancement revealed fault signatures more distinctly, improving subsequent CNN learning. This demonstrated valuable synergies between signal processing and computer vision domains: while time-frequency methods generate specialized signal representations, general image processing techniques further enhance fault feature visibility.

2.5.3 Texture-Based Feature Extraction from Time-Frequency Images

An alternative to deep learning approaches, some recent work combines time-frequency images with classical texture descriptors for feature extraction. Zhang et al. (2020) [41] applied Local Binary Patterns (LBP)—a texture descriptor quantifying local pixel intensity patterns—to time-frequency images. LBP captures local texture patterns characteristic of different fault types, generating compact feature vectors (typically 59 dimensions) suitable for traditional classifiers (SVM, random forest) without deep learning infrastructure.

This hybrid approach achieved 98.33% classification accuracy on standard datasets, demonstrating that texture analysis of time-frequency images provides effective fault signatures. The approach offers practical advantages: computational requirements remain low, interpretability is higher (texture patterns have intuitive visual meaning), and implementation requires minimal deep learning expertise. This makes the approach accessible to practitioners without specialized deep learning backgrounds, though raw accuracy trails state-of-the-art deep learning methods by 1–2%.

2.6 Advanced Feature Engineering Strategies

2.6.1 Physics-Informed Neural Networks with Embedded Time-Frequency Transforms

Recent work recognizes that embedding time-frequency transforms directly within neural network architectures enhances both interpretability and performance [8]. Time-Frequency Networks (TFN) integrate physically meaningful time-frequency transforms as explicit network layers. Rather than applying time-frequency analysis as preprocessing disconnected from learned features, the transform becomes an integral learned component.

This architecture design yields multiple benefits: (1) intermediate layer outputs provide time-frequency representations that can be visualized and interpreted; (2) the network is constrained to respect signal processing principles rather than learning entirely from data; (3) reduced sample complexity (fewer parameters require training) compared to fully learned approaches; (4) improved cross-condition generalization as physical principles transfer across operating conditions.

2.6.2 Attention Mechanisms for Multi-Source Feature Weighting

Recent state-of-the-art approaches employ channel-attention and spatial-attention mechanisms learning optimal feature combinations from multiple time-frequency sources. Feng et al. (2025) [7] integrated channel-attention modules that dynamically weight contributions from different time-frequency methods (CWT, STFT, HHT,

Wigner-Ville). These learned attention weights reveal which time-frequency methods prove most important for specific fault types.

Attention mechanisms address the challenge of multi-source fusion: naive concatenation of features from multiple methods may include irrelevant or noisy information from suboptimal transforms. Learned attention selectively emphasizes informative sources while de-emphasizing less relevant ones, improving overall representation quality.

2.6.3 Self-Supervised and Semi-Supervised Learning from Time-Frequency Representations

A critical practical challenge in machinery diagnostics is limited labeled data. While benchmark datasets (like Case Western Reserve University bearing data) provide thousands of labeled samples, real-world industrial systems typically generate sparse fault examples before bearing failure. Recent advances leverage time-frequency representations as basis for self-supervised learning, requiring minimal labeled data.

Liu et al. (2023) [9] developed self-supervised learning frameworks employing time-frequency images as primary features. The method learns representations by maximizing similarity between different augmentations of the same time-frequency image while minimizing similarity across different fault types. This approach requires minimal or no labeled data for representation learning, with labeling needed only for final classification layer training.

Similarly, Wang et al. (2025) [14] demonstrated semi-supervised fault diagnosis achieving ¿99% accuracy with minimal labeled samples (50 per fault type). The method enhanced standard supervised learning with pseudo-labeling and consistency regularization, exploiting abundant unlabeled vibration data while using sparse labeled data for classification.

3 Conclusion and Research Directions

Time-frequency analysis has established itself as foundational methodology for vibration signal feature engineering in rotating machinery diagnostics. The evolution from 2014 to 2025 demonstrates clear progression: from exploratory tool to manual feature extraction source, to image-based inputs for automated learning, to physics-informed and attention-based advanced architectures.

Contemporary best practice reflects these developments: adaptive signal decomposition preprocessing separates multi-component signals; multi-source time-frequency fusion captures complementary fault signatures; conversion to image format enables CNN-based feature learning; attention mechanisms learn optimal feature combinations; physics-informed architectures enhance interpretability; self-supervised frameworks minimize labeling requirements.

Significant research opportunities remain. Cross-condition generalization, particularly across different bearing types and machine geometries, remains challenging despite recent improvements. Computational efficiency for real-time embedded systems demands further optimization, though STFT-based approaches show promise. Physical interpretability of learned features, while improved through attention mechanisms and physics-informed networks remains incomplete. Real-world deployment in industrial systems faces practical challenges (sensor installation constraints, contaminated signals, incomplete datasets) requiring continued development.

Future research directions include: (1) domain adaptation techniques for transfer learning across different machines and conditions; (2) uncertainty quantification providing confidence estimates for diagnostic decisions; (3) explainability methods revealing which time-frequency regions drive fault classification; (4) hybrid systems combining classical signal processing principles with machine learning for enhanced interpretability; (5) online adaptation strategies responding to changing operating conditions.

The field demonstrates clear consensus that time-frequency representation constitutes essential feature engineering foundation for modern machinery diagnostics. Continued refinement of time-frequency methods, particularly in multi-source fusion and adaptive representation optimization, promises enhanced diagnostic accuracy and practical deployment capability.

References

[1]I. Bagri et al., "Vibration Signal Analysis for Intelligent Rotating Machinery Diagnosis and Prognosis: A Comprehensive Systematic Literature Review", Vibration **7**, 1013–1062 (2024).

[2]S. Li et al., "A systematic review of diagnosis methods for rolling bearing compound faults: research status, challenges, and future prospects", Meas. Sci. Technol. **36**, 012008 (2024).

[3]V. Kannan et al., "A Review of the Intelligent Condition Monitoring of Rolling Element Bearings", Machines **12**, 484 (2024).

[4]S. Neild et al., "A review of time-frequency methods for structural vibration analysis", Engineering Structures **25**, 713–728 (2003).

[5]Z. Feng et al., "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples", Mechanical Systems and Signal Processing **38**, 165–205 (2013).

[6]D. Verstraete et al., "Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings", Shock and Vibration **2017**, 5067651 (2017).

[7]T. Feng et al., "Rolling Based on Multi-Source Time–Frequency Feature Fusion with a Wavelet-Convolution, Channel-Attention-Residual Network-Bearing Fault Diagnosis Method", Sensors **25**, 4091 (2025).

[8]D. Chen et al., "A physics-informed neural network approach to fatigue life prediction using small quantity of samples", International Journal of Fatigue **166**, 107270 (2023).

[9]Y. Liu et al., "Self-supervised feature extraction via time–frequency contrast for intelligent fault diagnosis of rotating machinery", Measurement **210**, 112551 (2023).

[10]P. Flandrin, *Time-Frequency/Time-Scale Analysis* (Academic Press, Sept. 1998).

[11]N. Hess-Nielsen and M. Wickerhauser, "Wavelets and time-frequency analysis", Proceedings of the IEEE **84**, 523–540 (1996).

[12]I. Daubechies, *Ten lectures on wavelets* (SIAM, 1992).

[13]I. Makrouf et al., "Time-Frequency Analysis and Deep CNN-Based Fault Diagnosis for Rotating Machinery Under Variable Operating Conditions", in 2025 International Conference on Control, Automation and Diagnosis (ICCAD) (July 2025), pp. 1–6.

[14]J. Wang et al., "A Semi-supervised Fault Diagnosis Method for Rolling Bearing Using Time-Frequency Transform Enhanced Contrastive Learning", Eksploatacja i Niezawodność – Maintenance and Reliability **28**, `10.17531/ein/207015` (2025).

[15]D. Gabor, "Theory of communication. Part 1: The analysis of information", Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering **93**, 429–441 (1946).

[16]F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations", IEEE signal processing magazine **9**, 21–67 (2002).

[17]A. Mohammad-Alikhani et al., "Fault Diagnosis of Electric Motors by a Channel-Wise Regulated CNN and Differential of STFT", IEEE Transactions on Industry Applications **61**, 3066–3077 (2025).

[18]M. Zhang et al., "Physics-informed deep learning for structural vibration identification and its application on a benchmark structure", Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **382**, 20220400 (2023).

[19]R. Stockwell et al., "Localization of the complex spectrum: the S transform", IEEE Transactions on Signal Processing **44**, 998–1001 (1996).

[20]W. Zaman et al., "Centrifugal Pump Fault Diagnosis Based on a Novel SobelEdge Scalogram and CNN", Sensors **23**, 5255 (2023).

[21]G. Li et al., "Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and s-transform", Sensors **19**, 2750 (2019).

[22]N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", Proc. A **454**, 903–995 (1998).

[23]I. O. Olalere and O. A. Olanrewaju, "Tool and Workpiece Condition Classification Using Empirical Mode Decomposition (EMD) with Hilbert–Huang Transform (HHT) of Vibration Signals and Machine Learning Models", Applied Sciences **13**, 2248 (2023).

[24]Z. Wu et al., "The multi-dimensional ensemble empirical mode decomposition method", Advances in Adaptive Data Analysis **1**, 339–372 (2009).

[25]M. E. Torres et al., "A complete ensemble empirical mode decomposition with adaptive noise", in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (May 2011), pp. 4144–4147.

[26]J. VILLE, "Theorie et application dela notion de signal analysis", Cables et Transmissions **2**, 61–74 (1948).

[27]B. Boashash and P. Black, "An efficient real-time implementation of the Wigner-Ville distribution", IEEE Transactions on Acoustics, Speech, and Signal Processing **35**, 1611–1618 (1987).

[28]L. Cohen, "Time-frequency distributions-a review", Proceedings of the IEEE **77**, 941–981 (1989).

[29]O. Rioul and P. Flandrin, "Time-scale energy distributions: a general class extending wavelet transforms", IEEE Transactions on signal processing **40**, 1746–1757 (1992).

[30]L. Lin and J. Hongbing, "Signal feature extraction based on an improved EMD method", Measurement **42**, 796–803 (2009).

[31]K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition", IEEE Transactions on Signal Processing **62**, 531–544 (2014).

[32]Z. Li et al., "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive", Mechanical Systems and Signal Processing **85**, 512–529 (2017).

[33]C. Liu et al., "Planetary gears feature extraction and fault diagnosis method based on vmd and cnn", Sensors **18**, 1523 (2018).

[34]G. An et al., "A Parameter-Optimized Variational Mode Decomposition Investigation for Fault Feature Extraction of Rolling Element Bearings", Mathematical Problems in Engineering **2021**, 6629474 (2021).

[35]L. Wang and Y. Shao, "Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis", Mechanical Systems and Signal Processing **138**, 106545 (2020).

[36]C. Hu et al., "Fault diagnosis of rolling bearings based on variational mode decomposition and genetic algorithm-optimized wavelet threshold denoising", Machines **10**, 649 (2022).

[37]C. Gong and R. Peng, "A novel hierarchical vision transformer and wavelet time–frequency based on multi-source information fusion for intelligent fault diagnosis", Sensors **24**, 1799 (2024).

[38]Y. Yu et al., "MSIFT: A novel end-to-end mechanical fault diagnosis framework under limited & imbalanced data using multi-source information fusion", Expert Systems with Applications **274**, 126947 (2025).

[39]Y. LeCun et al., "Deep learning", Nature **521**, 436–444 (2015).

[40]X. Ding et al., "Transient Feature Extraction Based on Time–Frequency Manifold Image Synthesis for Machinery Fault Diagnosis", IEEE Transactions on Instrumentation and Measurement **68**, 4242–4252 (2019).

[41]S. Zhang et al., "Fault diagnosis of rotating machinery based on time-frequency image feature extraction", Journal of Intelligent & Fuzzy Systems **39**, 5193–5200 (2020).