

OPEN ACCESS

EDITED BY

Jesus Enrique Sierra Garcia,
University of Burgos, Spain

REVIEWED BY

Cheng Liu,
Shanghai Jiao Tong University, China
Petr Doležel,
University of Pardubice, Czechia

*CORRESPONDENCE

Fabing Liu,
✉ 1934760845@qq.com

RECEIVED 26 May 2025

REVISED 28 November 2025

ACCEPTED 09 January 2026

PUBLISHED 04 February 2026

CITATION

Zheng L, Liu F, Zuo S, Zhu X and Huang G
(2026) Identification of unknown crack
defects in wind turbine main shafts based on
acoustic signature and multi-scale
convolutional neural networks.
Front. Energy Res. 14:1635112.
doi: 10.3389/fenrg.2026.1635112

COPYRIGHT

© 2026 Zheng, Liu, Zuo, Zhu and Huang. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Identification of unknown crack defects in wind turbine main shafts based on acoustic signature and multi-scale convolutional neural networks

Liuyu Zheng¹, Fabing Liu^{1*}, Shihai Zuo¹, Xuefeng Zhu² and Guoyong Huang²

¹CGN New Energy Investment (Shenzhen) Co., Ltd., Yunnan Branch, Kunming, China, ²Faculty of Civil Aviation and Aeronautics, Kunming University of Science and Technology, Kunming, China

Introduction: Wind turbine main shaft crack detection is crucial for operational safety and maintenance planning. Conventional feature based diagnosis generalizes poorly to complex or unseen cracks, and deep learning is constrained by scarce and imbalanced defect data. This study proposes an acoustic signature driven multi-scale CNN (MSCNN) framework for identifying unknown main shaft crack defects.

Methods: A double threshold energy to zero-crossing (EZR) segmentation method is introduced to construct acoustic feature maps that capture both transient and steady-state crack characteristics, enhancing detection sensitivity and specificity. The MSCNN architecture automatically extracts multi-scale temporal features without manual feature engineering, while a novel segmentation strategy decomposes complex or unknown cracks into identifiable components for quantitative assessment.

Results: The proposed EZR-driven MSCNN framework achieves an average recognition accuracy of 90%, representing a 6.73% improvement over extreme learning machine (ELM) and a 3.36% improvement over single scale CNNs. Cross platform testing confirms robust adaptability, with accuracy ranging from 83.9% to 87.2% across different turbine models. Visualization analysis demonstrates improved separability of crack related acoustic features compared to conventional single-scale or handcrafted feature baselines.

Discussion: This work provides a practical and effective solution for wind turbine crack detection with enhanced capability for detecting diverse and previously unseen crack types in data scarce scenarios. The proposed framework demonstrates superior recognition stability and supports practical condition monitoring and early warning systems for wind turbine maintenance.

KEYWORDS

acoustic signature, crack detection, multi-scale convolutional neural network (MSCNN), unknown defect identification, wind turbine main shaft

1 Introduction

The global demand for safe, economical, and renewable energy has driven the rapid growth of wind energy (Hassan et al., 2024). By the end of 2025, global installed wind power capacity exceeded 700 GW (Kumar Dora et al., 2025). To meet rising energy demands, modern turbines are designed for higher power output, which subsequently requires higher reliability. Ensuring operational safety while controlling maintenance costs has become a critical focus (Gbashi et al., 2024). A robust condition monitoring and fault diagnosis system is essential to safeguard turbine reliability. As the core mechanical component, the main shaft endures complex forces, making it highly susceptible to fatigue cracks. A sudden fracture can cause severe equipment damage and economic losses (Santelo et al., 2022). Reliable crack detection is thus essential for safe operation, but it remains challenging. Cracks propagate unpredictably, interact with other defects, and often present as diverse or previously unclassified types, complicating the diagnostic process (Nejad et al., 2022).

Conventional crack detection methods typically involve signal acquisition, data analysis, and state classification. Ultrasonic sensors, for example, capture acoustic signals whose features are altered by defects (Cheng et al., 2020; Wang and Chen, 2023). Data analysis is critical for isolating these features under noise. Classical feature extraction methods include Wavelet Transform and Empirical Mode Decomposition (EMD) (Ding et al., 2019). More advanced approaches have focused on enhancing signals in noisy environments (Xia et al., 2020; Guo et al., 2019) or developing hybrid frameworks for complex patterns, such as extended cepstrum analysis or joint amplitude-frequency demodulation (Teng et al., 2019; Feng et al., 2019; Wang et al., 2019). Despite these advances, traditional methods often fail when confronting complex or unseen crack types. Their reliance on manually engineered features and predefined fault models limits their adaptability to the diverse defects found in real-world operations.

Deep learning (DL) has shown strong hierarchical feature learning capabilities, achieving success in complex recognition tasks (Hinton et al., 2006). Recent studies demonstrate that multi-scale convolutional neural networks (MSCNNs) are particularly effective at learning discriminative fault patterns from sensor data. Research shows MSCNNs can effectively capture features at different scales, achieving high accuracy and robustness against noise and varying loads (Peng et al., 2025; Chen et al., 2021; Zhao et al., 2025). However, applying DL to main shaft crack detection poses significant challenges. The operating environment of wind turbines yields datasets with limited defect diversity. Deep learning models, known for their high parameter counts, require large, balanced, and well-labeled datasets (Zhou and Wu, 2022). In practice: (1) Crack defects vary widely, making it impractical to collect all defect types; (2) Data acquisition and labeling are costly; and (3) Defect samples are far fewer than normal samples, leading to pronounced class imbalance. These factors hinder the development of generalizable models, especially for recognizing unknown crack types in data-scarce scenarios.

To address these challenges, this study proposes an acoustic-signature-driven multi-scale convolutional neural network (MSCNN) for identifying unknown crack defects in wind turbine main shafts. The key innovations are: (1) Acoustic

Signature Construction: A double-threshold energy to zero-crossing segmentation method is introduced to construct acoustic feature maps that capture both transient and steady state characteristics of cracks, enhancing sensitivity and specificity. (2) MSCNN-Based Feature Learning: A multi-scale CNN architecture automatically extracts features from acoustic maps at different temporal resolutions, capturing both fine-grained details and global patterns without manual feature engineering. (3) Recognition of Unknown Cracks: A segmentation strategy decomposes complex or unknown cracks into smaller components, enabling quantitative assessment and improving recognition accuracy for previously unseen defect types.

The remainder of this paper is organized as follows: Section 2 presents the proposed methodology, Section 3 describes the experimental setup, Section 4 reports and discusses the results, and Section 5 concludes with a summary and future research directions.

2 Methods

2.1 Individual acoustic signature extraction via energy to Zero Ratio (EZR)

The acoustic signature of a wind turbine main shaft is characterized using the EZR (Zhou and Wu, 2022; Wang et al., 2021), which effectively combines frame level energy and zero crossing information to highlight defect related patterns. A dual threshold segmentation algorithm is applied to localize high EZR regions, which are then normalized to form sample matrices for subsequent input into the MSCNN. The complete procedure is as follows.

Let $x(n)$ denote the acquired acoustic waveform. After windowing and framing with frame length N , the i th frame signal $x_i(n)$ is obtained. The frame energy is calculated as in Equation 1:

$$E_i = \sum_{n=1}^N x_i^2(n) \quad (1)$$

To mitigate the misinterpretation of transient noise as defect onset or offset, an enhanced energy measure introduces a robustness constant (Equation 2):

$$LE_i = \log_{10} \left(1 + \frac{E_i}{a} \right) \quad (2)$$

The robustness constant $\alpha = 0.003$ (range [0.001, 0.005]) was determined through pilot experiments on 100 training samples, providing an optimal balance between noise suppression and feature preservation. Values below 0.001 exhibited excessive sensitivity to noise, while values above 0.005 tended to over-smooth critical transient features.

To stabilize the calculation of the zero-crossing rate (ZCR) and eliminate minor zero drift artifacts, a central clipping operation is applied to the framed signal as in Equation 3, $x_i(n)$:

$$\tilde{x}_i(n) = \begin{cases} x_i(n), & |x_i(n)| \geq \delta \\ 0, & |x_i(n)| < \delta \end{cases} \quad (3)$$

The center-cropping threshold δ is a dimensionless scaling parameter. The value $\delta = 0.7$ was selected from the range [0.5,

1.0] based on maximizing drift artifact removal effectiveness (91% effectiveness) while minimizing signal distortion (<3.5% relative error).

Following central clipping, the ZCR of each frame is computed as in Equation 4, where the sign function is defined in Equation 5:

$$ZCR_i = \sum_{n=1}^N |\text{sign}[\tilde{x}_i(n)] - \text{sign}[\tilde{x}_i(n-1)]| \quad (4)$$

$$\text{sign}[\tilde{x}_i(n)] = \begin{cases} 1, & |\tilde{x}_i(n)| \geq 0 \\ -1, & |\tilde{x}_i(n)| < 0 \end{cases} \quad (5)$$

The Energy to Zero Ratio (EZR) is then defined as in Equation 6:

$$EZR_i = \frac{LE_i}{ZCR_i + b} \quad (6)$$

The regularization parameter $b = 0.03$ (range [0.01, 0.05]) ensures numerical stability when the zero-crossing rate (ZCR) approaches zero (condition number < 500), avoiding the overflow errors observed with $b \leq 0.02$ while preventing the systematic bias (<0.5%) associated with $b \geq 0.04$.

The dual-threshold segmentation mechanism adopts a two-level decision process to reliably identify defect-related acoustic events. The high threshold T_2 is set to the 95th percentile of the EZR distribution across all training samples, yielding $T_2 = 0.035$. This percentile-based rule robustly triggers candidate defect waves while limiting false positives from stochastic noise spikes. The low threshold T_1 is fixed at 0.015 ($\approx 43\%$ of T_2), determined via receiver operating characteristic (ROC) analysis to balance detection sensitivity and false-alarm rate. Operationally, when an EZR excursion exceeds T_2 , the algorithm searches bidirectionally (backward and forward in time) for the nearest intersections with T_1 , which define the temporal boundaries of the acoustic signature. This hysteresis strategy captures complete defect signatures despite transient EZR fluctuations. Thresholds were validated by a grid search over $T_2 \in [0.025, 0.045]$ and $T_1/T_2 \in [0.3, 0.5]$ on the validation set; the selected pair ($T_2 = 0.035, T_1 = 0.015$) achieved the highest F1-score (0.92), confirming their effectiveness for wind-turbine main-shaft defect detection.

For individual acoustic signature extraction, the ultrasonic signal is processed frame-by-frame to compute energy and EZR, from which a smoothed EZR sequence is obtained. The dual-threshold scheme then delineates segment boundaries: when EZR first exceeds T_2 , a significant event is flagged; the start point is set by backtracking to where EZR falls below T_1 , and the end point is set when EZR again drops below T_1 . This procedure yields stable boundaries and shows strong robustness to noise. To meet the MSCNN input requirements, each extracted segment is resampled to a fixed duration and amplitude-normalized to ensure consistency in temporal length and signal magnitude. In this design, T_2 provides precise event triggering, whereas T_1 supplies hysteresis, suppressing noise while preserving the completeness of defect information.

This procedure, illustrated in Figure 1, exemplifies the extraction of individual acoustic features from a wind turbine main shaft using the EZR method. The left panel shows the raw ultrasonic echo signal, which includes the transmitted wave, intrinsic structural wave, and backwall reflection from the shaft. By applying the dual threshold EZR segmentation method, high EZR segments correlated with potential defects are isolated (as highlighted by the middle arrow).

The right panel displays the resulting acoustic signature, comprising both the intrinsic structural response and the moderate crack reflection, which serves as the standardized input for subsequent MSCNN based feature learning.

2.2 MSCNN feature learning

To address the limitations of conventional CNN based crack detection models such as restricted feature expressiveness from fixed size convolution kernels and insufficient multi-dimensional feature extraction this study proposes a MSCNN tailored to the characteristics of acoustic wave signals (Huang and Wang, 2019; Chen et al., 2024; Fu et al., 2020). The MSCNN employs parallel convolutional branches with kernels of varying sizes to extract features at different temporal scales. Small kernels capture fine grained, short term details, whereas large kernels extract broader, long term structural patterns, enabling the network to model the diverse morphology and topology of crack related signals. The framework diagram of the multi - scale convolutional neural network is shown in Figure 2.

A CNN is a deep, feed forward network characterized by convolutional operations, sparse connections, and weight sharing. A one dimensional architecture is adopted because: (i) the EZR feature already encodes time-frequency information; (ii) 1D inputs better preserve temporal structure of transient crack signatures; and (iii) computational efficiency is critical for real-time monitoring. In MSCNN, convolution kernels of different sizes operate in parallel, each followed by a pooling layer, allowing simultaneous extraction of multi scale temporal features. Formally, the convolution operation is expressed as in Equation 7:

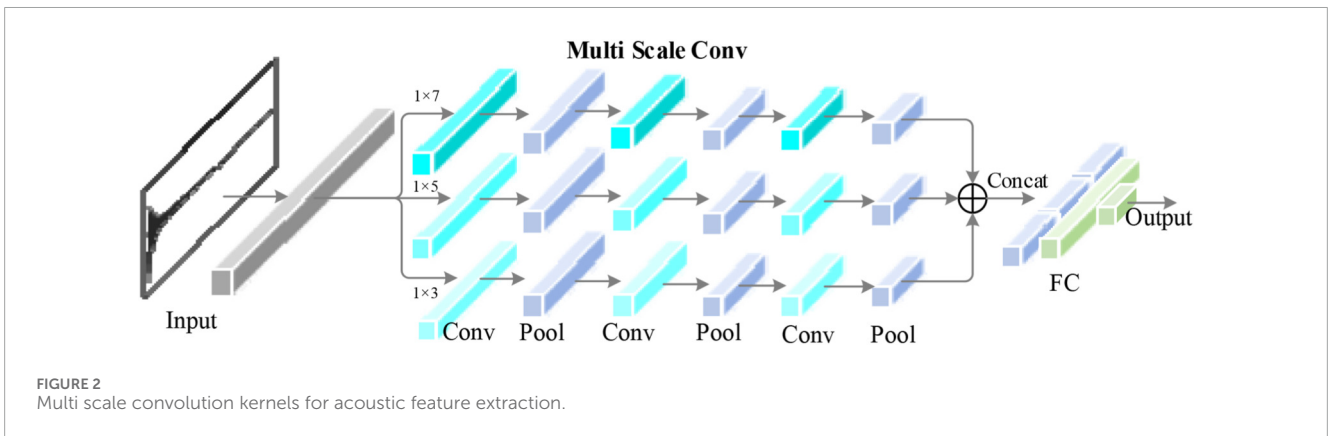
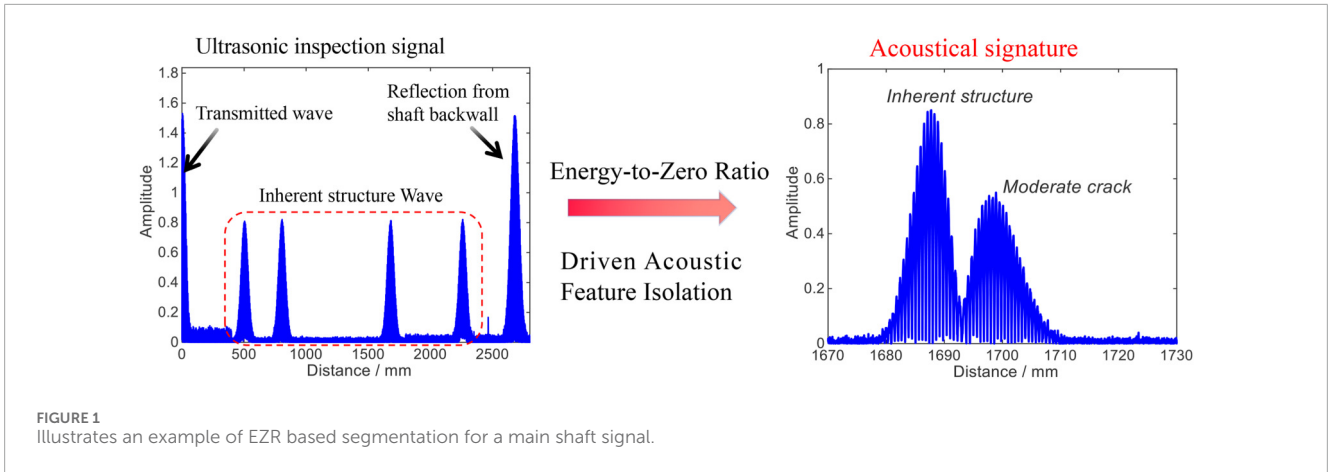
$$fea_{i,j}^d = f\left(b_i^{(d)} + \sum_{i=1}^d \langle W_i^{(d)}, x_{t+d-i}^{(d-1)} \rangle\right) \quad (7)$$

where $x_{t+d-i}^{(d-1)}$ is the input 1D sequence, $x_{t+d-i}^{(d-1)}$ is the weight tensor with F_l output channels and F_{l-1} input channels, $b^{(d)} \in R^{F_l}$ is the bias term, $d \in \{3, 5, 7\}$ is the kernel size for capturing multi-scale patterns, and different strides match the temporal scale of interest. The activation function $f(\cdot)$ is ReLU, chosen because: (i) its sparsity aligns with the clustered sparse energy distribution of acoustic signals; (ii) it suppresses low amplitude noise leakage compared to Leaky ReLU; and (iii) combined with batch normalization, it achieves faster convergence and smaller generalization gaps on our dataset without observable dying neuron effects.

Following convolution, the max-pooling operation is defined in Equation 8:

$$P_{(m,n)}^h = \max_{(n-1)g < t < ng} \{fea_{(m,t)}^h\}, n = 1, 2, \dots \quad (8)$$

where $fea_{(m,t)}^h$ denotes the activation of the h th neuron in the m th feature map of at layer t , g is the pooling kernel width, and $P_{(m,n)}^h$ is the output of the pooling operation. Max-pooling serves multiple purposes: (i) reducing temporal dimensions by half while retaining salient features, (ii) providing translational invariance to minor temporal shifts in crack signatures, and (iii) acting as an implicit regularizer by reducing feature map complexity. Batch normalization, applied before each activation,



stabilizes training by normalizing layer inputs, which is particularly beneficial in data-scarce scenarios where gradient estimates from small batches can be unstable. Together, these techniques reduce overfitting risk and accelerate convergence, as evidenced by our ablation studies. This multi-scale framework enhances temporal feature diversity, enabling the model to robustly identify both common and complex crack patterns in wind turbine main shafts.

Outputs from all branches are concatenated into a unified feature vector, followed by batch normalization (BN) to stabilize training and accelerate convergence. The final representation is expressed as in Equation 9:

$$Y = f(g(c(b_{short}, b_{medium}, b_{long}))) \tag{9}$$

where b_{short} , b_{medium} , and b_{long} are the short, medium, and long term features, respectively; \oplus denotes the feature concatenation operation.

2.3 Proposed framework for unknown crack identification

To address the limited feature representation capability of conventional fixed-kernel CNNs in capturing multi-scale acoustic

characteristics of crack defects, we propose an integrated framework that combines ultrasonic guided wave acquisition, EZR-based signature extraction, and MSCNN classification. The MSCNN employs parallel convolutional branches with varying kernel sizes to extract features across multiple temporal scales, enabling robust identification of diverse crack types in wind turbine main shafts. Notably, the framework can recognize previously unseen compound crack patterns through similarity based decomposition, eliminating the need for exhaustive training data covering all possible defect combinations. The proposed framework comprises four main stages, progressing from raw ultrasonic signal acquisition to classification of previously unseen crack patterns, as illustrated in Figure 3.

Step 1: Extraction of individual acoustic signatures based on the EZR

Raw acoustic signals are segmented through windowing and framing operations to produce temporal frames of specified length with partial overlap. For each frame, the energy and zero-crossing count are computed to derive the EZR feature, which effectively captures localized acoustic variations under different structural states. A dual-threshold segmentation algorithm is then applied to isolate salient acoustic signatures, followed by z-score normalization and length standardization. This

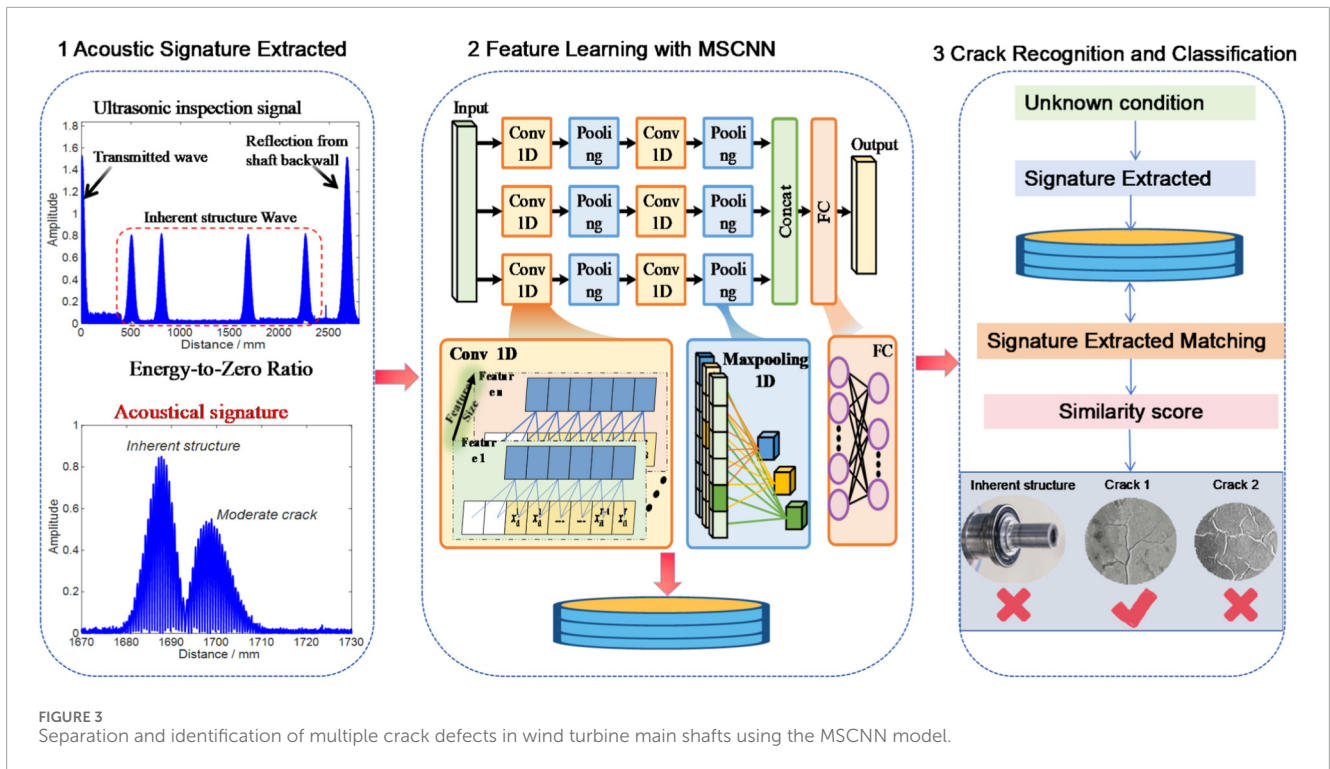


FIGURE 3 Separation and identification of multiple crack defects in wind turbine main shafts using the MSCNN model.

preprocessing stage enhances crack induced acoustic anomalies while suppressing ambient noise, producing standardized feature matrices suitable for subsequent neural network processing.

Step 2: Multi-scale feature construction using MSCNN

The normalized EZR matrices are fed into the MSCNN as one-dimensional time-series inputs, preserving temporal dependencies crucial for detecting transient defect signatures. The MSCNN employs three parallel convolutional branches with different kernel sizes to capture features across multiple temporal scales. Each branch consists of successive Conv1D–MaxPool1D blocks with progressively increasing channel dimensions. Features extracted from all branches are concatenated into a unified high-dimensional vector, which is further refined by fully connected layers to enhance discriminative capability for crack identification.

Step 3: Detection and classification of unknown conditions

For unknown shaft conditions, the EZR based acoustic signature is extracted following the Step-1 procedure. The signal is subsequently segmented to isolate individual acoustic components, because complex or compound cracks typically manifest as multiple high-amplitude regions in the EZR sequence, each corresponding to a distinct structural anomaly. Empirically, different crack configurations exhibit characteristic EZR patterns: a healthy shaft presents a single peak reflecting the inherent structural response; single-crack states exhibit two peaks; and multi-crack states present three or more peaks. Each segmented component is independently input to the trained MSCNN to obtain a deep feature vector, which is then matched against the reference

acoustic-signature library using cosine similarity as defined in Equation 10:

$$\text{sim}(\mathbf{v}_q, \mathbf{v}_r) = \frac{\mathbf{v}_q \cdot \mathbf{v}_r}{|\mathbf{v}_q| |\mathbf{v}_r|} \tag{10}$$

$\mathbf{v}_q \in \mathbb{R}^d$ is the query vector, the deep feature of a segmented acoustic component from an unknown sample, with $\mathbf{v}_q = f(x)$, $f(\cdot)$ denote the trained MSCNN embedding function. $\mathbf{v}_r \in \mathbb{R}^d$ is the reference vector, the prototype vector of class in the reference library, computed from training embeddings.

The reference library comprises prototype feature vectors for the elementary crack categories: inherent structure, small crack, medium crack, and significant defect. If the maximum similarity falls below a predefined threshold (as specified in the experimental section), the component is classified as healthy; otherwise, it is assigned to the crack category with the highest similarity. For compound cracks, all constituent components are identified and their similarity scores are reported, yielding a comprehensive diagnosis.

Step 4: Classification output and validation

This decomposition based strategy enables the framework to identify previously unseen compound crack patterns by representing them as combinations of known elementary crack types. Specifically, the method can recognize novel multi-crack states using models trained only on basic single crack categories, thereby significantly improving data efficiency and scalability. The validation design for both known and unknown crack states is detailed in Tables 3, 4, demonstrating the framework’s strong generalization capability to unseen defect patterns.

TABLE 1 Physical characteristics of the wind turbine main shaft.

Parameter category	Specific parameter	Value/Description
Geometric dimensions	Total length	2,690 mm
	Maximum diameter	560 mm
	Minimum diameter	420 mm
	Central bore diameter	140 mm
	Cross sectional features	Flange arcs, chamfers, fillets, varying cross-sections
Material properties	Material type	42CrMo alloy steel
	Chemical composition	C: 0.38%–0.45%, Cr: 0.90%–1.20%, Mo: 0.15%–0.25%, Mn: 0.40%–0.70%, si: 0.17%–0.37%
	Ultrasonic wave propagation velocity	5,930 m/s

3 Experimental setup

3.1 Physical characteristics of wind turbine main shaft

The experimental investigation was carried out on a megawatt class wind turbine main shaft located at a wind farm in Yunnan Province, China. The physical characteristics of the tested main shaft are summarized in Table 1. The shaft is installed within a semi enclosed compartment characterized by limited space and densely arranged equipment. Due to structural constraints, inspection devices can only access the exposed end face near the blade side, while the remaining sections are enclosed within a protective casing. To enable testing without dismantling the shaft, the exposed end face was selected as the inspection location, and the pulse echo method was adopted for ultrasonic signal acquisition.

Considering the inspection depth, an excitation frequency of 4 MHz and a 34 mm diameter straight probe were selected to achieve an optimal balance between penetration depth and defect resolution, in line with industry standards for wind turbine component inspection. Figure 4 illustrates the experimental setup and ultrasonic signal acquisition process. The data acquisition system comprised high precision ultrasonic sensors, signal conditioners, and a data acquisition card with a 100 MHz sampling frequency to ensure capture of fine signal details.

3.2 Artificial crack creation and validation

To simulate realistic and variable crack conditions, artificial cracks of varying sizes (minor, moderate, and major) were introduced across the shaft, with their strategic distribution designed to replicate overlapping and heterogeneous defect conditions.

Three fabrication techniques were employed. This approach enabled the acquisition of acoustic signature signals representing both isolated and composite crack states. Artificial cracks with

precisely controlled dimensions were fabricated using three primary methods. Electric discharge machining was employed for small and medium sized cracks, utilizing a 0.2 mm wire to produce defects with high dimensional accuracy and well defined boundaries, while minimizing alterations to the surrounding material properties. Ultrasonic Impact Treatment was applied at selected locations to induce microcracks resembling fatigue induced natural cracks through high frequency impact loading. Mechanical Notching, performed with specialized 0.5 mm tungsten carbide tools, was used to create large cracks with controlled depth and geometry.

Based on statistical data from actual failure cases in the wind energy industry and the structural characteristics of the main shaft, three representative crack sizes were designed: (1) Minor cracks: length 5–10 mm, width 0.2–0.3 mm, depth 2–3 mm; (2) Moderate cracks: length 15–25 mm, width 0.3–0.4 mm, depth 4–6 mm; (3) Major defects: length 30–50 mm, width 0.5–0.7 mm, depth 8–12 mm. To ensure the representativeness of the artificial cracks, their ultrasonic reflection characteristics were compared with those of naturally occurring cracks in defective main shafts retrieved from in service wind turbines. A high degree of similarity (>90%) in waveform features, reflection amplitude, and spectral distribution confirmed that the fabricated cracks effectively replicated natural defect conditions.

3.3 Data acquisition and sample design

A total of eight operating states of the wind turbine main shaft were recorded, covering the healthy condition, single crack states, and multiple crack states. For each state, 400 acoustic signature samples were collected. The acoustic signature data were labeled according to structural components and crack sizes. Table 2 summarizes the operating states, their identifiers, detailed descriptions, and the corresponding extracted acoustic signatures.

As shown in Table 2, the inherent structure component is present in all eight states, yielding 3,200 samples in total (400 per state). Minor cracks appear in C2, C5, C6, and C8, providing 1,600 samples.

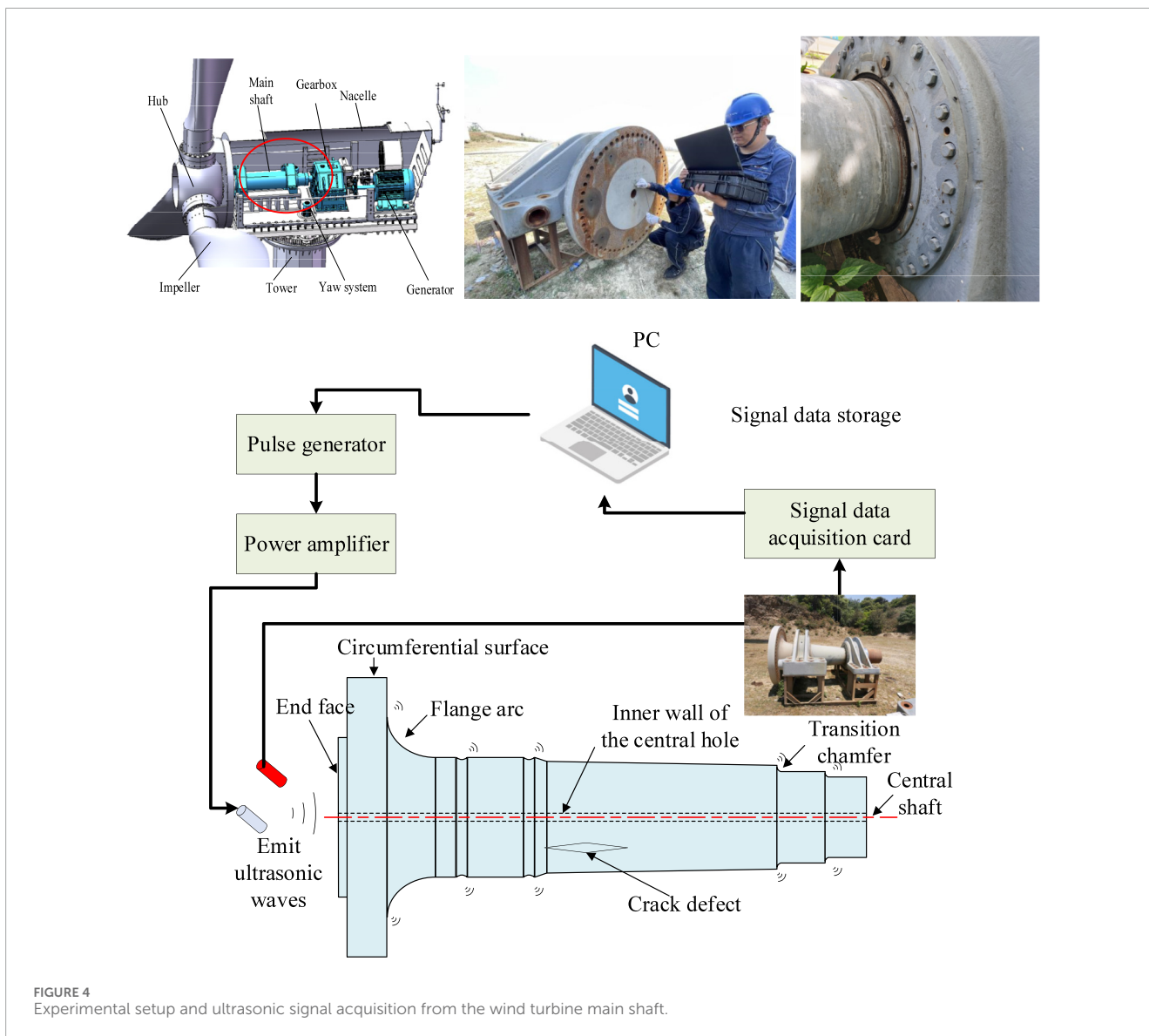


TABLE 2 Operational states and acoustic signatures of wind turbine main shaft.

Operating state	State ID	Detailed description	Extracted acoustic signature
Health	C1	Normal operation	Inherent structure
Single crack	C2	Inherent structure + minor crack	Inherent structure; minor crack
	C3	Inherent structure + moderate crack	Inherent structure; moderate crack
	C4	Inherent structure + major defect	Inherent structure; major defect
Multiple cracks	C5	Inherent structure + minor crack + moderate crack	Inherent structure; minor crack; moderate crack
	C6	Inherent structure + minor crack + major defect	Inherent structure; minor crack; major defect
	C7	Inherent structure + moderate crack + major defect	Inherent structure; moderate crack; major defect
	C8	Inherent structure + minor crack + moderate crack + major defect	Inherent structure; minor crack; moderate crack; major defect

TABLE 3 Summary of conditions and extracted signatures for multi-defect effect experiments.

Dataset	Operating states	Number of samples
Dataset A	C1; C2; C3; C4	Inherent structure: 400
		Minor crack: 400
		Moderate crack: 400
		Major defect: 400
Dataset B	C5; C6; C7; C8	Inherent structure: 400
		Minor crack: 400
		Moderate crack: 400
		Major defect: 400

Moderate cracks are present in C3, C5, C7, and C8, also totaling 1,600 samples. Major defects occur in C4, C6, C7, and C8, likewise totaling 1,600 samples.

To evaluate the model's capability in identifying unknown crack types, the eight operating states were divided into two distinct datasets as shown in Table 3.

As shown in Table 3, Dataset A (C1–C4) contains healthy and single-crack states and serves as the known-state dataset. Dataset B (C5–C8) contains multiple-crack states and serves as the unknown-state dataset.

Two validation experiments were designed: (1) Recognition of known crack defect states to verify model accuracy for known defects. (2) Recognition of unknown crack defect states—to evaluate generalization to unseen multiple-crack states.

As shown in Table 4, in Experiment 1, 70% of Dataset A was used for training and 30% for testing, resulting in 1,120 and 480 samples, respectively. In Experiment 2, all samples from Dataset A were used for training, while Dataset B served as the test set, each containing 1,600 samples. This rigorous experimental design ensured that the model was never exposed to composite crack data during training, thereby enabling an objective evaluation of its capability to learn from fundamental crack features and generalize to more complex scenarios. Such a design is crucial for validating the practical applicability of the proposed method, as it directly addresses the challenges encountered in real-world engineering applications.

3.4 MSCNN network architecture parameters

Table 5 presents the detailed network architecture parameters of the MSCNN, including the kernel sizes, strides, padding, channel configurations for each branch, and the output dimensions at each stage.

Table 5 presents the detailed parameters of the MSCNN architecture. The network employs a three branch design with kernel sizes of 7, 5, and 3 to extract multi-scale temporal features

from the 1×1024 EZR acoustic signatures. Each branch contains three consecutive Conv1D–MaxPool1D blocks with progressively increasing channels (32, 64, and 128). The outputs from the three branches are concatenated, followed by a global average pooling layer and a two-layer fully connected (FC) classifier for four-class defect recognition.

4 Experimental results

4.1 Individual acoustic signature extraction via EZR

The collected acoustic signals are first subjected to noise reduction processing, followed by adaptive extraction of acoustic signature features.

As shown in Figure 5, the horizontal axis represents the ultrasonic propagation distance along the wind turbine main shaft. In Figure 5a, the raw ultrasonic signal exhibits substantial random fluctuations and high-frequency noise, which obscure meaningful features and make the waveform appear chaotic. In contrast, Figure 5b presents the denoised signal, where the waveform is noticeably smoother, the main peaks are more distinct, and high-frequency noise is effectively suppressed. This enhancement improves both feature clarity and signal interpretability. Notably, pronounced peaks and discontinuities particularly near 1500 mm and 2000 mm are clearly visible after denoising, indicating potential defect-related features. Overall, the denoising process effectively removes high-frequency interference while preserving critical structural information, thereby producing a cleaner, more analytically useful signal suitable for feature extraction and defect localization.

To identify the onset and endpoint of high-threshold intervals corresponding to potential defects, a double-threshold segmentation method is applied to the EZR processed acoustic signal. Each individual acoustic signature is defined as a fixed-length segment starting from the onset point determined by the double-threshold method, enabling targeted feature extraction. Following segmentation, each signature is normalized, and a sample matrix is constructed for input into the MSCNN.

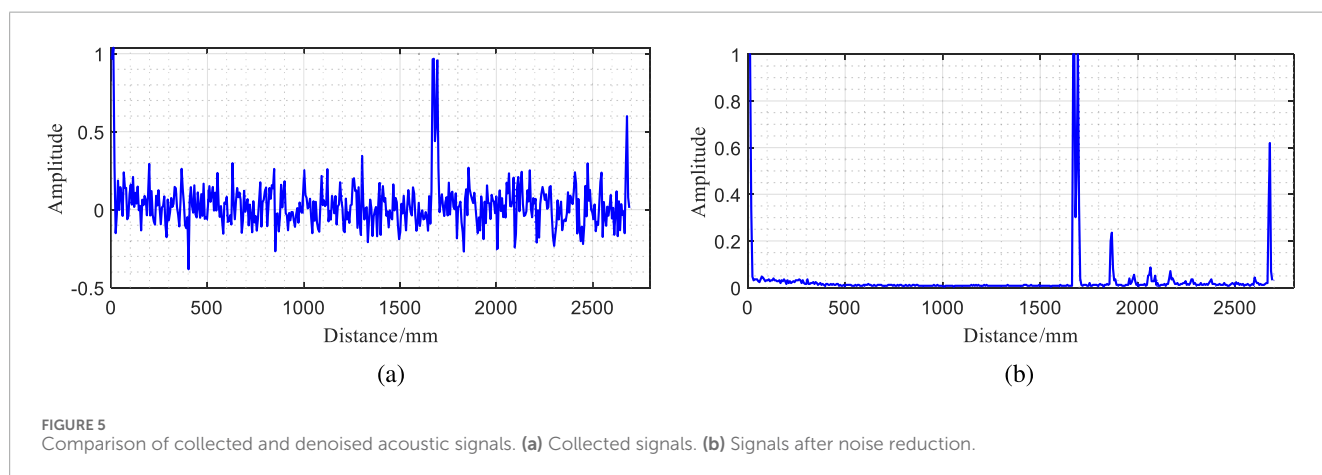
As illustrated in Figure 6, the proposed algorithm incorporates two level decision mechanism based on EZR. In the first stage, a high threshold T_2 is applied to the short time EZR curve to preliminarily identify candidate structural or defect waves, the segment between points A and B. In the second stage, a lower threshold T_1 is introduced. From point A, the algorithm searches leftward, and from point B, rightward, to locate points C and D where the signal intersects T_1 . These points define the start and end of the acoustic signature. To ensure uniform segment lengths, the waveform beyond T_1 is extended to a fixed duration. The segmented acoustic signal is thus divided into fixed length individual signatures, with their start and end positions marked by solid and dashed black lines in the figure. Experimental results demonstrate that this method effectively distinguishes between defect-free and defective signals, providing a robust dataset for subsequent defect identification and classification. The resulting segmented acoustic signatures are shown in Figure 7.

TABLE 4 Training and testing sample sets for validation experiments.

Validation purpose	Training set	Testing set	Samples (training: testing)
Known crack defect recognition	Dataset A	Dataset A	1120: 480
Unknown crack defect recognition	Dataset B	Dataset B	1600:1600

TABLE 5 MSCNN network architecture parameters.

Layer type	Branch	Kernel size	Stride	Padding	Channels (per block)	Output shape (per stage)
Input	-	-	-	-	1	1 × 1024
Conv1D + max Pool1D ×3	A	7	1	3	32 → 64 → 128	32 × 1024 → 64 × 512 → 128 × 256 → 128 × 128
Conv1D + max Pool1D ×3	B	5	1	2	32 → 64 → 128	32 × 1024 → 64 × 512 → 128 × 256 → 128 × 128
Conv1D + max Pool1D ×3	C	3	1	1	32 → 64 → 128	32 × 1024 → 64 × 512 → 128 × 256 → 128 × 128
Concat	All	-	-	-	384	384 × 128
Global avg Pool1D	-	-	-	-	384	384 × 1
FC1	-	-	-	-	256	256 × 1
FC2	-	-	-	-	4	4 × 1



As shown in Figure 7, the four types of acoustic signatures correspond to the inherent structure, minor crack, moderate crack, and major defect. In Figure 7a, a single prominent peak represents the inherent structure, the normal ultrasonic propagation path in a structurally intact shaft indicating no interference from cracks or defects. In Figure 7b, a minor secondary peak near 1700 mm suggests a small defect, such as a shallow crack, with limited structural impact. Figure 7c reveals a more pronounced secondary peak in the same region, indicating a medium-sized crack with higher energy reflection. In Figure 7d, the secondary peak near

1700 mm exhibits significantly greater amplitude than in the small and medium defect cases, indicating a major defect or severe material flaw that may require urgent intervention. Across all cases, the inherent structure is consistently observed, underscoring its role as a fundamental ultrasonic feature of the main shaft. Furthermore, the amplitude of the reflected peak increases with defect severity from small to large defects demonstrating that peak amplitude in ultrasonic signals serves as a reliable indicator for defect sizing and severity assessment. This finding provides a valuable basis for non-destructive evaluation of main shaft integrity in wind turbines.

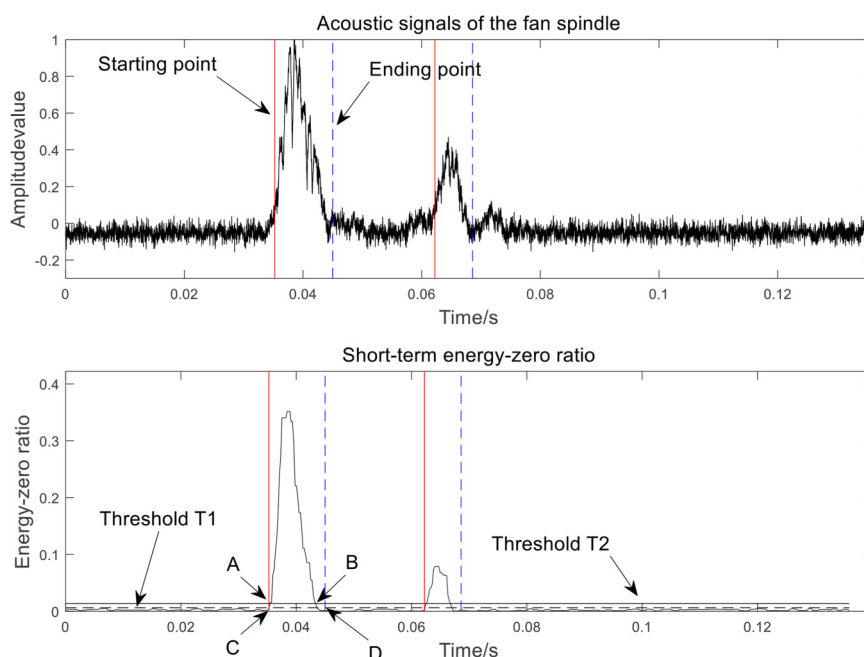


FIGURE 6 Segmentation of acoustic signature using EZR.

4.2 Model performance evaluation

4.2.1 Comparison of experimental results for multi-scale feature extraction

To assess the effectiveness of the improved MSCNN in extracting features from various crack defects, as well as in defect separation and identification, comparative experiments were conducted between a standard CNN with a fixed size convolution kernel and the proposed MSCNN model. The experimental analysis focuses on the performance differences between the two architectures. The respective loss functions and recognition accuracies are presented in Figure 8.

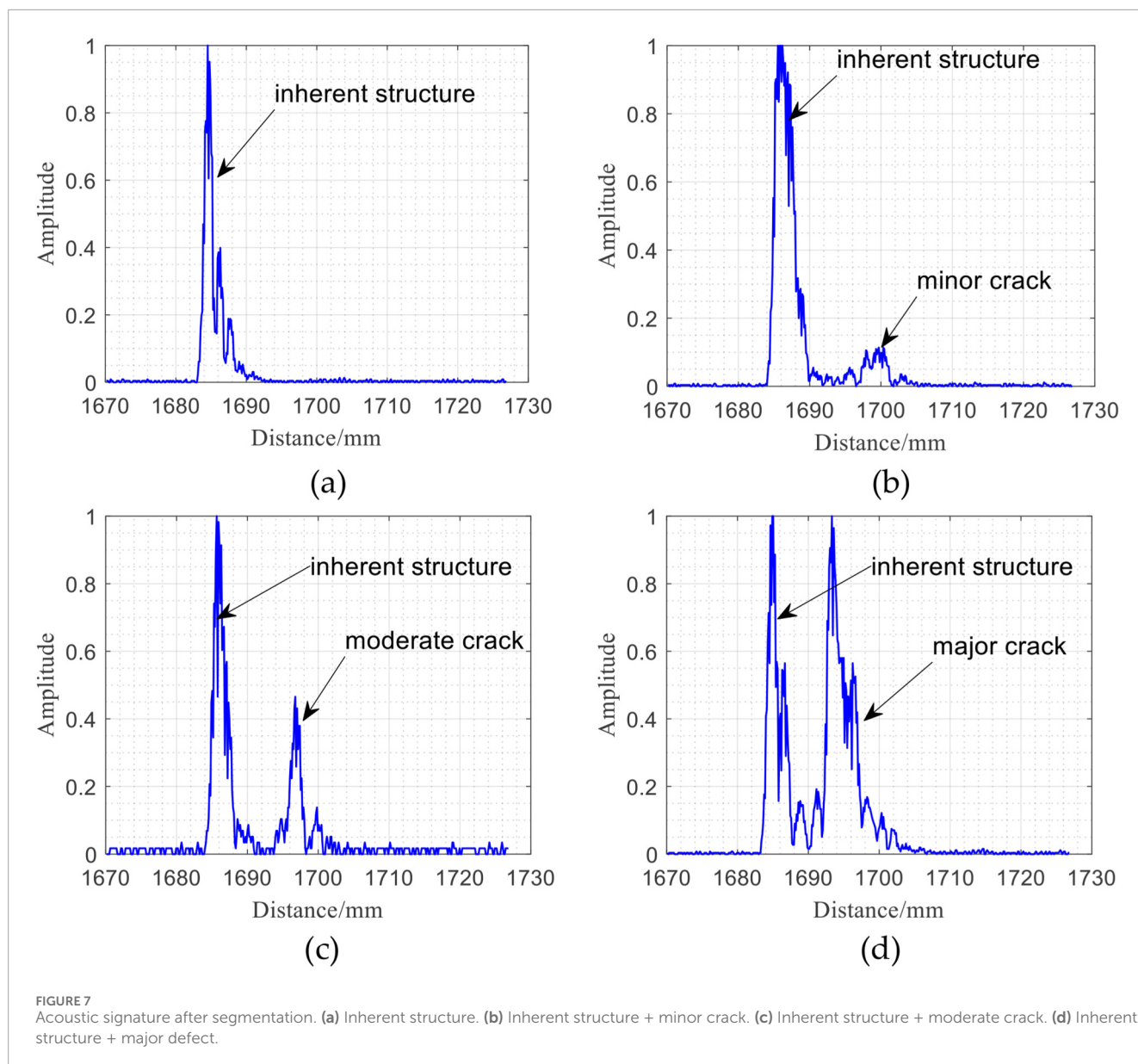
As illustrated in Figure 8, the loss function and recognition accuracy curves for both MSCNN and CNN models are plotted over 200 training iterations for both training and testing datasets. The MSCNN achieves recognition accuracies exceeding 90% for both datasets after only 20 iterations, with the corresponding loss function dropping below 0.1. The performance curves then stabilize, indicating convergence without overfitting. In contrast, the CNN requires more than 40 iterations to reach stability, demonstrating that MSCNN converges faster and requires less training time in practical applications. The MSCNN achieves a peak accuracy of 90.40%, outperforming the CNN's maximum accuracy of 85.26%, indicating the CNN's limitations in correctly classifying certain samples. This performance gain confirms MSCNN's superior ability to extract parallel multi-scale features, which is critical for accurate multi crack detection in real world diagnostic scenarios. The results further indicate that crack defect features exhibit distinct multi-scale temporal characteristics, often masked by noise in the raw signal. A single scale CNN struggles to capture such

complex patterns, whereas the MSCNN, by leveraging multi-scale convolution, effectively captures these variations, thereby improving detection and classification performance.

4.2.2 Comparison between traditional machine learning algorithms and MSCNN

To further evaluate the proposed deep learning approach for identifying crack defects in wind turbine spindles, a comparison was conducted with traditional machine learning methods. Previous studies (Chen et al., 2017; Wang et al., 2020; Zhou et al., 2016) have commonly applied discrete wavelet transform (DWT) to preprocess acoustic or vibration signals, followed by extracting features such as Mel frequency cepstral coefficients (MFCCs) or energy based descriptors. These were then classified using conventional algorithms such as support vector machines (SVM), back propagation (BP) neural networks, or ELM. In this study, these same features were used as input to SVM, BP, and ELM classifiers, and their recognition results are summarized in Figure 9.

As shown in Figure 9, the average recognition accuracies of the traditional models—SVM, BP, and ELM—are 76.45%, 79.56%, and 81.80%, respectively. These values are consistently lower than those of the deep learning models, including CNN and MSCNN. Among all models, MSCNN demonstrates the highest recognition performance, clearly surpassing traditional methods. This highlights the advantage of deep architectures, which can automatically learn and extract discriminative features from acoustic signals, overcoming the limitations of manually engineered features.



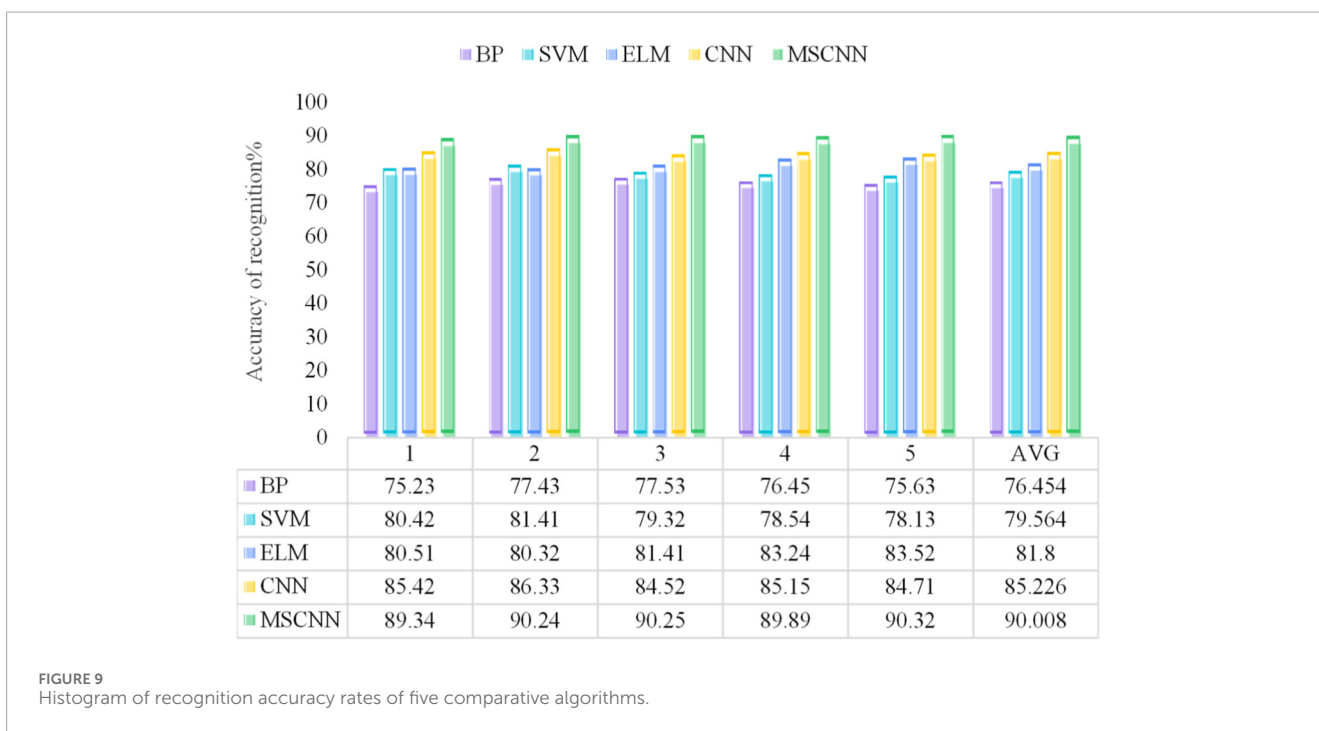
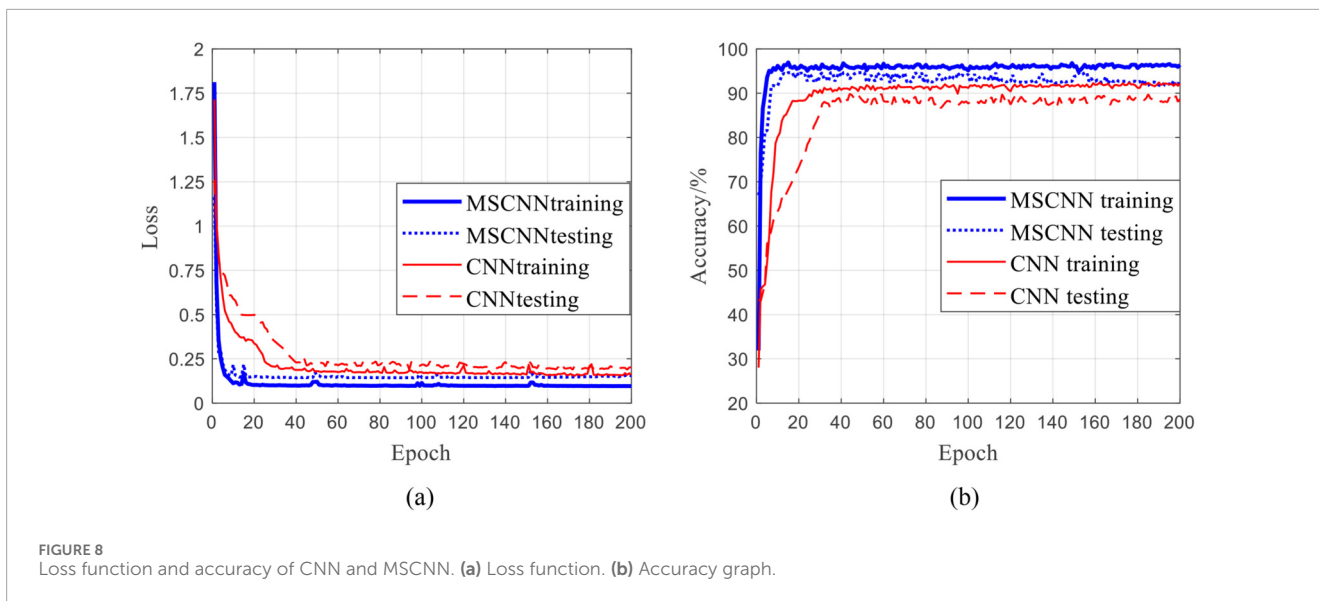
Unlike traditional algorithms, deep learning models capture features across multiple levels and temporal scales, enhancing classification accuracy. Specifically, the MSCNN achieves an average recognition accuracy of 90.008%, representing a 6.7% improvement over the best performing traditional model and a 10.01% gain compared to the standard CNN. By using convolution kernels of varying sizes to extract acoustic features in parallel at different temporal scales, MSCNN achieves richer and more discriminative feature representations, which are particularly effective for complex, multi type defects.

The experimental results confirm that MSCNN provides more comprehensive and accurate multi-level acoustic feature extraction, leading to significantly improved recognition performance. The use of multi-scale convolution kernels allows MSCNN to better capture transient signal characteristics, further validating its superiority over both traditional machine learning and single-scale CNN approaches.

4.2.3 Ablation study of MSCNN architectural components

A series of ablation experiments were conducted to quantify the contribution of key MSCNN components to overall performance. By individually removing or altering elements such as the number of branches, convolution kernel sizes, input feature types, network depth and width, fusion methods, and regularization strategies, their impact on the 10 run average accuracy was systematically evaluated.

Table 6 summarizes the ablation results, averaged over ten runs. The baseline configuration comprising three branches ($k \in \{7, 5, 3\}$), EZR input, GAP fusion, and dropout of 0.5 achieves an average accuracy of 89.39% and serves as the reference. Removing the multi-scale structure (A0) causes the largest performance drop (-5.04%), underscoring the critical role of multi-scale feature extraction in capturing discriminative patterns. Replacing the EZR input with raw time-domain signals yields a -5.26% decline, confirming the robustness and discriminative power of EZR



features. Reducing the number of branches from three to two (A3) leads to a -2.64% decrease, indicating that additional branches enhance feature diversity. Likewise, decreasing network depth (B1) or width (C1) reduces accuracy by -2.79% and -4.07% , respectively; conversely, increasing them (B3, C3) offers negligible gains ($+0.09\%$ and -0.25%), suggesting that the baseline already strikes a sound balance between complexity and generalization. Changing the fusion method from GAP to flattening (D2) reduces accuracy by -2.45% , and removing dropout (D4) decreases it by -2.89% , highlighting the importance of both fusion design and regularization. Adding an additional feature channel (EZR + ZCR, F2) provides only a marginal gain ($+0.03\%$), indicating limited

benefit relative to core components such as the multi-scale design and EZR input. Overall, these results demonstrate that multi-scale convolution, EZR input, balanced network depth and width, and GAP-based fusion are the most influential factors for achieving high recognition accuracy in crack defect identification.

4.3 Effect of multi-scale feature extraction

The experimental results demonstrate that the crack feature maps extracted by the MSCNN model at different scales exhibit strong discriminative ability. Compared with single-scale

TABLE 6 Ablation study on key architectural components of MSCNN.

Id	Variant	Avg. Acc (%)	Δ Acc	Key change
Ours	Baseline (3-branch; k = 7/5/3; EZR; GAP; dropout 0.5)	89.39	-	-
A0	Single branch (k = 3)	84.35	-5.04	Remove multi-scale
A3	Dual branch (k = 7,5)	86.75	-2.64	Reduce one branch
B1	Depth = 2 blocks	86.60	-2.79	Shallower network
B3	Depth = 4 blocks	89.48	+0.09	Deeper network
C1	Channels = 16→32→64	85.32	-4.07	Narrow network
C3	Channels = 64→128→128	89.14	-0.25	Wider network
D2	Flatten instead of GAP	86.94	-2.45	Change fusion method
D4	No dropout	86.50	-2.89	Remove regularization
F1	Raw time domain input	84.13	-5.26	Remove EZR
F2	EZR + ZCR (2-channel)	89.42	+0.03	Add extra feature

convolution, the multi-scale convolution approach significantly enhances the accuracy of crack detection.

As shown in Figure 10, the t-distributed Stochastic Neighbor Embedding (t-SNE) method is used to visualize the dimensionally reduced feature distributions obtained through different feature extraction methods applied to acoustic signals from the wind turbine main shaft (van der Maaten and Hinton, 2008). These include the raw input signal, features extracted by a CNN, and features extracted by the MSCNN model trained on either known or unknown defect types. Figure 10a shows that the raw signal lacks discernible structure, with feature points for various defects and inherent structural components appearing scattered and overlapping. This confirms that without feature extraction, differentiating between defect types is difficult, underscoring the necessity for advanced feature extraction methods. Figure 10b presents the feature distribution obtained by the CNN model. Compared with the raw signal, the CNN extracted features exhibit preliminary clustering for inherent structure, minor defects, moderate defects, and major defects. However, overlap remains—particularly between minor and moderate defects—indicating that while CNN improves separability, it still struggles with complex defect distinctions. In Figure 10c, the MSCNN model trained on known defect types yields distinctly separated clusters for each class. Feature points corresponding to inherent structure, minor, moderate, and major defects are tightly grouped, with minimal overlap. Notably, the boundary between minor and moderate defects is significantly clearer, demonstrating the superior feature extraction and classification performance of MSCNN when guided by known defect labels. Figure 10d shows the feature distribution generated by the MSCNN model without prior knowledge of defect types. While separability remains superior to that achieved by CNN, it is less distinct than in the known-defect scenario. Some overlap between defect classes persists, particularly between minor and moderate defects. This suggests

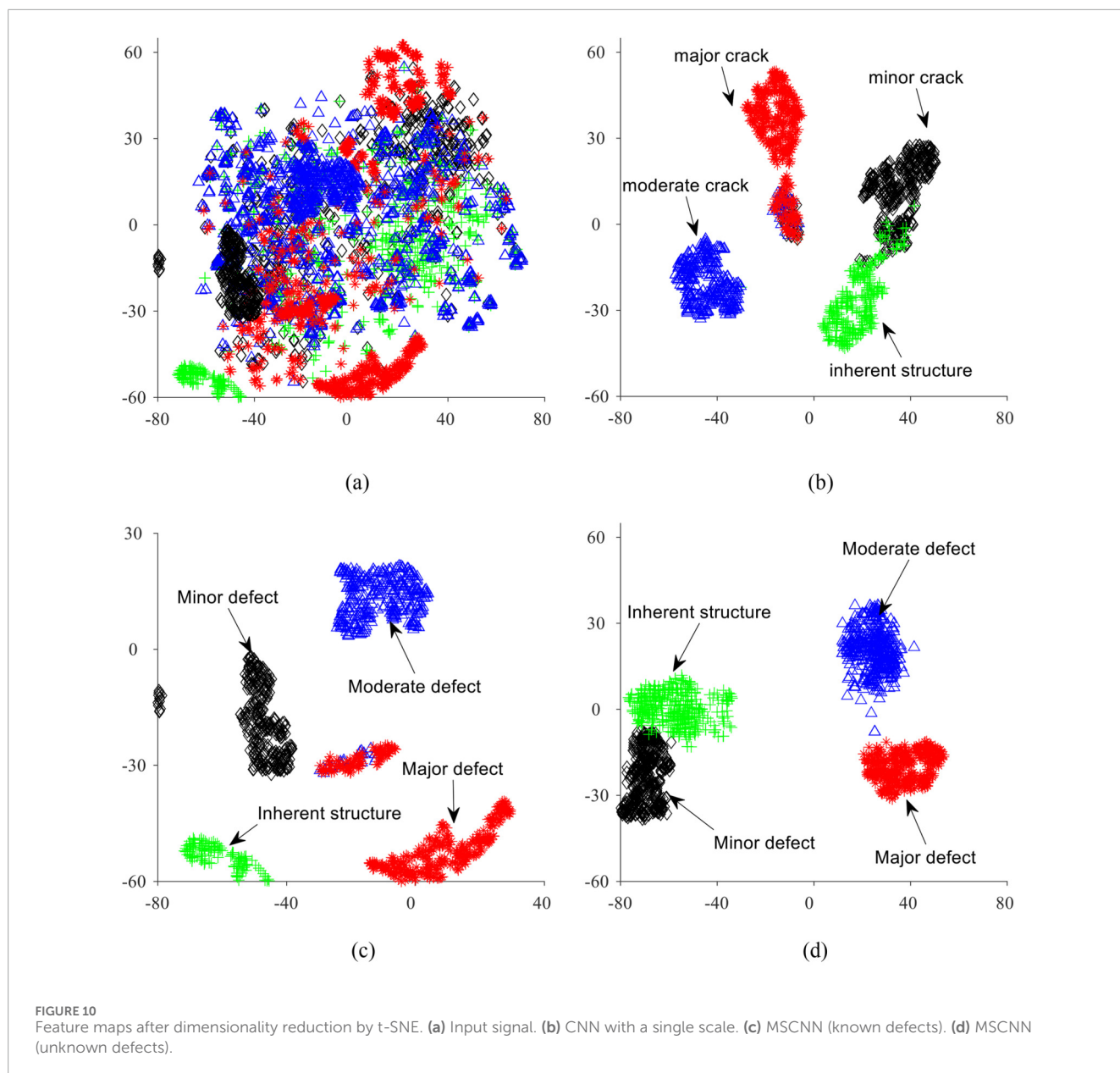
that although MSCNN can extract discriminative features in unsupervised contexts, the guidance of known labels markedly improves classification performance.

5 Conclusion

This study presents an intelligent crack detection method for wind turbine main shafts that leverages acoustic signature analysis and a MSCNN. The proposed EZR based segmentation algorithm effectively isolates crack features and demonstrates robustness in identifying both single and composite cracks. Experimental results show that multi scale feature learning outperforms single scale methods, achieving an average recognition accuracy of 90%, representing a 6.73% improvement over traditional models such as ELM and a 3.36% gain over single scale CNNs.

Despite these promising results, several practical challenges remain. Sensor installation during maintenance is time consuming, requiring 4–6 h per turbine. Computational efficiency for real time applications also needs further optimization, with current inference times of 47 m on standard hardware and 215 m on edge devices. Cross platform validation confirmed good adaptability, with accuracy ranging from 83.9% to 87.2% across different turbine models. However, environmental testing revealed performance degradation under extreme conditions, highlighting the need for compensation and adaptation techniques.

Future work will focus on improving computational efficiency for edge devices, integrating additional sensor modalities, and extending the system to predict crack propagation. Developing decision support algorithms for autonomous maintenance recommendations will also be a priority. Overall, the proposed method offers a practical and effective solution for wind turbine fault diagnosis, though further refinements are required for large scale industrial deployment.



Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LZ: Investigation, Writing – original draft, Software, Writing – review and editing. FL: Conceptualization, Writing – review and editing, Writing – original draft. SZ: Conceptualization, Writing – original draft, Investigation. XZ: Data curation, Project administration, Formal Analysis, Methodology, Writing – review and editing, Conceptualization, Writing – original draft. GH: Investigation, Supervision, Project administration, Writing – original draft, Methodology.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Yunnan Provincial Natural Science Foundation (Grant No. 202401AU070158), the Research Fund of Kunming University of Science and Technology (Grant No. KKZ3202465076), and an Industry–University Collaborative Research Project (Contract No. HZ2024K0212A).

Conflict of interest

Authors LZ, FL, and SZ were employed by CGN New Energy Investment (Shenzhen) Co., Ltd., Yunnan Branch.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that this work received funding from Industry–University Collaborative Research Project (Contract No. HZ2024K0212A). The funder had the following involvement in the study: study design and data collection. The funder had no role in the data analysis or interpretation, manuscript preparation, or the decision to publish.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

References

- Chen, J., He, Z., Zi, Y., Zhang, W., and Wang, Y. (2017). Wavelet transform-based feature extraction for fault diagnosis of wind turbine blades. *Renew. Energy* 102, 275–287.
- Chen, X., Zhang, B., and Gao, D. (2021). Bearing fault diagnosis based on multi-scale CNN and LSTM model. *J. Intelligent Manuf.* 32, 971–987. doi:10.1007/s10845-020-01600-2
- Chen, X., Lou, W., Zhao, W., Yang, G., Ding, K., and Zhang, J. (2024). A rolling bearing fault diagnosis method via 2D feature map of CSCoh after denoising and MSCNN under different conditions. *J. Vib. Control* 30 (5–6), 1241–1253. doi:10.1177/10775463231158739
- Cheng, J., He, C., Lyu, Y., Zheng, Y., Xie, L., and Wu, L. (2020). Ultrasonic inspection of the surface crack for the main shaft of a wind turbine from the end face. *NDT E Int.* 114, 102283. doi:10.1016/j.ndteint.2020.102283
- Ding, C., Zhao, M., Lin, J., and Jiao, J. (2019). Multi-objective iterative optimization algorithm based optimal wavelet filter selection for multi-fault diagnosis of rolling element bearings. *ISA Trans.* 88, 199–215. doi:10.1016/j.isatra.2018.12.010
- Feng, K., Wang, K., Zhang, X., Ni, Q., Zuo, M. J., and Chen, D. (2019). Specifying roller-bearing clearance for wind turbine gearboxes: an experimental and theoretical investigation. *IEEE Access* 7, 103911–103922.
- Fu, Q., Li, S., and Wang, X. (2020). MSCNN-AM: a multi-scale convolutional neural network with attention mechanisms for retinal vessel segmentation. *IEEE Access* 8, 163926–163936. doi:10.1109/access.2020.3022177
- Gbashi, S. M., Olatunji, O. O., Adedeji, P. A., and Madushele, N. (2024). From academic to industrial research: a comparative review of advances in rolling element bearings for wind turbine main shaft. *Eng. Fail. Anal.* 163, 108510. doi:10.1016/j.engfailanal.2024.108510
- Guo, J., Zhen, D., Li, H., Shi, Z., Gu, F., and Ball, A. D. (2019). Fault feature extraction for rolling element bearing diagnosis based on a multi-stage noise reduction method. *Measurement* 139, 226–235. doi:10.1016/j.measurement.2019.02.072
- Hassan, Q., Viktor, P., Al-Musawi, T. J., Mahmood Ali, B., Algburi, S., Alzoubi, H. M., et al. (2024). The renewable energy role in the global energy transformations. *Renew. Energy Focus* 48, 100545. doi:10.1016/j.ref.2024.100545
- Hinton, G. E., Osindero, S., and Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18 (7), 1527–1554. doi:10.1162/neco.2006.18.7.1527
- Huang, Y., and Wang, Y. (2019). “Multi-format speech perception hashing based on time-frequency parameter fusion of energy zero ratio and frequency band variance,” in *Proceedings of the 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, Xiamen, China, 18–20 October 2019 (IEEE), 243–251. doi:10.1109/EITCE47263.2019.9094822
- Kumar Dora, B., Bath, S., Mitra, A., Ernst, D., Halinka, A., Zychma, D., et al. (2025). The global electricity grid: a comprehensive review. *Energies* 18 (5), 1152. doi:10.3390/en18051152
- Nejad, A. R., Keller, J., Guo, Y., Sheng, S., Polinder, H., Watson, S., et al. (2022). Wind turbine drivetrains: state-of-the-art technologies and future development trends. *Wind Energy Sci.* 7 (1), 387–411. doi:10.5194/wes-7-387-2022
- Peng, C., Li, H., Gui, W., Tang, Z., and Yuan, X. (2025). Fault diagnosis method for rotating machinery based on MSCNN-MGAT. *IEEE Trans. Instrum. Meas.* 74, 2540511. doi:10.1109/tim.2025.3587368
- Santelo, T. N., de Oliveira, C. M. R., Maciel, C. D., and de A. Monteiro, J. R. B. (2022). Wind turbine failures review and trends. *J. Control Autom. Electr. Syst.* 33, 1–17. doi:10.1007/s40313-021-00789-8
- Teng, W., Ding, X., Cheng, H., Han, C., Liu, Y., and Mu, H. (2019). Fault diagnosis for a wind turbine planetary gearbox via novel method based on an extended cepstrum and MOMEDA. *Appl. Sci.* 9, 4355.
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Wang, T., and Chen, Y. (2023). Review of acoustic emission signal analysis for structural health monitoring. *Sensors* 23 (1), 312.
- Wang, W., Xu, F., Li, J., Huang, L., and Guo, W. (2019). Order spectrum analysis for planetary gearbox fault detection via joint amplitude and frequency demodulation. *Shock Vib.* 2019, 9086538.
- Wang, X., Zhang, Y., Liu, Z., and Li, H. (2020). Fault diagnosis of wind turbine gearbox using ELM and wavelet packet energy entropy. *Measurement* 165, 108076.
- Wang, Y., Huang, Y., Zhang, R., and Zhang, Q. y. (2021). Multi-format speech biotransforming based on energy to zero ratio and improved lp-mmse parameter fusion. *Multimed. Tools Appl.* 80, 10013–10036. doi:10.1007/s11042-020-09701-z
- Xia, J., Huang, X., Wang, X., and Qiao, H. (2020). A fault diagnosis approach for gears using improved spectral kurtosis, ensemble intrinsic time-scale decomposition and correlated feature selection. *Appl. Sci.* 10, 1879.
- Zhao, D., Tian, C., Fu, Z., Zhong, Y., Hou, J., and He, W. (2025). Multi-scale convolutional neural network combining BiLSTM and attention mechanism for bearing fault diagnosis under multiple working conditions. *Sci. Rep.* 15, 13035. doi:10.1038/s41598-025-96137-w
- Zhou, Z., and Wu, Y. (2022). Few-shot learning for intelligent fault diagnosis: a survey. *IEEE Trans. Ind. Inf.* 18 (6), 3762–3772.
- Zhou, D., Yang, Q., Guo, Y., and Huang, S. (2016). Wind turbine blade damage detection using wavelet packet decomposition and BP neural network. *Mech. Syst. Signal Process.* 70–71, 103–115.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.