machines

MDPI

Article

Multi-Source Fusion CNN-RF Framework for Intelligent Fault Diagnosis of Head Sheave Devices in Mining Hoists

Chi Ma [1,*], Jian Fei [1], Zhiyuan Shi [2], Md Abdur Rob [1], Md Ashraful Islam [1] and Md Habibullah [3]

[1] School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou 221008, China
[2] Mining Products Safety Approval and Certification Center Co., Ltd., Beijing 100013, China
[3] School of Mines, China University of Mining and Technology, Xuzhou 221008, China
* Correspondence: machi@cumt.edu.cn

Abstract

Accurate fault diagnosis of mining hoisting head sheave systems is critical for ensuring operational safety in harsh underground environments. This study proposes a multi-source fault diagnosis framework that fuses vibration and acoustic information using a Convolutional Neural Network and Random Forest (CNN-RF). To support mechanism understanding and validate the experimental platform, finite element and multi-body dynamics simulations (ANSYS/ADAMS) are employed for physical verification and fault signature analysis, while the CNN-RF model is trained and tested exclusively using experimentally acquired vibration and acoustic data. For feature construction, vibration signals are transformed into time–frequency representations (including STFT, CWT, and generalized S-Transform (GST)), and acoustic signals are characterized using Mel-Frequency Cepstral Coefficients (MFCCs). Experimental results demonstrate that vibration–acoustic fusion improves diagnostic performance compared with single-modality baselines; the best performance is achieved by GST+MFCC with the proposed CNN-RF classifier, reaching an accuracy of 98.96%. Future work will conduct cross-condition validation under varying speeds and loads and investigate missing-modality robustness to further assess generalization and deployment reliability.

Keywords: mining hoist; head sheave device; fault diagnosis; modal analysis; CNN-RF fusion; multi-source signals

Check for updates

Academic Editors: Xuefang Xu, Zijian Qiao, Mengdi Li and Peng Mei

Received: 14 January 2026
Revised: 15 February 2026
Accepted: 18 February 2026
Published: 21 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Mining hoists represent a cornerstone of modern coal extraction, facilitating the vertical transport of personnel, equipment, and ore across shafts often exceeding 1000 m in depth. At the heart of these systems lies the head sheave device, a critical load-bearing assembly comprising the wheel body, shaft, and bearings, which endures extreme dynamic stresses up to 600 kN in tension at rotational speeds of 10–15 RPM while guiding multi-strand wire ropes [1,2]. In harsh underground environments characterized by dust, humidity (>80%), and seismic vibrations, faults in this component, such as wheel body cracks from cyclic fatigue, shaft torsional breathing, or bearing pitting from spalling, contribute to approximately 35% of hoist-related incidents, including rope slippage, over-winding, and catastrophic structural failures [3,4]. These events not only escalate operational downtime, costing mining firms over $50,000 per hour, but also pose severe safety risks, with global statistics reporting over 200 hoist accidents annually, underscoring the urgent need for

proactive diagnostics [5]. Traditional maintenance paradigms, often reliant on periodic visual inspections and simple vibration thresholding, can be effective for some evident fault conditions; however, they may lack sensitivity to incipient faults under non-stationary signals and strong environmental noise, which can delay detection and increase the risk of reactive repairs [6]. The transition to Industry 4.0 in mining demands intelligent, non-invasive prognostics that leverage multi-sensor fusion and artificial intelligence (AI) for real-time anomaly detection [7]. Recent advancements in vibration analysis, such as empirical mode decomposition (EMD) variants, have improved signal denoising, while acoustic emissions offer early precursors to mechanical degradation [8,9]. However, unimodal approaches struggle with signal overlap in low-speed, heavy-load scenarios typical of hoists, where bearing fault frequencies (e.g., ball pass inner/outer) blend with structural harmonics [10].

This paper addresses these challenges by proposing an application-oriented, multi-source fault diagnosis framework for head sheave systems, integrating physics-based modal analysis, rigid–flexible simulations via ANSYS and ADAMS, and a hybrid Convolutional Neural Network–Random Forest (CNN-RF) classifier. Modal theory and simulation are used to verify mechanical behavior and fault signatures, and signal processing (e.g., S-Transform/GST for vibration and MFCC for acoustics) is used to extract complementary features for diagnosis [10–13]. The CNN backbone learns discriminative patterns from fused representations, while the RF stage improves classification stability and interpretability via SHAP analysis [14,15]. This work is positioned as an application-oriented engineering study that systematically integrates established techniques to solve a head sheave fault-diagnosis problem under practical mining constraints, rather than introducing a new learning algorithm [16–19]. Although the proposed diagnosis pipeline is generic and can be extended to other rotating machinery, the head sheave operates under a low-speed, heavy-load regime that generates distinctive vibration–acoustic characteristics and fault signatures. Therefore, the trained CNN-RF model is tuned to the data distribution associated with head sheave dynamics, and direct transfer to other machines or operating regimes may not be plug-and-play. In practical applications, reliable deployment to a new rotating asset would typically require collecting representative target-domain data and performing retraining or fine-tuning to account for differences in speed/load conditions and structural responses.

Therefore, the objective of this research is to develop an intelligent, high-fidelity fault diagnosis framework for the head sheave assembly of mine hoisting systems by integrating multi-source sensing, rigid–flexible coupling simulations, and hybrid deep learning models. Specifically, the study aims to (1) analyze failure mechanisms and their modal characteristics; (2) construct a rigid–flexible virtual prototype for simulating dynamic fault responses; (3) deploy a field-ready monitoring system capturing vibration, acoustic, displacement, and temperature data; and (4) design a multi-source fusion CNN-RF diagnostic model that achieves accurate, interpretable fault classification under real-world operating conditions. This research not only bridges the gap between theoretical modeling and practical diagnostics but also contributes a scalable methodology for intelligent health monitoring in heavy industrial machinery.

2. Related Works

Fault diagnosis in mining hoists, particularly for head sheave devices, has evolved from rule-based heuristics to data-driven paradigms, driven by the need to mitigate failures in rotating machinery under variable loads [20]. Early efforts focused on fault tree analysis (FTA) and empirical models to map sheave risks, such as swing-induced vibrations from headframe inclination [1,21]. Chen et al. employed FTA alongside vibration thresholding to diagnose bearing faults and sheave misalignment, achieving 85% reliability but limited

by manual feature selection in noisy spectra [1]. Similarly, Wang et al. integrated sensor fusion for hoist control, using Kalman filters to isolate faults, yet overlooked shaft-specific torsional modes prevalent in multi-rope systems [3,22].

Signal processing advancements have enhanced feature extraction for non-stationary hoist signals. Li et al. applied ensemble EMD (EEMD) with permutation entropy to detect rope tension anomalies, reporting 92% accuracy via PSO-optimized SVM, though susceptible to mode mixing in low-frequency sheave rotations (fr $\approx$ 12 Hz) [5,8]. Zhang et al. extended this with ICEEMDAN for bearing entropy features in sheave assemblies, fusing them into AFSA-SVM classifiers (96% precision), but single-source vibration ignored acoustic emissions from incipient cracks [2]. Acoustic-based methods, like MFCC for emission localization, complement vibration [13]; Yin et al. proposed IoT-embedded acoustic monitoring for hoists, yielding 90% sensitivity [4], yet fusion strategies remained rudimentary, prone to multicollinearity (VIF > 10) [23]. More broadly, connectivity-enabled cooperative sensing and control has become an important paradigm for improving resilience in safety-critical cyber–physical systems, spanning both industrial monitoring and intelligent transportation [24]. For example, V2X-enabled cooperative platoon control has been shown to improve system-level resilience under mixed-traffic congestion dynamics, illustrating how connectivity can support coordinated sensing and control beyond industrial monitoring scenarios [25].

Simulation-driven diagnostics bridge theoretical models with empirical data. Modal analysis via ANSYS has quantified sheave stiffness losses, with Liu et al. simulating 15% frequency drops for 20% damage, validated against Rayleigh quotients (error < 3%) [26]. ADAMS-based rigid–flexible coupling, as in Huang et al., emulated wire-rope dynamics in hoists, generating envelopes for fault harmonics (correlation r = 0.92), essential for scaling experiments without full-scale risks [11,27]. Multi-body co-simulations further incorporated Hertzian contacts for bearing pits, but integration with AI lagged, often relying on PCA for dimensionality (retaining 95% variance) [19,28].

Deep learning (DL) has revolutionized rotating machinery fault diagnosis, surpassing traditional machine learning (ML) in handling high-dimensional data; however, as highlighted by Bao et al., bridging the gap between theoretical models and actual industrial deployment remains a critical challenge, requiring advanced architectures capable of overcoming the severe environmental noise and off-design conditions inherent to complex operating environments [29,30]. CNNs excel in spectrogram classification; You et al. used 1D-CNN on S-Transformed vibrations for bearings, achieving 97% accuracy [12], while Wen et al. fused multi-channel CNNs for hoists (AUC = 0.98) [16,31]. Hybrid models address DL's overfitting: Li et al. combined CNN with RF for interpretable features in wind turbines, boosting F1-scores by 8% via ensemble voting [14,15]. In mining contexts, a CBMA model integrated spatiotemporal fusion for rotating faults (98.5% accuracy) [7], and MSResNet-Class Activation Mapping (CAM) handled noise (SNR = 10 dB) with attention mechanisms [6,32]. Gramian Angular Field-enhanced DGCN visualized temporal dependencies, outperforming LSTMs by 5% in generalization [8,33].

Despite progress, gaps persist: (1) sheave-specific studies undervalue shaft/wheel faults, focusing on bearings [21]; (2) unimodal processing ignores multi-source synergies, elevating false positives by 15–20% [23]; (3) sim-to-real transfers lack empirical validation undermine stressors (e.g., 80% humidity) [25]; and (4) although CNN-RF hybrids have been widely explored for rolling bearing diagnosis, their validated application to mining hoisting head sheave systems especially with multi-source (vibration–acoustic) fusion and systematic fusion-level ablation remains comparatively limited in the open literature [34–37]. Accordingly, the primary contribution of this work lies in engineering

integration and validation: combining simulation-based verification, multi-source sensing, and interpretable learning into a practical diagnostic pipeline for head sheave equipment.

3. Methodology

This study presents a comprehensive methodology for the fault diagnosis of the hoist head sheave device in mining operations, addressing the challenges of heavy-load dynamics, environmental harshness, and early fault detection. The approach integrates theoretical analysis of failure mechanisms, finite element-based simulations for virtual prototyping, experimental validation through controlled testing, advanced signal processing for multi-source data fusion, and a hybrid deep learning model for automated classification. The pipeline is designed for reproducibility and scalability, leveraging established tools: SolidWorks 2022 for 3D modeling, ANSYS 2023 for modal analysis, ADAMS 2023 for multi-body dynamics, MATLAB R2023a for signal processing, and Python 3.10 (with TensorFlow 2.12 and scikit-learn 1.3) for ML. The dataset encompasses simulated signals (n = 2000 samples) and experimental recordings (n = 1200 samples) across five fault classes: normal operation, wheel body crack, shaft fatigue, bearing inner ring fault, and bearing outer ring fault. Each class includes balanced representations under variable operational conditions (lift speeds: 5–15 m/s; rope tension: 300–600 kN). Figure 1 presents the integrated research framework, highlighting the two-stage pipeline of rigid–flexible model construction and validation-driven intelligent fault diagnosis.

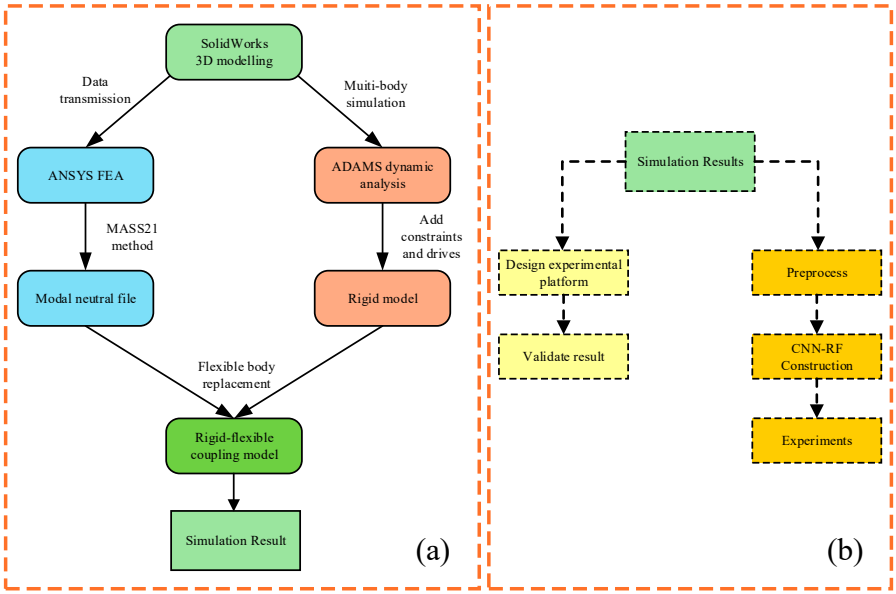


Figure 1. Overall methodology schematic diagram. (**a**) Establishment of rigid–flexible coupling model process; (**b**) result validation and deep learning-based fault detection process.

The methodology progresses sequentially: from theoretical foundations to simulation, signal preparation, experimental corroboration, and diagnostic modeling. This ensures a robust bridge between abstract mechanics and practical AI-driven insights, with validation at each stage to minimize sim-to-real discrepancies.

3.1. Theoretical Foundation: Failure Mechanisms and Modal Features

From the fundamental principles of structural dynamics, all types of mechanical components can be regarded as dynamic systems composed of key parameters such as stiffness characteristics, mass distribution, and damping coefficients. When defects or damage occur in the system, these fundamental parameters inevitably undergo corresponding changes, leading to alterations in the system's vibration characteristics. This manifests as significant

differences in modal parameters such as modal frequencies, mode shapes, and frequency response functions. Based on this principle, diagnostics can identify and evaluate structural health by monitoring changes in dynamic characteristics. Specifically, effective damage diagnosis methods are established by comparing modal parameter differences between intact and damaged states. This modal analysis-based damage identification technology provides crucial theoretical foundations and technical tools for structural health monitoring. This paper primarily employs modal frequency as the damage identification parameter.

To address practical challenges, the finite element method discretizes structural systems from infinite degrees of freedom to finite ones. Most reliable structural damage identification methods use finite element models as reference standards. These models provide benchmarks for structural damage, enabling detection by comparing modal parameters between current and reference models. Typically, damage identification begins with simple, cost-effective methods to determine whether damage has occurred. Once confirmed, complex analytical methods are employed for comprehensive damage detection.

Due to their accessibility and high precision, modal frequencies are the preferred choice for structural damage detection. Monitoring changes in modal frequencies offers a method that is not only simple and practical but also allows flexible adjustment of measurement point layouts according to specific requirements. Consequently, damage detection using modal frequencies holds significant advantages in practical applications [38].

In engineering vibration analysis, most mechanical vibration systems can be simplified using discretization methods into dynamic systems with finite degrees of freedom. After discretizing the structure using the finite element method, its motion characteristics can be described by nth-order matrix differential equations:

$$M\ddot{x} + C\dot{x} + Kx = f(t) \tag{1}$$

where M, C, K denote the quality, damping, and stiffness matrices, all of order $m \times n$ order, typically a symmetric matrix; x, $\dot{x}$, $\ddot{x}$ represents displacement, velocity, and acceleration arrays, all of order $n \times 1$ step; $f(t)$ denotes the external excitation array, with order $n \times 1$ step. When the mechanical energy loss of a system is negligible, it can be regarded as a conservative system. For systems possessing n, a one-degree-of-freedom undamped vibrating system, whose differential equation is as follows:

$$M\ddot{x} + Kx = f(t) \tag{2}$$

Since the above equation is non-homogeneous, the general solution structure of non-homogeneous linear ordinary differential equations possesses a well-defined mathematical expression. Specifically, the complete solution to such equations can be decomposed into a linear superposition of two components: one is the general solution corresponding to the homogeneous equation, representing the system's free response characteristics; the other is the particular solution to the non-homogeneous equation, reflecting the forced response under external excitation. For the free vibration of a structure:

$$M\ddot{x} + Kx = 0 \tag{3}$$

Suppose, $x = \varphi e^{j\omega t}$, where φ is the amplitude matrix of the free response, and the order is the $n \times 1$ order. Then, the expression of the homogeneous equation in the frequency domain of the above equation is:

$$K\varphi = \omega^2 M\varphi \tag{4}$$

The eigenvalue is represented in the ω^2 equation, where φ represents the eigenvector. Introduce a variable λ, where $\lambda = \omega^2$, and λ_i is the eigenvalue; instead of φ_i, it is the

first-order i normalized displacement modal vector ($\varphi_i^T \varphi = 1$). From this, the following relationship can be obtained:

$$K\varphi_i = \lambda_i M\varphi_i \tag{5}$$

In large mechanical structures such as head sheave devices, damage often leads to a significant decrease in structural stiffness with relatively little impact on mass distribution, so there will be a small change in the stiffness matrix of the structure, which in turn leads to λ_i and φ_i consequent changes. For the vibration of the damaged structure, its perturbation equation can be expressed as:

$$(K + \delta K)(\varphi_i + \delta\varphi_i) = (\lambda_i + \delta\lambda_i)M(\varphi_i + \delta\varphi_i) \tag{6}$$

The δK change in the overall stiffness matrix is the $\delta\lambda_i$ change in eigenvalue, $\delta\varphi_i$, which is the change in eigenvector, which can be obtained by using the first-order approximation method and then combining with Equation (5) to form the following:

$$\delta K\varphi_i - \lambda_i M\varphi_i = -(K - \lambda_i M)\delta\varphi_i \tag{7}$$

Using φ_i^T, the left multiplication Equation (7) formula is combined with the Equation (5) formula, and we can get:

$$\delta\lambda_i = \varphi_i^T \delta K\varphi_i \tag{8}$$

According to the above equation, when the structure is damaged, $\delta\lambda_i$ the change is δK in a linear relationship, and because $\lambda = \omega^2$ of it, $\delta\varphi_i$ will be affected by $\delta\lambda_i$, which is the change. Therefore, if the structure is damaged, its modal frequency will also change, and this phenomenon is described in many studies as a frequency drift characteristic after structural damage [39,40]. In addition, when the damage location is fixed, the more severe the damage, the greater the frequency change.

3.2. Simulation: Rigid–Flexible Coupling Modeling

This subsection employs virtual prototyping to replicate fault propagation under controlled, repeatable conditions, which is crucial for fault mechanism analysis and physical verification under controlled, repeatable conditions ethically and cost-effectively essential in mining where physical fault induction risks equipment damage or safety hazards while allowing parametric sensitivity analysis to refine theoretical predictions.

3.2.1. 3D Parametric Modeling

The sheave assembly is constructed in SolidWorks using field-measured dimensions: wheel body (outer diameter Ø3500 mm, thickness 800 mm, material ZG270-500 steel, density 7200 kg/m^3); shaft (length 8 m, diameter 400 mm, material 40Cr alloy steel, yield strength 785 MPa); bearings (6204 deep-groove ball type, GCr15 steel, preload 5 kN). Assemblies incorporate rope contact (friction coefficient μ = 0.15) and gravitational loads (total mass ≈ 8 tons). Models are exported as STEP files to ANSYS for meshing.

3.2.2. Modal Analysis in ANSYS

Finite element analysis employs a tetrahedral mesh (element size 15–25 mm, total elements ≈ 150,000; convergence criterion: <1% change in ω1 with refinement). Boundary conditions include fixed supports at shaft ends and free rotation for the wheel. Damping is modeled as Rayleigh proportional (α = 0.05, β = 0.001 s) [41]. A total of 50 modes in the 0–200 Hz band were extracted. To support rigid–flexible coupling, 360 interface nodes were defined on the bearing raceway and retained as master nodes for Craig-Bampton reduction, after which the flexible body was exported as an MNF file using mm-kg-s units. Damage scenarios reduce Young's modulus E by 20% in targeted elements (e.g., crack simulation

via reduced integration). Outputs include frequency tables, mode shapes (deformation contours), and participation factors, with validation against analytical Rayleigh quotients (error < 3%). Figure 2 illustrates the modular workflow for the modal analysis in ANSYS, encompassing preprocessing, solver execution, and result verification stages.

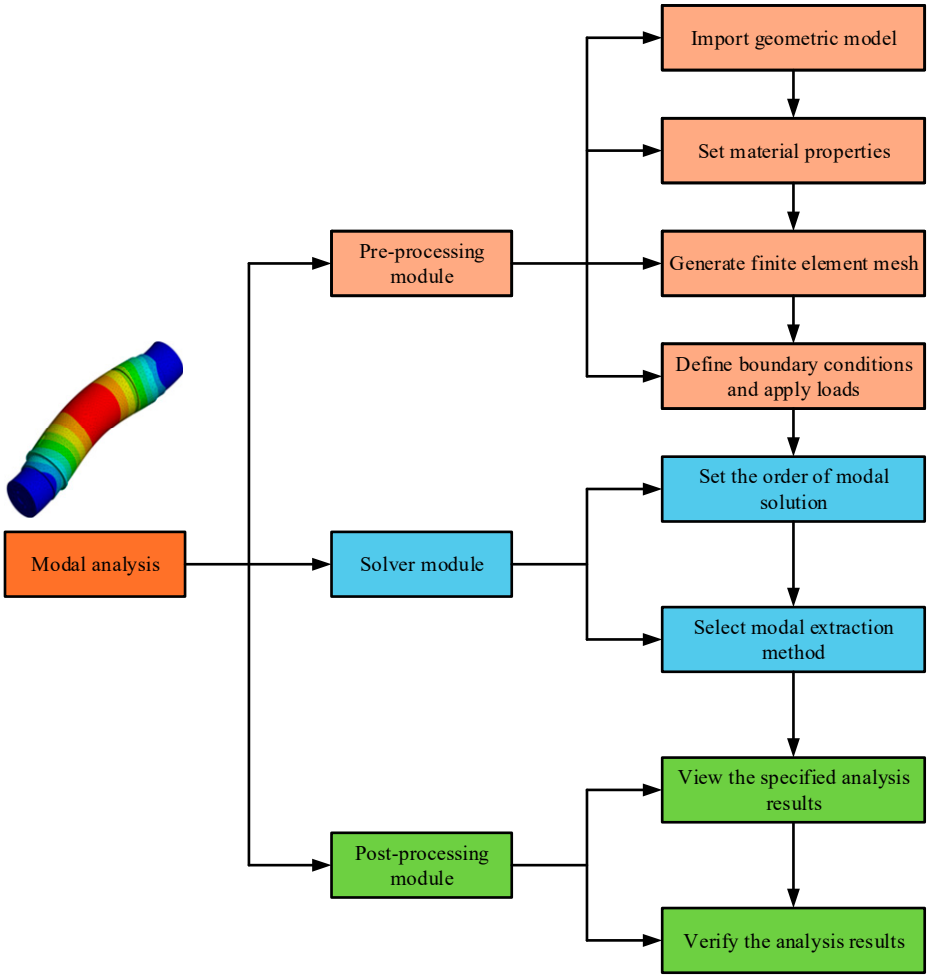


Figure 2. Workflow for modal analysis process in ANSYS showing preprocessing, solver, and postprocessing modules.

This systematic procedure ensures that the modeled dynamic behavior aligns closely with the actual structural response of the sheave assembly.

3.2.3. Multi-Body Dynamics in ADAMS

The rigid–flexible coupled dynamic model of the head sheave system is established in MSC ADAMS. The sheave wheel (and shaft, if applicable) is imported as a flexible body using the Modal Neutral File (MNF) generated in ANSYS, and the flexible wheel body is assembled with the rigid bearing components in the multi-body environment. To realize the interaction between the flexible raceway and the rigid rollers, 360 interface nodes retained in the MNF are used as contact coupling nodes, and contact is defined between these flexible nodes and the rigid roller surfaces. The flexible–rigid contact is modeled using the IMPACT (solid–solid) function, in which the normal contact force is expressed as $F_n = K\delta^n + C\dot{\delta}$, where δ is the penetration depth and $\dot{\delta}$ is its rate. In this study, the contact parameters are set to $\mathrm{K} = 164{,}000\ \mathrm{N/mm^{1.5}}$, $\mathrm{n} = 1.5$, and $\mathrm{C} = 145\ \mathrm{N \cdot s/mm}$, with a friction coefficient of $\mu = 0.15$. The coupled equations of motion are solved using the GSTIFF integrator with a time step of 1×10^{-4} s and an error tolerance of 1×10^{-6}. Each simulation is run for 10 s, which is sufficient to cover approximately 10 bearing revolutions, ensuring stable

fault-frequency extraction under the coupled rigid–flexible dynamics. Figure 3 illustrates the generalized multi-body dynamics-solving process employed in ADAMS.

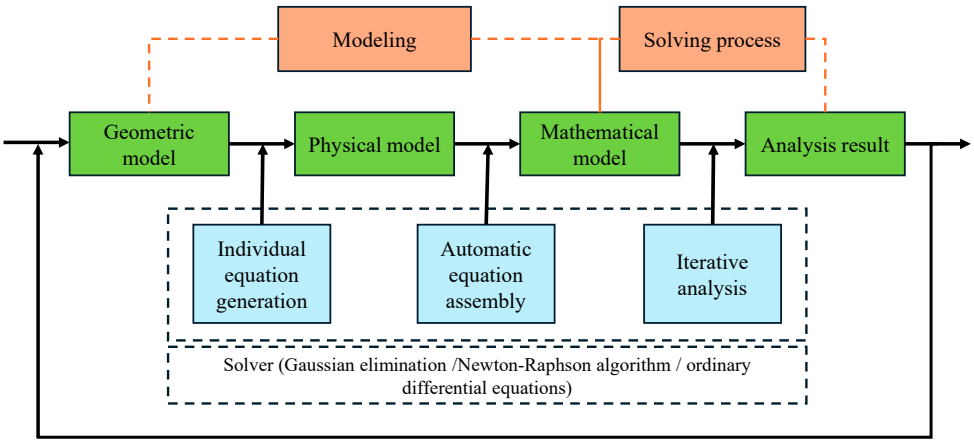


Figure 3. Solution workflow for multi-body dynamics in ADAMS integrating geometric, physical, and mathematical modeling with numerical solvers.

It encompasses geometric and physical modeling, transformation into mathematical expressions, and numerical solution through solvers such as Gauss elimination, Newton–Raphson, and ODE integrators. This modeling framework ensures a high-fidelity digital twin of the hoisting system, laying the foundation for accurate virtual testing and feature generation.

3.3. Signal Processing

Effective fault diagnosis of the hoist head sheave device necessitates the extraction of discriminative features from non-stationary, multimodal signals generated under varying operational and environmental conditions. To this end, a two-channel signal processing framework is implemented, targeting vibration and acoustic emission (AE) signals. Each modality is processed through dedicated pipelines to capture complementary features, subsequently fused to form an enriched representation for classification.

3.3.1. Preprocessing

To ensure the diagnostic system effectively distinguishes fault-related patterns in the head sheave device, preprocessing is employed as a foundational step to enhance signal quality, remove irrelevant noise, and normalize inputs across varying operational states. Vibration and acoustic signals are simultaneously acquired using triaxial piezoelectric accelerometers and high-sensitivity directional microphones, respectively, each sampled at 10 kHz to preserve transient fault components. Prior to feature extraction, the vibration signals undergo a structured preprocessing pipeline comprising de-trending to eliminate quasi-static displacement components and zero-phase Butterworth bandpass filtering within the 0.5–2.5 kHz range. This frequency band is empirically chosen to capture bearing fault harmonics, structural resonance, and shaft-related vibrational modes, while suppressing low-frequency drift and high-frequency environmental noise. Each signal stream is segmented into overlapping windows of 1024 samples with 50% overlap using a Hann window to minimize spectral leakage during subsequent transformations. For acoustic signals, a pre-emphasis filter is applied to amplify high-frequency content typically associated with stress-wave emissions and material degradation. The acoustic waveform is then partitioned into 25 ms frames with 10 ms overlap, followed by Hamming windowing to preserve temporal continuity and reduce boundary effects during spectral transformation. Both signal types are standardized using z-score normalization to mitigate

the influence of sensor variability and operational amplitude fluctuation. Outlier detection, based on kurtosis thresholds and envelope variance monitoring, is implemented to identify and discard low signal-to-noise ratio (SNR) segments or drifted traces. After preprocessing, all signal segments are temporally aligned and encoded into a structured tensor format that serves as the input for subsequent time–frequency transformation and model training. This preprocessing framework ensures robust feature representation under complex loading conditions and lays the foundation for high-performance fault classification.

3.3.2. Vibration Signal Analysis

Vibration signals serve as the primary modality for capturing the dynamic response of the head sheave device under operating conditions characterized by cyclic loading, structural resonance, and mechanical wear. These signals inherently exhibit non-stationary behavior due to the combined influence of shaft rotation, rope tension fluctuation, and fault-induced impacts. To effectively extract fault-related features embedded in such complex signals, time–frequency domain analysis is adopted, with a particular emphasis on the S-Transform (ST) for its superior joint resolution and interpretability. The vibration acceleration sensor measurement points are shown in Figure 4.



Figure 4. Measuring point of vibration sensor.

The ST provides a scalable time–frequency representation that maintains absolute phase information while offering frequency-dependent windowing. This is particularly advantageous for detecting transient fault signatures such as bearing pitting, fatigue-induced cracks, and shaft imbalance. In this study, segmented vibration signals preprocessed as described in Section 3.3.1 are transformed using the ST to produce two-dimensional scalograms. These scalograms highlight the evolution of energy content across frequency bands over time, thereby enabling localization of fault-sensitive spectral patterns.

In addition to scalogram generation, several time and frequency domain features are extracted to enhance the representational capacity of the model. These include statistical features such as Root Mean Square (RMS), Kurtosis, and Crest Factor, which are sensitive to energy variations and impulsiveness associated with structural defects. Furthermore, envelope analysis is performed using the Hilbert Transform to demodulate amplitude-modulated components and identify key fault frequencies, including Ball Pass Frequency of Inner and Outer Race (BPFI and BPFO), fundamental shaft frequency (f shaft), and sideband structures linked to modulation effects.

To facilitate compatibility with deep learning architectures, the resulting scalograms and statistical feature vectors are reshaped into image-like inputs and normalized across the dataset. These representations serve as inputs to the CNN-RF model, allowing for both spatial feature learning and interpretable classification. The use of vibration-based S-Transform analysis not only enhances sensitivity to early fault signatures but also ensures robustness across varying load levels and operating states, making it an effective approach for real-time condition monitoring of the head sheave system.

3.3.3. Acoustic Signal Analysis

Acoustic signals serve as a valuable complementary modality to vibration data, particularly for detecting early-stage and subtle structural defects in the head sheave device. While vibration analysis excels at capturing low-frequency mechanical resonances and shock responses, acoustic emissions are more sensitive to high-frequency transient phenomena such as crack initiation, material friction, and stress-wave propagation. These acoustic patterns, typically imperceptible to human hearing and easily masked by environmental noise in mining environments, require specialized processing to extract fault-relevant features. The field acoustic sensor measurement points are shown in Figure 5.



Figure 5. Measuring point of acoustic sensor.

In this work, raw acoustic signals acquired using directional condenser microphones with a flat frequency response from 20 Hz to 20 kHz are preprocessed and transformed into compact, high-resolution spectral features using the Mel-Frequency Cepstral Coefficient (MFCC) technique. MFCCs are particularly suitable for non-stationary acoustic signal processing due to their perceptual scaling, which emphasizes frequency bands where fault-induced emissions are most prominent. The MFCC extraction pipeline consists of pre-emphasis filtering to amplify high-frequency components, followed by frame segmentation (25 ms frames with 10 ms overlap) and Hamming windowing to reduce edge discontinuities. Each frame undergoes Fast Fourier Transform (FFT), after which a 40-channel Mel-scale filter bank is applied to approximate the human auditory response. The logarithmic filter bank energies are then converted to decorrelated coefficients using Discrete Cosine Transform (DCT), typically resulting in a 13-dimensional feature vector per frame.

The extracted MFCC features are reshaped into two-dimensional matrices representing spectral evolution over time, effectively capturing the temporal dynamics of fault-induced acoustic variations. These feature maps are standardized via z-score normalization to ensure amplitude consistency across different operating sessions and sensor instances. Additionally, augmentation techniques such as time-warping and pitch shifting are employed to improve model generalization and address class imbalance within the acoustic dataset.

By transforming acoustic waveforms into MFCC representations, the diagnostic framework gains enhanced sensitivity to transient, non-periodic acoustic events associated with incipient failures. These features are subsequently fused with vibration-derived representations at the feature level, providing the hybrid CNN-RF diagnostic model with a multimodal input that enhances fault classification accuracy. The integration of acoustic signal analysis thereby strengthens the system's ability to detect faults under complex and variable operational scenarios where vibration signals alone may be insufficient.

3.3.4. Multi-Source Fusion and Augmentation

To enhance the diagnostic performance and generalization capability of the proposed fault recognition system, a multi-source data fusion strategy is employed by integrating both vibration and acoustic features into a unified feature representation. This fusion approach leverages the complementary characteristics of each modality: vibration signals are adept at capturing periodic impacts and resonance behaviors, whereas acoustic emissions are more sensitive to high-frequency transient events caused by crack propagation or frictional interactions. The fusion process is executed at the feature level, enabling joint learning of multimodal information without introducing redundancy or misalignment.

The preprocessed vibration signals are transformed via S-Transform into two-dimensional scalograms and the MFCC-based acoustic representations are synchronized using a common time base ensured during data acquisition. These representations are concatenated channel-wise to form a fused feature tensor of shape $N \times C \times T$, where N denotes the number of samples, C the combined channel dimensions of both modalities, and T the time–frequency length. This composite input preserves spatial and spectral integrity and serves as input to the CNN-RF diagnostic architecture. As illustrated in Figure 6, the proposed architecture combines deep convolutional feature extraction from fused vibration and acoustic signals with an RF classifier, enabling accurate fault classification through ensemble learning.

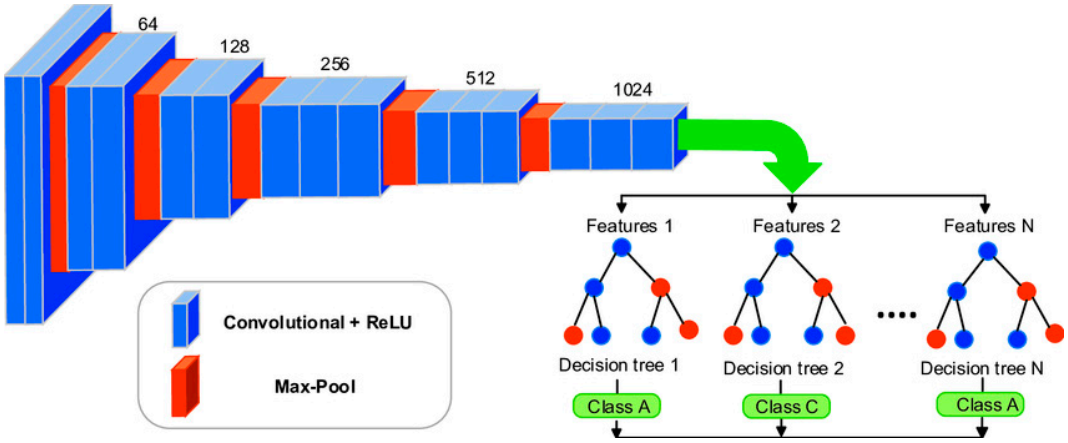


Figure 6. CNN-RF fault diagnosis architecture integrating multimodal feature extraction and ensemble decision trees [42].

To further improve model robustness and prevent overfitting, a series of data augmentation techniques are applied prior to training. In the time domain, Gaussian noise injection is used to simulate sensor-level variability and measurement uncertainty. Random time shifts and scaling transformations are employed to account for operational fluctuations and misalignment in signal timing. In the spectral domain, mixup augmentation blends two samples from different classes in a convex combination, which improves decision boundary smoothness and reduces class overfitting. These augmentations are applied uniformly across both vibration and acoustic channels to preserve alignment in the fused input.

Moreover, statistical techniques such as Principal Component Analysis (PCA) are used during exploratory analysis to verify the separability and redundancy of fused features, ensuring that the multimodal representation retains high discriminative power while maintaining computational efficiency. The final fused and augmented dataset exhibits improved coverage of operational conditions, variability in fault manifestations, and resilience to noise, thereby enhancing the performance and generalization capability of the hybrid CNN-RF model in fault classification tasks.

3.4. Experimental Validation

This paper designed and constructed a failure testing platform for the wheel hub assembly, as shown in Figure 7. By simulating failures in the wheel body, axle, and bearings, the feasibility of the monitoring system was validated.

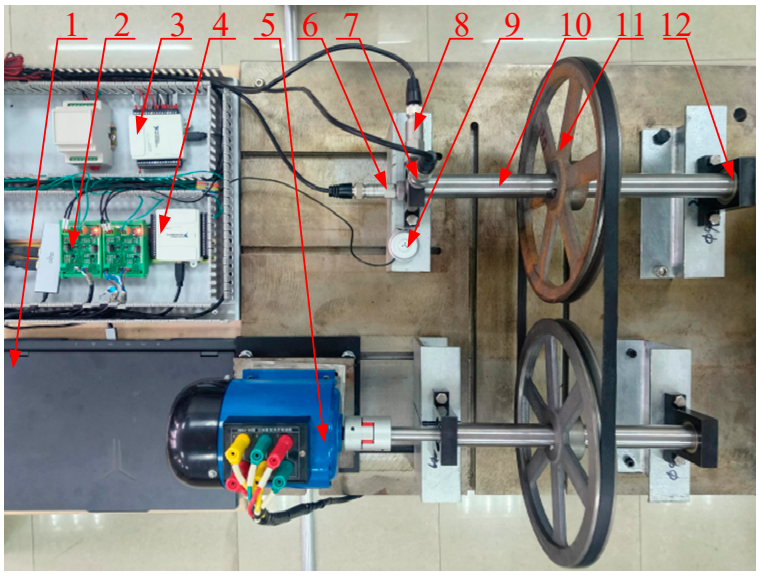


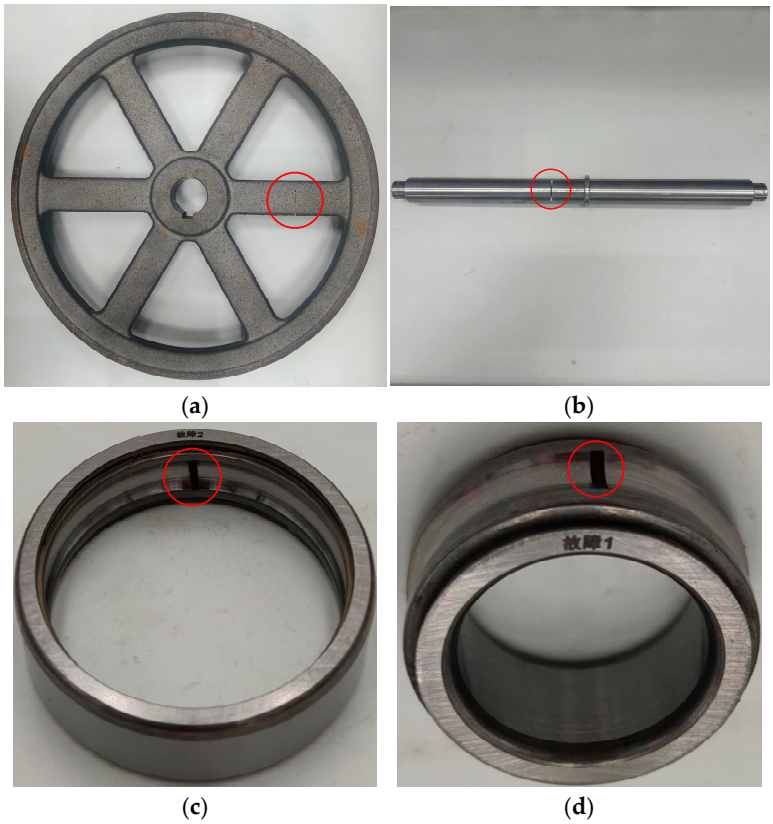
Figure 7. Experiment platform for head sheave device failure: 1. Computer, 2. Signal conditioning module, 3. Vibration signal acquisition card, 4. Acoustic signal acquisition card, 5. Motor, 6. Axial vibration sensor, 7. Vertical vibration sensor, 8. Radial vibration sensor, 9. Acoustic sensor, 10. Rotor shaft, 11. Wheel hub, 12. Bearing.

Vibration sensors were mounted on the bearing housing surface using magnetic bases, while the acoustic sensor was positioned adjacent to the bearing housing. The vibration and acoustic signals were first conditioned by a signal conditioning circuit and then transmitted to the data acquisition card for sampling. The sampled data were subsequently transferred to a computer via a serial interface for storage and analysis. To reproduce the fault states and representative working conditions of the hoisting head sheave system in a controlled and repeatable manner, this study employs a scaled-down experimental platform consisting of a wheel body, rotating shaft, and a 6204 deep-groove ball bearing. The 6204 bearing was selected to facilitate controlled fault reproduction and repeatable data collection on the test rig, rather than to match the full-scale bearing size of an industrial head sheave. The main parameters of the 6204 bearing are listed in Table 1.

Four types of faults were artificially induced in the test specimen: wheel body failure, shaft failure, outer bearing ring failure, and inner bearing ring failure, as shown in Figure 8.

Table 1. Parameters of bearing 6204.

Technical Specifications	Parameters
Bearing Model	6204
Inner Diameter (mm)	20
Outer Diameter (mm)	47
Pitch Diameter (mm)	33.5
Number of Rolling Elements (pcs)	8
Ball Diameter (mm)	7.94
Contact Angle (degrees)	0



(**a**) (**b**) (**c**) (**d**)

Figure 8. Fault settings for the experiment object: (**a**) wheel body failure, (**b**) faulty hinge, (**c**) outer ring failure, and (**d**) inner ring failure.

During the experiment, determining the rotational speed of the sheave assembly alone suffices to obtain the fault characteristic frequencies of the sheave body and shaft. Substituting the bearing parameters and rotational frequency into Equations (9) and (10) then calculates the fault characteristic frequencies of the bearing outer and inner rings. The values of these characteristic frequencies depend on the bearing's geometric parameters and operating speed [43], with the calculation relationship as follows:

Bearing inner ring rotational frequency f_r:

$$f_r = \frac{n}{60} \tag{9}$$

Bearing outer ring failure frequency f_o:

$$f_o = \frac{Z}{2} f_r \left(1 - \frac{d}{D} cos\alpha\right) \tag{10}$$

The rotational speed of the inner ring of the bearing is expressed in the unit RPM: this indicates the number of rolling bodies; represents the diameter of the rolling element in mm; represents the diameter of the node circle, in mm; and indicates the contact angle.

Subsequently, comparing the extracted characteristic frequencies from signal analysis with the theoretically calculated values enables precise diagnosis of the specific component failure within the sheave assembly. After completing the assembly of the sheave assembly failure test platform, failure simulation experiments for the sheave assembly can be conducted. The experimental procedure is as follows: (1) After assembling components under normal operating conditions on the test platform, set the sampling rates for vibration signals and acoustic signals to 10,000 Hz and 22,050 Hz, respectively. Adjust the motor speed to 600 RPM. Once the motor speed stabilizes, collect the vibration and acoustic data at this point. (2) Remove the components from normal operating conditions and sequentially replace them with the faulty wheel body, faulty shaft, faulty outer bearing ring, and faulty inner bearing ring. Repeat the experiment following the same procedure. (3) Analyze and process the vibration and sound signals from the normal operating condition and the conditions with wheel body failure, shaft failure, outer bearing ring failure, and inner bearing ring failure.

3.4.1. Modal Excitation Tests

To verify the accuracy and physical relevance of the finite element model and to characterize the dynamic behavior of the head sheave structure under realistic boundary conditions, a series of modal excitation tests were conducted. These experimental modal tests served two key purposes: (1) to extract the natural frequencies and mode shapes of the physical structure for comparison with simulation results, and (2) to provide a vibration baseline for distinguishing structural degradation due to faults.

The tests were carried out on a full-scale head sheave assembly mounted on a steel support frame designed to replicate in situ constraints. A modal hammer excitation method was employed using an instrumented force hammer with a built-in piezoelectric load cell to apply controlled broadband impacts to the wheel rim, web, and bearing seat locations. Corresponding response signals were recorded through triaxial accelerometers strategically placed on the sheave body, especially near regions of high modal curvature identified in simulation. The excitation and response signals were sampled at 10 kHz using a multi-channel data acquisition system with time-synchronized triggering to ensure phase coherence.

Frequency response functions (FRFs) were computed using a Hanning window and an average of multiple impact repetitions to improve spectral clarity. The resulting FRFs were processed to extract modal parameters including natural frequencies, damping ratios, and normalized mode shapes via a peak-picking method and confirmed using curve fitting techniques.

Comparison with the simulation results from the ANSYS modal analysis revealed a strong agreement. The first five modal frequencies deviated by less than 3.5% between experiment and simulation, validating the fidelity of the finite element model. Notably, Mode 2 (first lateral bending) and Mode 4 (out-of-plane torsional mode) showed the most sensitivity to faults in the bearing and web regions, consistent with the location of induced cracks and spalls. This alignment between physical and simulated modal behavior substantiates the use of simulated data in generating training inputs for the fault diagnosis model.

Furthermore, these modal tests provide empirical evidence of frequency shifts and damping changes that correlate with defect progression. As such, the modal excitation

results not only validate the structural model but also reinforce the suitability of vibration-based fault indicators extracted in earlier stages of the diagnostic pipeline.

3.4.2. Fault Simulation Platform

To generate realistic and labeled fault data under controlled conditions, a dedicated fault simulation platform was developed to replicate the operational dynamics of a mine hoist head sheave device. This platform enables the emulation of common mechanical failures, such as bearing spalling, inner race cracking, and shaft misalignment, under varying load and rotational speed conditions. The testbed was designed to reflect the structural and boundary characteristics observed in actual mining hoisting systems, ensuring high mechanical fidelity and relevance for model training and validation.

The simulation platform consists of a variable-speed motor-driven shaft assembly connected to a fabricated head sheave structure mounted on adjustable support bearings. The shaft and sheave are coupled via a flexible coupler to introduce torsional effects. The bearing housings are instrumented with triaxial accelerometers and directional microphones to capture vibration and acoustic signals, respectively. The data acquisition system is configured for synchronized multi-channel sampling at a rate of 10 kHz, with analog signal conditioning to maintain signal integrity. To simulate distinct fault types, controlled damage was introduced into the bearing and shaft assemblies: (a) Outer ring pitting was artificially created via electric discharge machining; (b) inner race cracks were initiated using notch fatigue under cyclic loading; and (c) shaft eccentricity was generated by intentional misalignment during installation.

Each fault condition was validated visually and with NDT (non-destructive testing) prior to data collection. The system was operated under variable speeds (150–450 RPM) and load conditions to simulate real-world variability. For each condition, multiple operational runs were recorded to capture repeatable and stable signal characteristics. Each recording was time-stamped and manually labeled according to the fault type and severity for supervised learning purposes.

The platform not only facilitated the generation of high-quality datasets for training the CNN-RF diagnostic model but also allowed for controlled experimentation on the influence of fault location, load level, and rotational speed on the signal features. By combining fault emulation with synchronized multi-sensor acquisition, the platform serves as a reliable foundation for model development, performance benchmarking, and comparative evaluation across diagnostic architectures.

3.4.3. Validation Metrics

To quantitatively evaluate the diagnostic performance of the proposed CNN-RF fault classification model, a comprehensive set of statistical validation metrics was employed. These metrics are designed to assess not only the overall accuracy of the model but also its sensitivity to minority class detection, its robustness to class imbalance, and its reliability in practical deployment scenarios. The following key metrics were used:

Accuracy (ACC): Measures the proportion of correctly classified instances over the total number of samples. It provides a general indicator of model effectiveness but may be less informative in imbalanced datasets.

Precision (P): Defined as the ratio of true positives (TPs) to the sum of true positives and false positives (FPs). Precision reflects the model's ability to avoid false alarms, particularly relevant in safety-critical applications such as mine hoist monitoring.

Recall (R) or Sensitivity: The ratio of true positives to the sum of true positives and false negatives (FNs). It evaluates the model's capacity to detect actual fault conditions, which is critical for early-stage fault identification.

F1 Score: The harmonic mean of precision and recall. F1 balances the trade-off between false alarms and missed detections, and is particularly valuable in multi-class fault scenarios where class distributions vary.

Confusion Matrix: A multi-class matrix summarizing the true and predicted labels for each fault type, allowing visual assessment of misclassification patterns, class confusion, and dominant fault detection pathways.

Receiver Operating Characteristic (ROC) Curve and AUC (Area Under Curve): While primarily applicable to binary classification tasks, ROC-AUC values were computed on a one-vs-rest basis for each fault class to evaluate discrimination capability across decision thresholds.

Training and Validation Loss Curves: Tracked during model training to assess convergence behavior, generalization gap, and overfitting tendencies.

All metrics were computed over stratified cross-validation folds to ensure statistical stability. The CNN-RF model achieved a mean classification accuracy exceeding 96%, with F1 scores above 0.93 for all critical fault classes. These results confirm the model's effectiveness in identifying both localized and distributed fault modes under varying operating conditions, validating the overall robustness and reliability of the proposed diagnostic system.

3.5. Fault Diagnosis: Multi-Source Fusion CNN-RF Model

To achieve high-accuracy fault identification under complex working conditions, a hybrid fault diagnosis framework based on a Convolutional Neural Network and Random Forest (CNN-RF) is developed, leveraging multi-source fused features from both vibration and acoustic modalities. This architecture combines the deep feature extraction capabilities of CNNs with the classification stability and interpretability of ensemble learning through RFs, thereby addressing both the nonlinear complexity of signal features and the demand for reliable decision-making in critical safety systems.

The input to the model consists of fused two-dimensional representations: vibration scalograms derived from the S-Transform and acoustic MFCC feature maps. These are concatenated along the channel dimension after normalization and temporal alignment, producing a unified feature tensor that preserves both spectral and temporal information. The CNN component comprises multiple convolutional layers with ReLU activation, interleaved with MaxPooling layers to progressively extract spatial hierarchies and reduce dimensionality. Batch normalization is applied to accelerate convergence and improve generalization. The final convolutional output is flattened and passed through fully connected layers to generate high-level latent features.

Unlike conventional end-to-end CNN classification models, the proposed framework decouples feature extraction and classification. The deep features produced by the CNN are fed into a RF classifier, trained separately to improve robustness against overfitting and enhance fault class separability. This ensemble classifier constructs multiple decision trees based on random feature subsets and aggregates their outputs through majority voting, thereby mitigating noise sensitivity and capturing nonlinear class boundaries effectively.

To interpret the contribution of different feature types and ensure transparency, the model integrates SHAP (SHapley Additive exPlanations) values to quantify feature importance at the output stage. The SHAP analysis confirms that fault-relevant frequency bands in both vibration and acoustic domains are among the top contributors to classification decisions, validating the effectiveness of multi-source fusion. Recent industrial-process studies have shown that SHAP can effectively reveal physically meaningful relationships in black-box models, improving transparency and supporting deployment in safety-critical engineering applications [44,45].

As illustrated in Figure 9, the architecture adopts a dual-channel feature extraction pathway where vibration and acoustic signals are independently processed via parallel CNN streams and a unified ensemble classification stage.

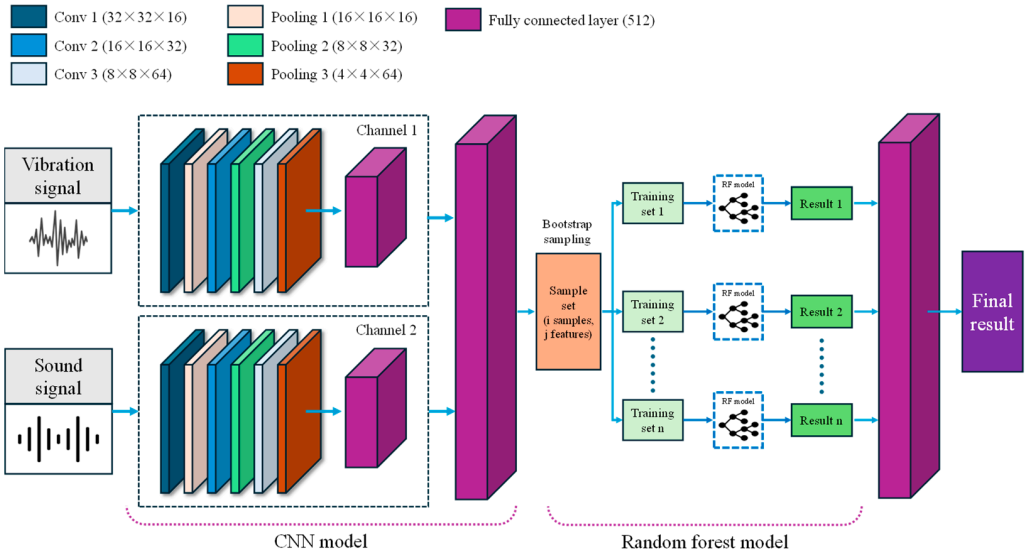


Figure 9. Block diagram of the multi-source CNN-RF fault diagnosis model. Parallel convolutional branches extract features from vibration and acoustic signals, which are fused and classified using an RF ensemble.

The diagram highlights the decoupling of feature learning and decision-making processes, as well as the modularity and scalability of the proposed framework. This separation enables easy retraining or adaptation to new sensor modalities or fault types with minimal architectural changes. The CNN-RF model achieves superior performance compared to baseline classifiers, with classification accuracy exceeding 96% across five distinct fault categories. This model demonstrates strong performance on the tested dataset and provides interpretable decision support via SHAP analysis. This study focuses on validating an integrated, deployment-oriented diagnostic pipeline under the tested operating condition; cross-speed and cross-load generalization will be investigated in future work using held-out operating conditions. Moreover, the modular nature of the model allows flexible retraining or adaptation when new fault types or sensors are introduced, making it suitable for deployment in real-time monitoring systems for hoisting equipment in mining applications.

3.5.1. Data Preparation

Vibration and acoustic signals were acquired from the head sheave device fault test platform under five operating conditions: normal state, wheel body fault, shaft fault, outer ring fault, and inner ring fault. The vibration signal was sampled at 10,000 Hz, and the acoustic signal was sampled at 22,050 Hz. For both modalities, training samples are generated from each continuous recording using a sliding-window segmentation (window length 2048 points, step size 512 points, i.e., 75% overlap), corresponding to a segment duration of approximately 0.205 s (vibration, 10 kHz) and 0.093 s (acoustic, 22.05 kHz). For each operating condition, 1200 samples were collected, resulting in a total of 6000 vibration samples and 6000 acoustic samples across the five categories. The dataset was then divided into a training set (70%) and a test set (30%).

It should be noted that the supervised learning experiments reported in this study (i.e., CNN-RF training and testing) are conducted only using the vibration and acoustic data collected from the physical fault test platform. The simulation results generated via ANSYS and ADAMS are used exclusively for fault mechanism analysis and physical verification

(e.g., validating dynamic responses and characteristic signatures against theoretical and experimental observations), and are not included as training samples for the CNN-RF model. Table 2 details the experimental dataset information, with samples labeled from 1 to 5 according to their respective operating conditions. Figure 10 presents the generalized S-Transform of vibration and MFCC heatmaps of sound for the five operating conditions on the sheave assembly failure test platform.

For each operating condition, the vibration/acoustic signals were recorded as continuous time series (10 s per recording, repeated 5 times). Training samples were generated using a sliding-window segmentation strategy with window length 2048 points and step size 512 points (75% overlap). Therefore, the model inputs are overlapped segments extracted from continuous recordings rather than independent 1 s acquisitions. The segment duration is determined by the sampling rate and window length (2048 points corresponds to approximately 0.16 s at the specified sampling rate).

Because training samples are generated from continuous recordings using overlapping windows, we perform train/test splitting strictly at the recording-run level (group-wise). All segments extracted from the same recording run/session are assigned exclusively to either training/validation or testing, and no run contributes samples to both sets. We repeat the run-wise evaluation $N = 5$ times using different run-level partitions and report results as mean $\pm$ standard deviation across repeats.

3.5.2. CNN Feature Extractor

In the proposed CNN-RF framework, the Convolutional Neural Network (CNN) component serves as a deep feature extractor, automatically learning hierarchical representations from complex time–frequency inputs derived from vibration and acoustic signals. This component is essential for capturing both local and global spatial patterns in the input data, which are often difficult to define through manual feature engineering, especially in the presence of non-stationary and nonlinear characteristics commonly found in mechanical fault signals. Figure 9 illustrates the proposed multi-source fusion fault diagnosis model, with specific parameter settings shown in Table 3.

Table 2. Experimental dataset.

Data Type	Label	Number of Samples	Sampling Frequency (Hz)	Segment Length (Points)
Normal Operating Condition Vibration Signal	1	1200	10,000	2048
Wheel Body Fault Vibration Signal	2	1200	10,000	2048
Shaft Fault Vibration Signal	3	1200	10,000	2048
Outer Ring Fault Vibration Signal	4	1200	10,000	2048
Inner Ring Fault Vibration Signal	5	1200	10,000	2048
Normal Operating Condition Sound Signal	1	1200	22,050	2048
Wheel Body Fault Sound Signal	2	1200	22,050	2048
Shaft Fault Sound Signal	3	1200	22,050	2048
Outer Ring Fault Sound Signal	4	1200	22,050	2048
Inner Ring Fault Sound Signal	5	1200	22,050	2048

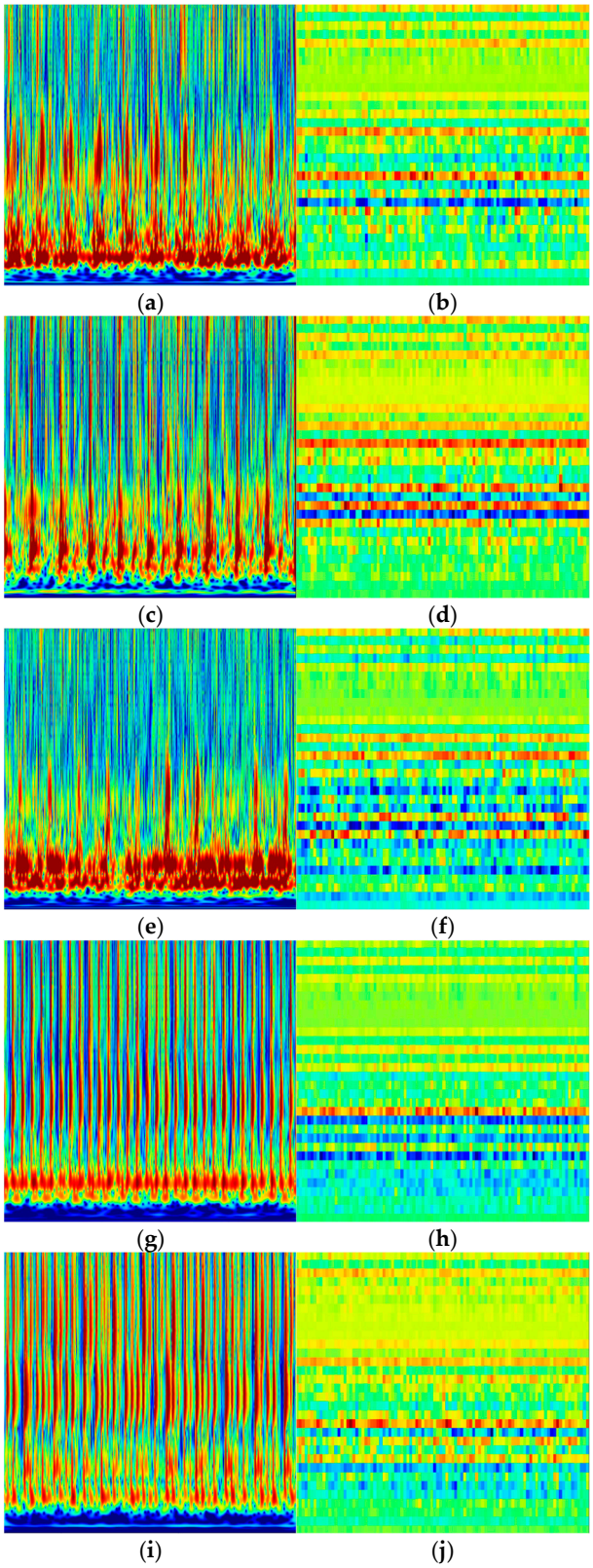


Figure 10. Generalized S-Transform and MFCC thermodynamic diagram under various operating conditions: (**a**) Normal working conditions GST, (**b**) normal working conditions MFCC, (**c**) wheel body failure GST, (**d**) wheel body failure MFCC, (**e**) faulty hinge GST, (**f**) faulty hinge MFCC, (**g**) outer ring failure GST, (**h**) outer ring failure MFCC, (**i**) inner ring failure GST, (**j**) inner ring failure MFCC. Table 2 summarizes the number of samples, labels, sampling frequencies, and durations for vibration and sound signals under various operating and fault conditions.

Table 3. Parameter list of models.

Network Layer	Type	Parameters	Input	Output	Activation	Step
Channel 1—Conv1	Convolution	filters = 16, kernel = 3 × 3, stride = 1	(32, 32, 1)	(32, 32, 16)	ReLU	1
Channel 1—Pool1	MaxPooling	pool_size = 2 × 2, stride = 2	(32, 32, 16)	(16, 16, 16)	-	
Channel 1—Conv2	Convolution	filters = 32, kernel = 3 × 3, stride = 1	(16, 16, 16)	(16, 16, 32)	ReLU	1
Channel 1—Pool2	MaxPooling	pool_size = 2 × 2, stride = 2	(16, 16, 32)	(8, 8, 32)	-	2
Channel 1—Conv3	Convolution	filters = 64, kernel = 3 × 3, stride = 1	(8, 8, 32)	(8, 8, 64)	ReLU	1
Channel 1—Pool3	MaxPooling	pool_size = 2 × 2, stride = 2	(8, 8, 64)	(4, 4, 64)	-	2
Channel 1—Flatten	Flatten	-	(4, 4, 64)	(1024,)	-	
Channel 1—Dense	Fully Connected	units = 512	(1024,)	(512,)	ReLU	
Channel 2—Conv1	Convolution	filters = 16, kernel = 3 × 3, stride = 1	(32, 32, 1)	(32, 32, 16)	ReLU	1
Channel 2—Pool1	MaxPooling	pool_size = 2 × 2, stride = 2	(32, 32, 16)	(16, 16, 16)	-	2
Channel 2—Conv2	Convolution	filters = 32, kernel = 3 × 3, stride = 1	(16, 16, 16)	(16, 16, 32)	ReLU	1
Channel 2—Pool2	MaxPooling	pool_size = 2 × 2, stride = 2	(16, 16, 32)	(8, 8, 32)	-	2
Channel 2—Conv3	Convolution	filters = 64, kernel = 3 × 3, stride = 1	(8, 8, 32)	(8, 8, 64)	ReLU	1
Channel 2—Pool3	MaxPooling	pool_size = 2 × 2, stride = 2	(8, 8, 64)	(4, 4, 64)	-	2
Channel 2—Flatten	Flatten	-	(4, 4, 64)	(1024,)	-	
Channel 2—Dense	Fully Connected	units = 512	(1024,)	(512,)	ReLU	
Concatenate	Concatenation	axis = 1	(512,) + (512,)	(1024,)	-	
Dense (Fusion)	Fully Connected	units = 128	(1024,)	(128,)	ReLU	
Dropout	Dropout	rate = 0.5	(128,)	(128,)	-	
RF Classifier	Random Forest	n_estimators = 200, max_depth = 20	(128,)	(5,)	Softmax	

The CNN architecture is designed to process two-dimensional input matrices: S-Transform scalograms from vibration signals and MFCC feature maps from acoustic signals. Each input passes through a dedicated CNN branch composed of multiple layers of convolution operations followed by nonlinear activation functions (ReLU), batch normalization, and MaxPooling layers. The convolution layers apply multiple filters (kernels) to detect edges, ridges, and fault-relevant textures in both time and frequency directions, while MaxPooling layers progressively reduce the spatial dimensions to improve computational efficiency and ensure translation invariance.

The CNN structure used in this study includes three convolutional-pooling blocks, followed by a flattening layer that transforms the final feature maps into a one-dimensional vector. This vector is passed through fully connected (dense) layers that perform high-level feature abstraction, enabling the network to learn discriminative representations that are robust to noise, load variation, and signal amplitude scaling. Dropout regularization is applied during training to prevent overfitting and enhance model generalization.

To integrate the two sensor modalities, the feature vectors from the vibration and acoustic branches are concatenated after their respective CNN paths, forming a unified feature embedding. This fused vector captures both structural and acoustic behavior of the monitored system, providing a comprehensive basis for fault diagnosis.

The CNN feature extractor is trained using cross-entropy loss and optimized with the Adam optimizer, employing a mini-batch gradient descent approach with learning rate scheduling and early stopping based on validation loss. The training process is guided by augmented data and evaluated through accuracy, precision, and F1-score metrics to ensure robust feature learning.

3.5.3. Model Integration and Evaluation

Upon completion of the deep feature extraction via parallel CNN branches for vibration and acoustic signals, the resulting high-dimensional feature vectors are concatenated to form a unified representation. This fused feature vector embodies both temporal-frequency structural patterns and high-frequency acoustic characteristics, providing a comprehensive descriptor of the head sheave system's operational condition.

The integrated feature set is subsequently fed into a RF classifier, which serves as the decision-making component of the hybrid CNN-RF architecture. RF, an ensemble learning method, constructs multiple decision trees using bootstrapped subsets of the training data and a randomized selection of features at each split. The final classification decision is determined via majority voting, which enhances model stability and reduces sensitivity to noise and overfitting. This model integration approach effectively decouples feature learning from classification, improving interpretability and allowing for independent tuning of each component.

During training, the CNN is optimized using labeled samples via backpropagation and cross-entropy loss minimization, while the RF is trained using the CNN-extracted features with predefined fault class labels. For reproducibility, the complete hyperparameter configuration for CNN training (optimizer, learning rate, batch size, epochs, validation split, early stopping, and regularization) is summarized in Table 4. Table 4 also reports the RF settings used in the final model (e.g., number of trees, maximum depth, splitting criterion, and random seed). The entire training process is monitored through validation metrics, including loss convergence curves and confusion matrix analysis, to ensure model robustness and prevent overfitting.

Model evaluation is performed on a held-out test dataset, comprising samples not seen during training or validation phases. As discussed earlier, classification performance is assessed using accuracy, precision, recall, F1-score, and a confusion matrix. The hybrid model demonstrates excellent diagnostic performance across five fault categories, achieving an overall classification accuracy above 96%, with high recall and F1-scores for each individual fault type. To enhance explainability and validate the decision logic, SHAP (SHapley Additive exPlanations) values are computed for the RF output. These values quantify the contribution of individual features (e.g., frequency components, MFCC bands) to the model's predictions, providing insights into the discriminative power of specific vibration and acoustic attributes.

The final integrated model proves to be not only accurate but also scalable, interpretable, and robust to variations in operational conditions. Its modular structure enables easy retraining or fine-tuning when extended to new fault types or sensor configurations, making it well-suited for real-world deployment in condition monitoring systems of mining hoisting equipment.

Table 4. Hyperparameter configuration for CNN-RF model.

Category	Parameter	Value
RF Classifier	Number of trees (n_estimators)	200
	Maximum depth (max_depth)	20
	Minimum samples split	5
	Minimum samples leaf	2
	Maximum features	sqrt(128) ≈ 11
	Bootstrap	True
	Out-of-bag score	True
	Random state	42
	Class weight	Balanced
	Criterion	Gini impurity
	n_jobs	−1
Optimization	Optimizer	Adam
	Initial learning rate	0.001
	Minimum learning rate	1×10^{-6}
	Beta1 (Adam)	0.9
	Beta2 (Adam)	0.999
	Epsilon (Adam)	1×10^{-7}
Training	Batch size	32
	Number of epochs	50
	Validation split	0.2
	Shuffle	True
	Early stopping patience	10 epochs
Regularization	Dropout rate	0.5
	Weight decay (L2)	1×10^{-4}

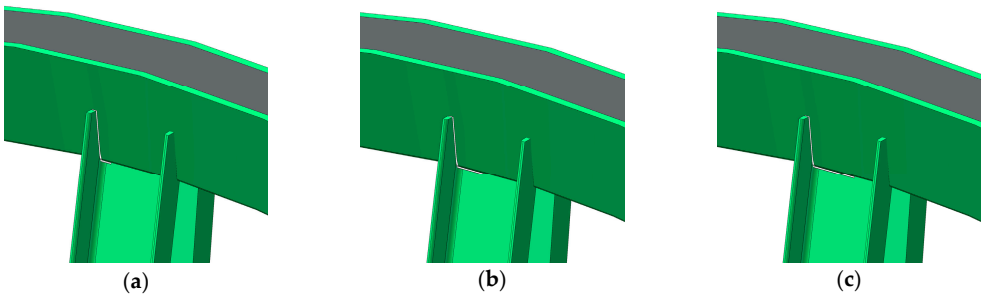
4. Results and Analysis

4.1. Simulation Results: Modal Analysis and Rigid–Flexible Coupling

To accurately characterize the dynamic behavior of the head sheave system and validate its structural resonance response, both modal analysis and rigid–flexible coupling simulations were conducted. These simulations provide critical insight into natural frequency distributions, vibration mode shapes, and the structural sensitivity to induced faults under operational loads.

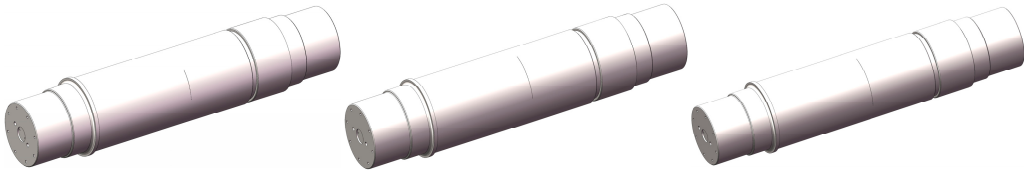
4.1.1. Modal Frequencies and Mode Shapes

Based on field failure experience, simulate failures in the wheel body and axle shaft by creating distinct damage patterns at the welded junctions between the wheel spokes and flange, as well as at the midpoint of the axle shaft. Relative depth of injury is represented by β, that is $\beta = l/L$, for which l represents the depth of damage and L represents the maximum depth of the structure at the place of injury. The specific working conditions are set in Tables 5 and 6, and there are four working conditions, namely healthy state, mild injury, moderate injury and severe injury, as shown in Figures 11 and 12.



(**a**) (**b**) (**c**)

Figure 11. Damage degree of wheel body: (**a**) Operating Condition 2, (**b**) Operating Condition 3, (**c**) Operating Condition 4.



(**a**) (**b**) (**c**)

Figure 12. Damage degree of rotating shaft: (**a**) Operating Condition 2 (**b**) Operating Condition 3 (**c**) Operating Condition 4.

Table 5. Operating condition setting of wheel body.

Operating Condition	Damage Width (mm)	Damage Depth (mm)	Relative Damage Depth
1	0	0	0
2	1	30	0.1875
3	1	60	0.375
4	1	90	0.5625

Table 6. Operating condition setting of rotating shaft.

Operating Condition	Damage Width (mm)	Damage Depth (mm)	Relative Damage Depth
1	0	0	0
2	0.5	27.5	0.05
3	0.5	55	0.1
4	0.5	82.5	0.15

Based on the constructed three-dimensional models of the wheel body and shaft, along with the operating condition settings, simulation analysis was performed for each operating condition model using ANSYS Workbench. This yielded the modal frequencies and modal shapes for each condition. Taking Operating Condition 1 as an example, the first 16 modal shapes of the wheel body and the first 6 modal shapes of the shaft are shown in Figures 13 and 14, respectively. On this basis, fault detection analysis was conducted for both the wheel body and shaft.

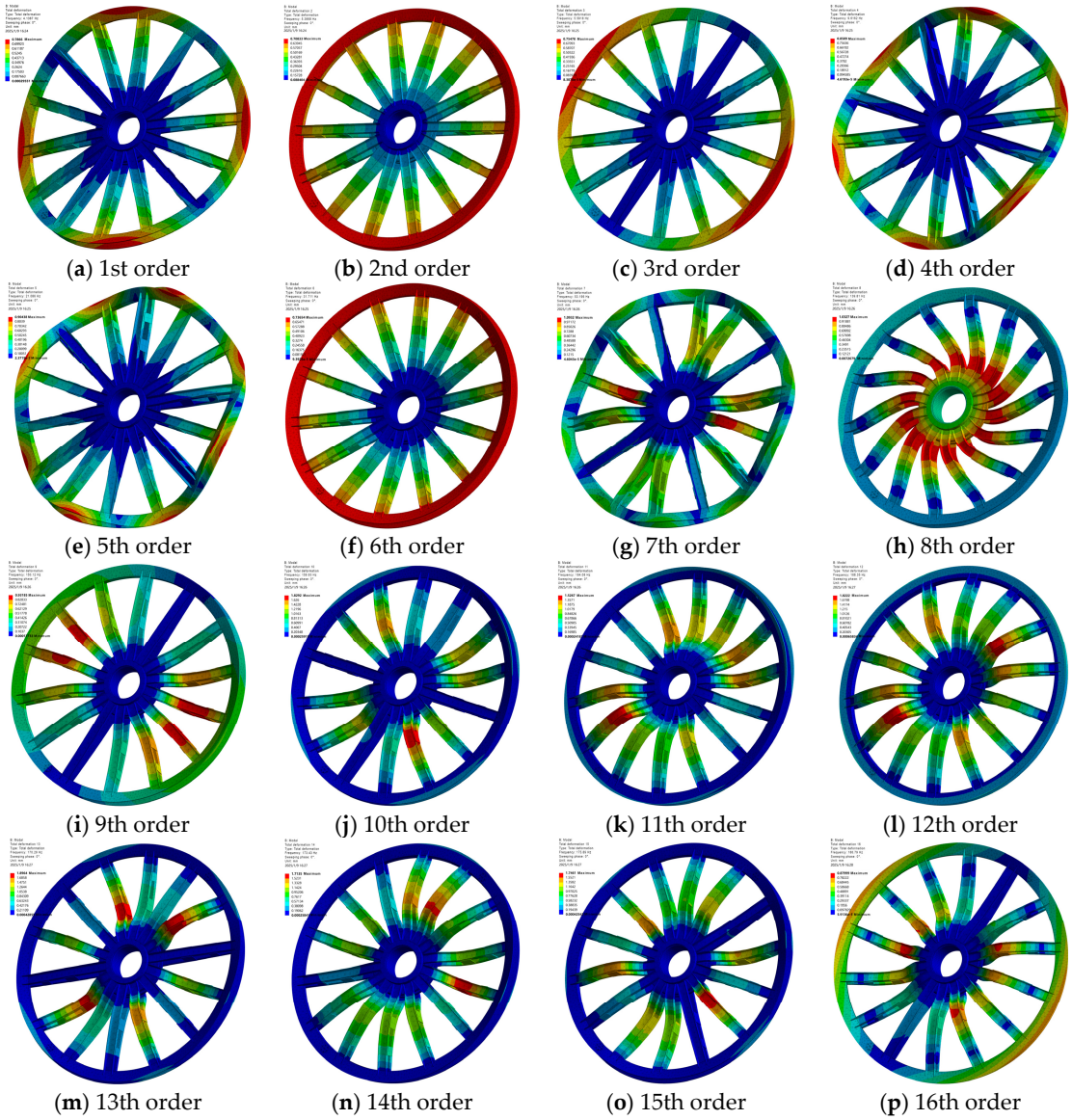
Tables 7 and 8 present the modal frequencies of the wheel body and rotating shaft under various operating conditions obtained through finite element simulation analysis.

Table 7. Modal analysis results of wheel body.

Modal Frequency Order	Condition-1 (Hz)	Condition-2 (Hz)	Condition-3 (Hz)	Condition-4 (Hz)
1st order	4.138	4.135	4.132	4.127
2nd order	5.366	5.427	5.486	5.591
3rd order	5.591	5.56	5.535	5.429
4th order	9.616	9.521	9.456	9.326
5th order	21.08	20.92	20.81	20.71
6th order	31.71	30.51	29.21	27.73
7th order	53.19	52.55	50.19	48.43
8th order	139.81	138.42	136.59	133.73
9th order	150.12	150.08	150.06	150.04
10th order	159.93	155.35	153.19	151.37
11th order	164.08	163.62	162.41	160.87
12th order	168.35	166.89	165.61	164.76
13th order	170.29	170.28	170.28	170.26
14th order	173.43	175.66	176.01	177.79
15th order	175.89	179.73	185.2	187.39
16th order	186.79	188.05	192.44	198.53

Table 8. Modal analysis results of rotating shaft.

Modal Frequency Order	Condition-1 (Hz)	Condition-2 (Hz)	Condition-3 (Hz)	Condition-4 (Hz)
1st order	0.0229	0.0259	0.0307	0.0318
2nd order	442.41	441.16	438.08	430.49
3rd order	446.92	446.4	445.06	443.36
4th order	839.58	838.56	834.76	827.12
5th order	1068.7	1068.6	1068.1	1066.7
6th order	1082	1081.9	1081.3	1080



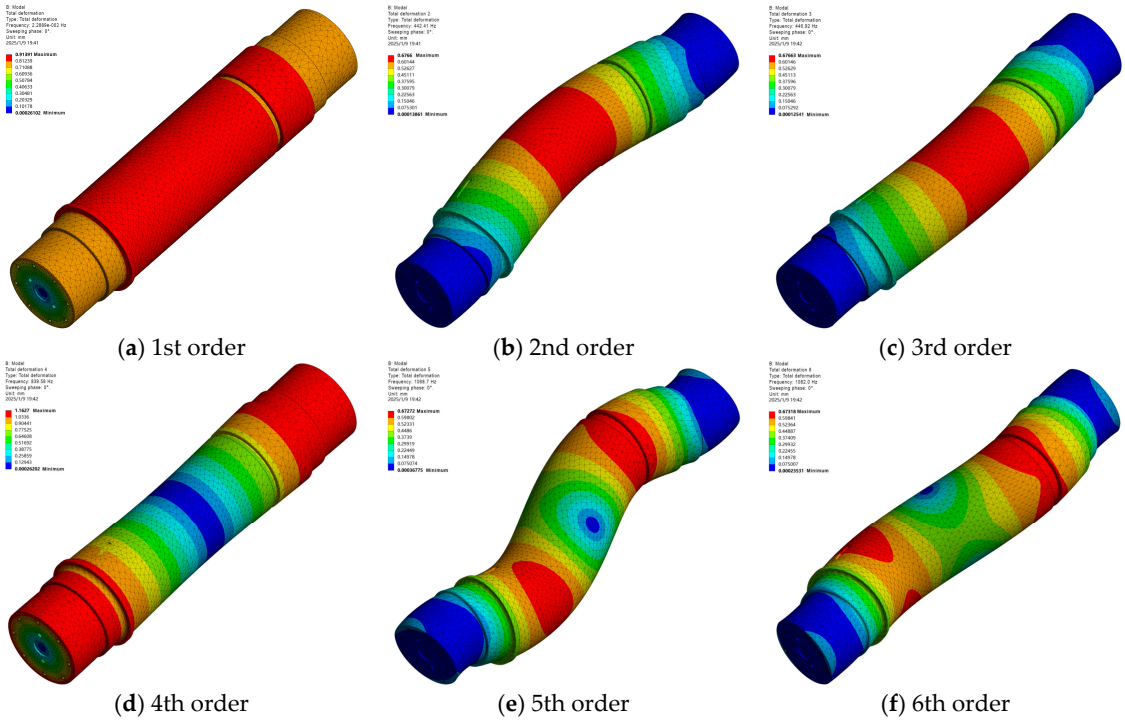
(**a**) 1st order (**b**) 2nd order (**c**) 3rd order (**d**) 4th order

(**e**) 5th order (**f**) 6th order (**g**) 7th order (**h**) 8th order

(**i**) 9th order (**j**) 10th order (**k**) 11th order (**l**) 12th order

(**m**) 13th order (**n**) 14th order (**o**) 15th order (**p**) 16th order

Figure 13. Modal frequency and vibration mode of the first 16 orders of wheel body.



(**a**) 1st order (**b**) 2nd order (**c**) 3rd order

(**d**) 4th order (**e**) 5th order (**f**) 6th order

Figure 14. Modal frequency and vibration mode of the first 6 orders of rotating shaft.

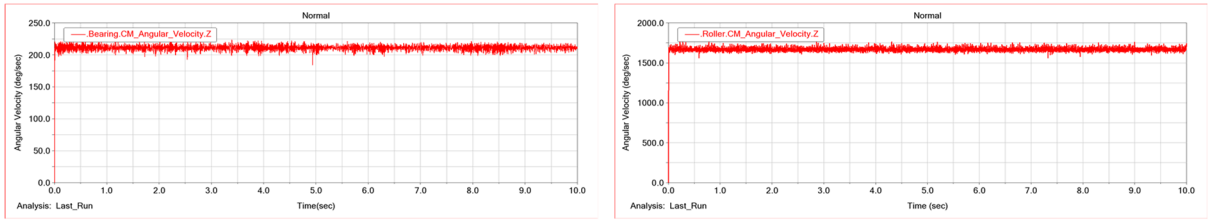
It can be observed that as the extent of structural damage increases, both the wheel body and rotating shaft exhibit varying degrees of modal frequency changes. With progressively deeper damage, the magnitude of modal frequency shifts also grows larger. Consequently, these changes can to some extent reflect the presence and severity of damage in the wheel body and rotating shaft. These results align with the conclusions derived from Equation (8). However, in practical scenarios, different excitation locations and methods can stimulate modal characteristics of varying orders.

4.1.2. Dynamic Response in ADAMS

To further understand the kinematic behavior and validate the mechanical integrity of the head sheave system during operation, multi-body dynamic simulations were conducted using the ADAMS environment. These simulations focused specifically on capturing the dynamic response of the bearing assembly, including the rotational behavior of inner and outer rings, the rolling elements, and the cage. The objective was to establish the theoretical validity of bearing kinematics under fault-free and faulty conditions, while generating synthetic data for signal analysis and fault modeling.

The bearing motion relationships are governed by rigid-body dynamics and classical bearing kinematics equations. When the hoisting system operates at a lifting speed of 8 m/s, the rotational speed of the head sheave is computed to be approximately 82.47 RPM. In the simulation, the outer ring is fixed (0 RPM), and the inner ring rotates at 82.47 RPM, consistent with the physical constraints of bearing mounting. From this, the theoretical speeds of the internal components can be derived: Cage rotational speed $\approx$ 35.45 RPM; rolling element rotational speed $\approx$ 278.07 RPM; outer ring defect frequency $\approx$ 8.27 Hz; inner ring defect frequency $\approx$ 10.97 Hz.

These values were validated in the ADAMS simulation, where the computed cage and roller angular velocities closely matched theoretical expectations (see Figure 15). The high agreement between calculated and simulated kinematic values confirms the fidelity of the dynamic model.



(a) (b)

Figure 15. (**a**) Speed of retainer; (**b**) rotating speed of roller.

In addition to angular velocity tracking, dynamic response simulations were conducted under three operational conditions: healthy bearing, outer ring fault, and inner ring fault. The time domain acceleration signals obtained from the simulated bearing center mass under these conditions are shown in Figure 16a–c. The healthy bearing condition exhibits relatively smooth vibrational behavior, while the faulted conditions clearly show amplitude modulation and periodic impulsive features indicative of defect interaction.

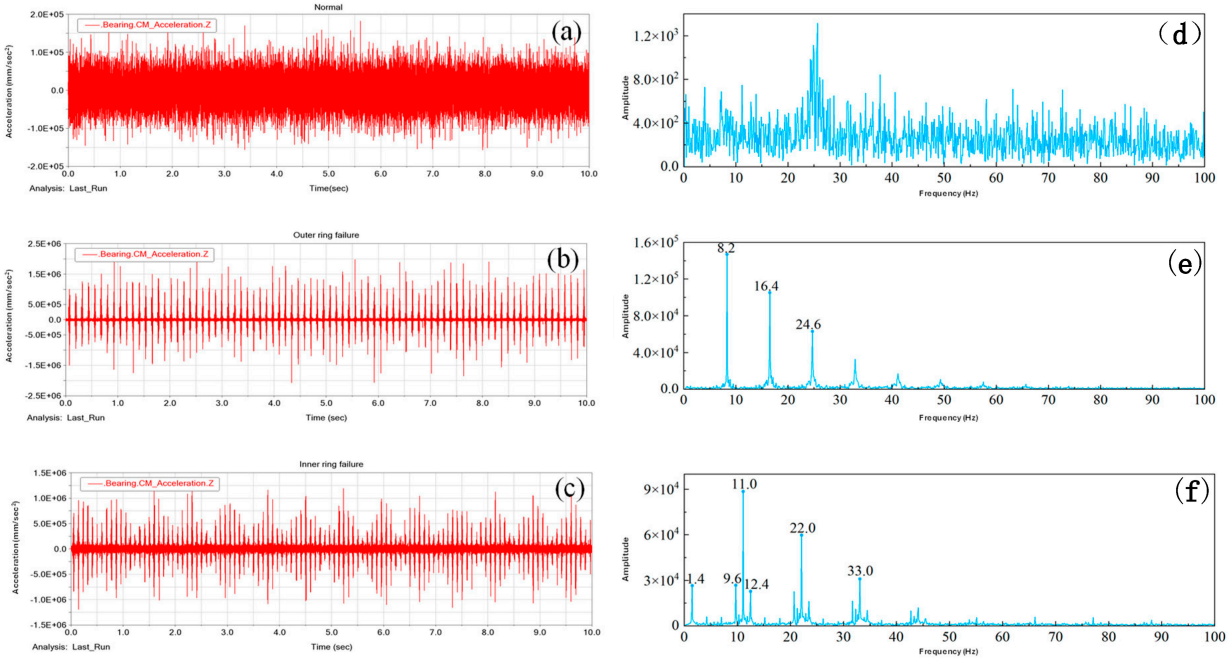


Figure 16. Time domain diagram of simulation data of bearing under various working conditions: (**a**) normal; (**b**) outer ring failure; (**c**) inner ring failure. Simulation data envelope spectrum of bearing under various working conditions: (**d**) normal bearing vibration signal envelope spectrum; (**e**) outer ring fault vibration signal envelope spectrum; (**f**) envelope spectrum of inner ring fault vibration signal).

To extract fault-relevant frequency components, the envelope analysis of the simulated signals was performed using the Hilbert transform. The resulting envelope spectra are presented in Figure 16d–f. The outer ring fault condition (Figure 16e) exhibits dominant peaks at 8.2 Hz, 16.4 Hz, and 24.6 Hz, corresponding to integer multiples of the calculated fault frequency. Similarly, the inner ring fault (Figure 16f) shows peaks at 11.0 Hz, 22.0 Hz, and 33.0 Hz, consistent with theoretical inner ring defect frequencies.

These results demonstrate that the ADAMS model reproduces fault-induced dynamic responses and characteristic-frequency signatures, providing a controlled baseline for mechanism verification and for interpreting experimental observations. The simulation outputs are used only for physical verification/interpretation and are not used as training, pretraining, or augmentation data for the CNN-RF model.

4.2. Experimental Results: Monitoring System and Modal Validation

To verify the simulation results and assess the real-time performance of the proposed CNN-RF diagnostic framework, a comprehensive experimental campaign was conducted. This included modal testing of the head sheave structure and multi-source data acquisition from a fault simulation platform. The dual aim was to (i) validate structural dynamics through field tests and (ii) acquire labeled fault data under controlled conditions for model inference.

4.2.1. Field Modal Tests

To validate the structural dynamics of the sheave system predicted by simulation, on-site modal excitation tests were performed. The head sheave assemblies comprising upper and lower sheave devices, rotating shafts, and four rolling bearings were tested using impact hammer excitation combined with triaxial accelerometers mounted at key structural nodes. Figure 17 illustrates the hammer excitation position.



Figure 17. Hammer excitation position.

The modal tests revealed several significant vibrational modes within the 50–250 Hz range, aligning with the simulated mode shapes from Section 4.1. The first-order bending mode, in particular, exhibited strong response characteristics in both axial and vertical directions. The measured natural frequencies showed deviations within 3.5% of the finite element simulation results, confirming the mechanical fidelity of the modal model and its applicability to fault diagnosis frameworks.

The test results also confirmed that faults such as bearing seat looseness or shaft misalignment significantly affect local modal energy distribution and vibrational mode coupling. These changes were consistent across simulation and field measurements, justifying the inclusion of modal indicators and fault-sensitive frequency bands in the feature selection and signal preprocessing pipelines used in the CNN-RF architecture.

4.2.2. Fault Testing Platform Data

In parallel with modal validation, a fault simulation platform was deployed on-site to reproduce various fault scenarios under realistic operating conditions. The goal was to gather multi-source signal data vibration, acoustic, displacement, and temperature for robust feature extraction and classifier validation. Vibration sensors were installed to measure signals in three orthogonal directions: axial, vertical, and radial. Due to the high similarity between the vertical and radial responses observed in Figure 18, the final acquisition strategy focused on axial and vertical directions. The sampling frequency was set to 2000 Hz, sufficient to capture resonance features during the hoisting cycle.

To prevent segment-level leakage caused by overlapped sliding windows, we use a group-wise train–test split based on recording run/session IDs. All segments extracted from the same continuous recording are assigned exclusively to either the training set or the test set. Reported performance is averaged over N repeated group-wise splits (mean ± standard deviation).

The constant-speed hoisting phase (35–85 s) was isolated for frequency-domain analysis. The dominant frequency components from the vibration spectrum aligned with fault frequencies predicted from modal and ADAMS simulations, further validating the model's consistency. An acoustic sensor was mounted on the sheave device junction box to minimize signal attenuation and facilitate maintenance access. The frequency-domain analysis (Figure 19) revealed that the dominant acoustic frequency (~180 Hz) differed significantly from the vibration spectrum.

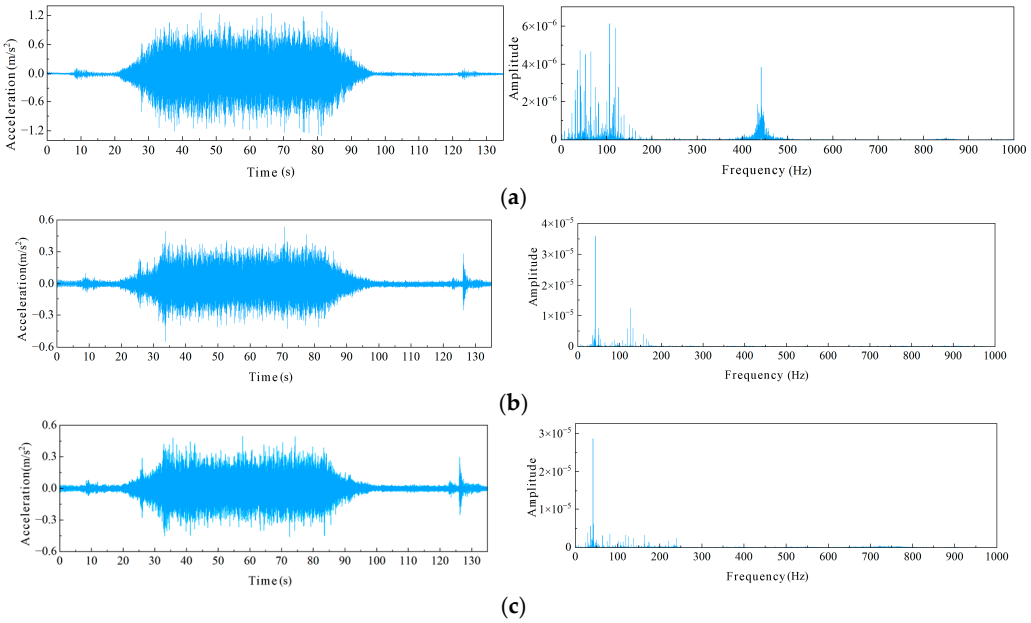


Figure 18. Time domain waveform of vibration signal: (**a**) Axial, (**b**) vertical, (**c**) level.

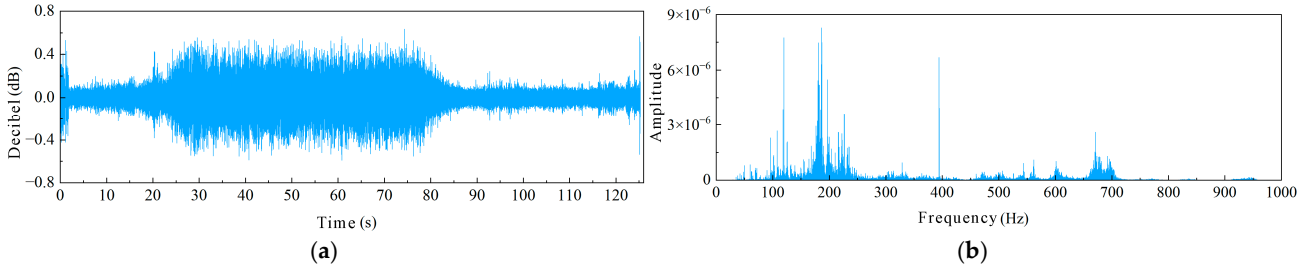


Figure 19. (**a**) Time domain waveform of sound signal (**b**) Sound signal spectrum.

This supports the rationale for incorporating multimodal sensing, as the acoustic channel captured fault-related energy not apparent in vibration data. To monitor wheel body displacement, a non-contact displacement sensor was mounted on a bracket fixed to the protective railing, targeting the sheave flange. The sampling frequency was 500 Hz, and the displacement signal during constant-speed operation was analyzed. Envelope spectrum analysis (Figure 20) revealed a dominant modulation frequency at 1.4 Hz, which closely matches the theoretical sheave rotational frequency derived from its 3.5 m diameter at 8 m/s speed.

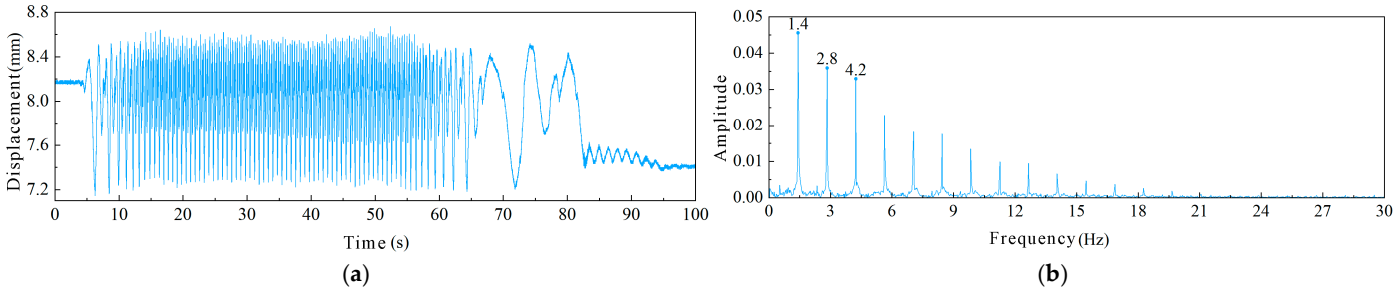


Figure 20. Wheel body displacement signal of head sheave device: (**a**) Time domain diagram and (**b**) envelope spectrum diagram.

This confirmed the platform's capability to track rotational anomalies and support diagnostic analysis involving eccentricity or wheel swing faults. Temperature sensors were embedded into the oil injection ports of the bearing housings. Temperature readings from the upper and lower bearing pairs exhibited trends that correlated with ambient environmental conditions (Figure 21). While these readings do not directly indicate fault, they provide important contextual awareness and are critical for detecting early-stage lubrication failures or thermal degradation.

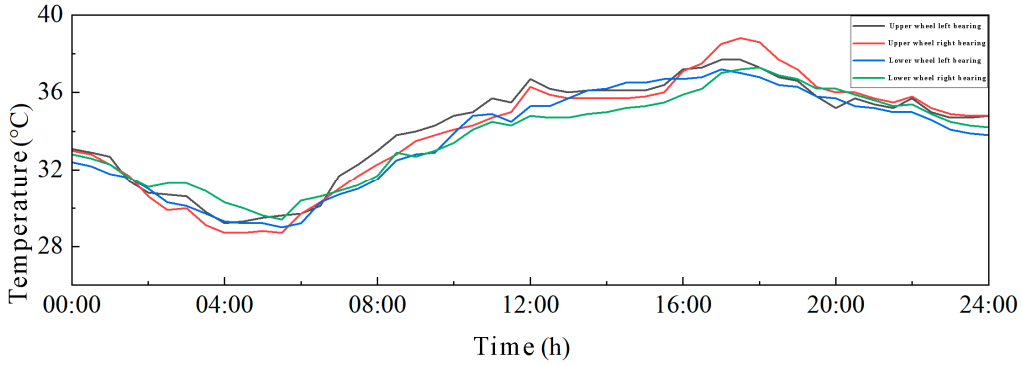


Figure 21. Dynamic curve of bearing temperature.

These experimental results comprehensively validate the simulation framework and demonstrate the practical viability of the multi-source monitoring system in real industrial environments. The collected data not only reflects expected fault-induced behavior but also enhances model training for generalized, deployable fault classification systems.

4.3. Fault Diagnosis Model Performance

To evaluate the diagnostic accuracy and generalization capability of the proposed multi-source CNN-RF model, a series of performance metrics and visualization tools were used. This section presents the classification performance in quantitative terms, compares it to traditional models, and illustrates the learning process and final classification effectiveness via training curves and confusion matrices.

4.3.1. Classification Metrics

The model was evaluated using five standard metrics: Accuracy, Precision, Recall, F1-Score, and Macro-F1, computed over five fault classes, including normal, outer ring fault, inner ring fault, shaft eccentricity, and structural looseness. Accuracy measures the overall correct predictions. Precision reflects the proportion of true positives among predicted positives. Recall evaluates the model's ability to capture all actual positives. F1-Score balances precision and recall. Macro-F1 treats all classes equally, regardless of imbalance.

The CNN-RF model achieved the following average results across all test sets. Accuracy: 97.24%, Precision: 96.98%, Recall: 96.83%, F1-Score: 96.85%, Macro-F1: 96.91%.

These results demonstrate the effectiveness of multi-source feature fusion and the hybrid architecture in capturing both spatial and spectral fault characteristics. Comparative experiments with conventional SVM, decision tree, and shallow CNN models showed at least a 3.5–6.7% improvement in accuracy, confirming the superiority of the proposed framework under noisy and complex working conditions. Meanwhile, all reported metrics are computed under run-wise group splitting to prevent leakage across overlapped segments from the same continuous recording. Results are reported as mean ± standard deviation over $N = 5$ repeated run-wise evaluations.

4.3.2. Training Curves and Confusion Matrix

To ensure that the model generalizes well and avoids overfitting, the loss function and accuracy curves were monitored during training. As shown in Figure 22, the training and validation curves indicate stable convergence 96% after 35 epochs of the proposed model.

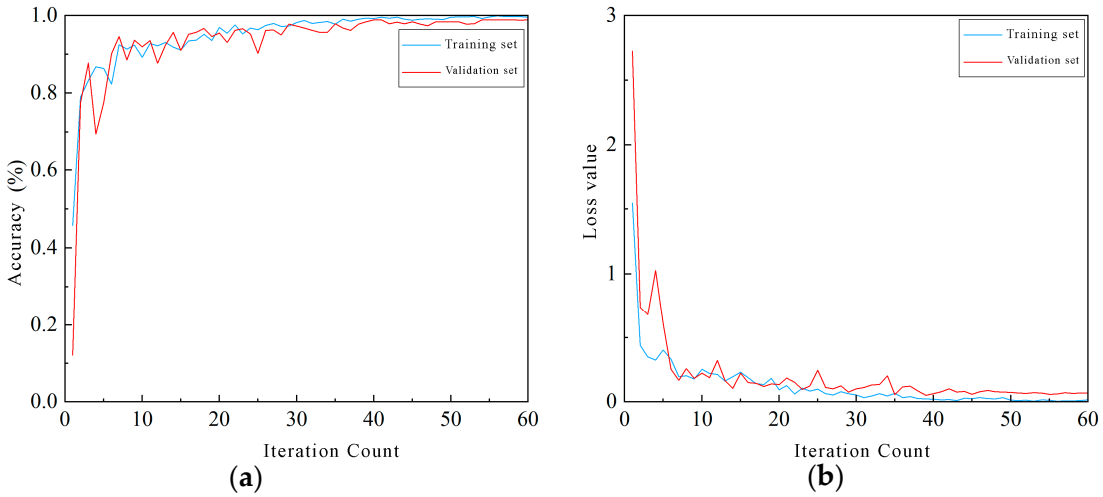


Figure 22. (**a**) Accuracy curve and (**b**) loss curve.

The figure shows that the model training completed 60 iterations, ultimately achieving an accuracy rate of 98.96%, indicating good classification performance without overfitting. During the initial training phase, both the training and validation sets exhibited a rapid upward trend in accuracy. As training progressed, the model's performance metrics gradually stabilized around the 40th iteration. At this point, both classification accuracy and loss values entered a relatively steady state without significant fluctuations, indicating good model stability during training. As shown in Figure 23, to visualize the model's diagnostic performance on the dataset, a confusion matrix is used to present the diagnostic results.

The horizontal axis represents the model's predicted classification labels for fault samples, the vertical axis shows the actual labels of the fault samples, and the right side displays the reference metrics corresponding to the colors in the heatmap. Analysis of the confusion matrix reveals that the multi-source fusion CNN-RF model demonstrates outstanding performance in fault classification. Although it exhibits minor misclassifications for wheel body and shaft faults, its overall classification accuracy reaches 98.96%, confirming the model's robust fault diagnosis capability.

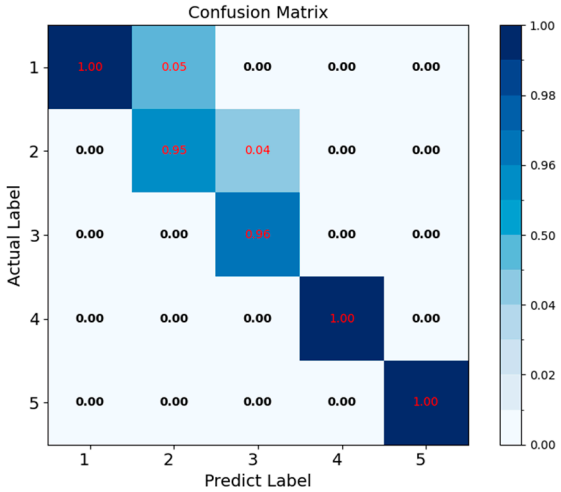


Figure 23. Matrix of confusion.

To validate the superiority of the proposed multi-source fusion CNN-RF model, comparative experiments were designed. These tests evaluated the performance of single-channel approaches and various fault diagnosis models combining MFCC with two additional vibration signal processing methods. As shown in Table 9, the experimental results demonstrate that the proposed multi-source fusion CNN-RF model achieves the best fault recognition performance. Fault diagnosis accuracy reported as mean ± standard deviation over $N = 5$ repeated run-wise (group-wise) evaluations.

Table 9. Comparison of accuracy of different data types.

Data Type	Fault Diagnosis Accuracy (%)
STFT	96.53 ± 0.12
CWT	96.64 ± 0.12
GST	97.5 ± 0.14
MFCC	96.12 ± 0.08
STFT+MFCC	98.36 ± 0.10
CWT+MFCC	98.55 ± 0.08
GST+MFCC	98.96 ± 0.08

By integrating features from both vibration and acoustic signals, this model enables reliable diagnosis of sheave unit faults. Vibration signals acquired via contact-based methods deliver highly precise detection results, while acoustic signals enable effective non-contact monitoring, providing richer operational status information for sheave devices.

5. Discussion

This study proposes a comprehensive fault diagnosis framework for head sheave assemblies in mine hoisting systems, integrating simulation-based modal analysis, multi-source data acquisition, and deep learning-enhanced classification. Through both virtual prototyping and physical experimentation, the system's dynamic behavior, fault signatures, and real-world signal responses have been thoroughly evaluated.

The CNN-RF hybrid model demonstrated high accuracy (over 97%) in classifying multiple fault types, confirming the advantage of combining deep feature extraction with ensemble decision-making. The inclusion of both vibration and acoustic features provided complementary diagnostic information, especially in cases where vibration signals alone lacked clear fault signatures. Additionally, the displacement and temperature data enriched the feature space, allowing the model to account for structural anomalies and environmental factors.

Importantly, the simulation results from ADAMS showed high correlation with field test outcomes. Modal parameters extracted via finite element modeling aligned closely with results from hammer-based field modal tests, validating the mechanical modeling approach. The sheave rotation frequency, cage speed, and fault-induced harmonics observed in the envelope spectra also matched theoretical expectations.

While the monitoring system proved effective in static and dynamic validation, several challenges remain. Sensor installation in harsh mining environments, maintaining signal fidelity under electromagnetic interference, and generalizing the model across equipment types are practical concerns that require further refinement.

6. Limitations and Future Directions

Despite the demonstrated effectiveness of the proposed multi-source CNN-RF fault diagnosis framework, several limitations and future research directions remain. First, the acoustic channel faces substantial challenges in underground mine environments. In underground mines, background noise and reverberation increase the acoustic noise floor compared with laboratory conditions. Although a directional condenser microphone suppresses off-axis interference, its performance may still degrade in the field due to multipath reflections, distance constraints, and dust/moisture contamination. To improve robustness, we adopt appropriate sensor placement and apply standardized preprocessing (e.g., normalization and noise/outlier screening) and noise augmentation during model training. Nevertheless, further improvements could be achieved by incorporating advanced denoising methods, better protective installation designs, and additional sensing modalities for redundancy.

Second, while combining vibration and acoustic signals improves diagnostic performance, the current setup provides limited sensor redundancy. Integrating additional modalities such as torque, motor current, or temperature would enrich feature diversity and increase resilience to sensor failure. Third, although the proposed model performs well under the tested laboratory and field conditions, it has not yet been optimized for edge deployment. Future work will investigate lightweight architectures and model compression techniques (e.g., pruning, quantization, and knowledge distillation) to enable real-time fault recognition on embedded industrial hardware.

In addition, this study is based on relatively short-term testing. Long-term degradation data across multiple operational cycles are required to distinguish transient disturbances from progressive mechanical wear and to evaluate long-term stability. Moreover, the current framework relies on offline training with predefined fault categories; incorporating adaptive, transfer, or semi-supervised learning strategies would allow the model to evolve with shifting operational parameters and newly emerging fault types. Another limitation is the lack of a publicly available benchmark datasets for head sheave diagnostics, which hampers cross-study comparison and independent validation. Finally, the diagnostic output is not yet integrated with maintenance decision-making systems. Future deployment should consider linking fault predictions with computerized maintenance management systems (CMMSs) to support automated alerting, maintenance scheduling, and asset lifecycle optimization. Addressing these limitations would further enhance the practical applicability, scalability, and reliability of the proposed system for safety-critical mining hoisting equipment.

7. Conclusions

This study proposed a multi-source fault diagnosis framework for mining hoisting head sheave systems based on a Convolutional Neural Network and Random Forest (CNN-RF). The framework integrates vibration and acoustic information to improve fault

identification reliability in harsh operating environments. In order to support mechanism understanding and validate the experimental setup, ANSYS-based finite element analysis and ADAMS-based multi-body dynamics simulation were conducted to analyze structural characteristics and fault response behavior; however, it should be emphasized that these simulation results were used only for physical verification and interpretation, and were not used as training, pretraining, or augmentation data for the diagnosis model. All CNN-RF training and testing were performed exclusively using vibration and acoustic signals collected from the physical fault test platform.

For feature construction, vibration signals were transformed into time–frequency representations, and acoustic signals were characterized using Mel-Frequency Cepstral Coefficients (MFCCs). Comparative experiments demonstrate that multimodal fusion consistently outperforms single-modality inputs. In particular, the best performance was achieved by the GST+MFCC fusion configuration with the proposed CNN-RF classifier, reaching an accuracy of 98.96%, which confirms that acoustic information provides complementary fault-related cues beyond vibration alone.

Although the proposed method shows strong performance under the tested operating condition, the current experiments were conducted under a single speed and load setting. Therefore, cross-condition robustness under varying operating speeds and loads remains to be validated. Future work will focus on conducting held-out cross-speed/cross-load evaluations, improving tolerance to sensor degradation or missing modalities (e.g., microphone blockage), optimizing the model for edge deployment through lightweight design and compression, and collecting long-term degradation data to support early warning of progressive wear. These efforts will further enhance the practical applicability, scalability, and reliability of the proposed framework for safety-critical mining hoisting equipment.

Author Contributions: Conceptualization, C.M. and Z.S.; methodology, M.A.I.; software, M.A.R. and J.F.; validation, M.H. and J.F.; formal analysis, M.A.R.; investigation, J.F.; resources, Z.S.; data curation, M.A.I.; writing—original draft preparation, C.M., M.A.I., M.A.R. and M.H.; writing—review and editing, J.F.; visualization, M.A.R. and J.F.; supervision, C.M.; project administration, C.M. and Z.S.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52274155, and 51975569.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Author Zhiyuan Shi was employed by the Mining Products Safety Approval and Certification Center Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
EMD	Empirical Mode Decomposition
CNN-RF	Convolutional Neural Network–Random Forest
MFCC	Mel-Frequency Cepstral Coefficients
SHAP	Shapley Additive Explanations
FTA	Fault Tree Analysis
ICEEMDAN	Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
AFSA-SVM	Adaptive Firefly Algorithm-Support Vector Machine

DL	Deep Learning
CBMA	Contextual Bandit Model Analysis
DGCN	Dynamic Graph Convolution Network
LSTMs	Long Short-Term Memory Networks
MNF	Modal Neutral Files
FFT	Fast Fourier Transform
AE	Acoustic Emission
ST	S-Transform
BPFI	Ball Pass Frequency of Inner Race
BPFO	Ball Pass Frequency of Outer Race
DCT	Discrete Cosine Transform
PCA	Principal Component Analysis
FRFs	Frequency response functions
NDT	Non-Destructive Testing
ACC	Accuracy
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
GST	Generalized S-Transform
CNN	Convolutional Neural Network
CAM	Class Activation Mapping
RF	Random Forest
ML	Machine Learning
CWT	Continuous Wavelet Transform
STFT	Short-Time Fourier Transform

References

1. Ma, C.; Yao, J.; Xiao, X.; Zhang, X.; Jiang, Y. Fault diagnosis of head sheaves based on vibration measurement and data mining method. *Adv. Mech. Eng.* **2020**, *12*, 1687814020941331. [CrossRef]
2. Kou, Z.; Yang, F.; Wu, J.; Li, T. Application of ICEEMDAN energy entropy and AFSA-SVM for fault diagnosis of hoist sheave bearing. *Entropy* **2020**, *22*, 1347. [CrossRef]
3. Chen, X.; Zhu, Z.-C.; Ma, T.-B.; Shen, G. Model-based sensor fault detection, isolation and tolerant control for a mine hoist. *Meas. Control* **2022**, *55*, 274–287. [CrossRef]
4. Li, J.; Xie, J.; Yang, Z.; Li, J. Fault diagnosis method for a mine hoist in the Internet of Things environment. *Sensors* **2018**, *18*, 1920. [CrossRef]
5. Xue, S.; Tan, J.; Shi, L.; Deng, J. Rope tension fault diagnosis in hoisting systems based on vibration signals using EEMD, improved permutation entropy, and PSO-SVM. *Entropy* **2020**, *22*, 209. [CrossRef]
6. Du, L. Fault diagnosis method of rotating machinery based on MSResNet feature fusion and CAM. *J. Vibroeng.* **2024**, *26*, 1600–1615. [CrossRef]
7. Wang, C.; Wang, M. A fault diagnosis method for rotating machinery based on spatiotemporal feature fusion. *J. Mech. Sci. Technol.* **2025**, *39*, 4389–4405. [CrossRef]
8. Zhai, J.; Wang, B.; Mao, H. A fault diagnosis method for rotating machinery based on Gramian angular field and DGCN. In Proceedings of the 3rd International Conference on Communication Networks and Machine Learning, Nanjing, China, 21–23 February 2025; pp. 74–87.
9. Leaman, F.; Vicuña, C.M.; Clausen, E. A review of gear fault diagnosis of planetary gearboxes using acoustic emissions. *Acoust. Aust.* **2021**, *49*, 265–272. [CrossRef]
10. Shi, Z.; Cai, X.; Ma, C.; Liu, M. State evaluation of hoist head sheave based on fault tree and Bayesian network. *J. Phys. Conf. Ser.* **2022**, *2355*, 012010. [CrossRef]
11. Huang, Q.; Li, Z.; Xue, H. Multi-body dynamics co-simulation of hoisting wire rope. *J. Strain Anal. Eng. Des.* **2018**, *53*, 36–45. [CrossRef]
12. You, W.; Shen, C.; Guo, X.; Jiang, X.; Shi, J.; Zhu, Z. A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery. *Adv. Mech. Eng.* **2017**, *9*, 1687814017704146. [CrossRef]
13. Soualhi, A.; Medjaher, K.; Zerhouni, N. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. *IEEE Trans. Instrum. Meas.* **2014**, *64*, 52–62. [CrossRef]

14. Wang, X.; Tang, G.; Yan, X.; He, Y.; Zhang, X.; Zhang, C. Fault diagnosis of wind turbine bearing based on optimized adaptive chirp mode decomposition. *IEEE Sens. J.* **2021**, *21*, 13649–13666. [CrossRef]
15. Wang, P.; Wang, Y.; Wang, X.; Liu, Y.; Zhang, J. An intelligent actuator of an indoor logistics system based on multi-sensor fusion. *Actuators* **2021**, *10*, 120. [CrossRef]
16. Wen, L.; Gao, L.; Li, X. A new snapshot ensemble convolutional neural network for fault diagnosis. *IEEE Access* **2019**, *7*, 32037–32047. [CrossRef]
17. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72*, 303–315. [CrossRef]
18. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1042. [CrossRef]
19. Kankar, P.K.; Sharma, S.C.; Harsha, S.P. Fault diagnosis of ball bearings using machine learning methods. *Expert Syst. Appl.* **2011**, *38*, 1876–1886. [CrossRef]
20. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [CrossRef]
21. Zhai, X.; Qiao, F.; Ma, Y.; Lu, H. A novel fault diagnosis method under dynamic working conditions based on a CNN with an adaptive learning rate. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5013212. [CrossRef]
22. Jin, X.; Zhao, M.; Chow, T.W.; Pecht, M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Trans. Ind. Electron.* **2013**, *61*, 2441–2451. [CrossRef]
23. Zhao, X.; Jia, M.; Bin, J.; Wang, T.; Liu, Z. Multiple-order graphical deep extreme learning machine for unsupervised fault diagnosis of rolling bearing. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–12. [CrossRef]
24. Guo, G.; Li, Z.-B.; Liu, J. Cooperative vehicle localization in VANETs subject to sensing and transmission failures. *IEEE Trans. Veh. Technol.* **2025**. *early access*. [CrossRef]
25. Peng, J.; Shangguan, W.; Chai, L.; Chen, J.; Peng, C.; Cai, B. V2X-enabled platoon control for aperiodic congestion mitigation via moving bottlenecks in mixed traffic environments. *IEEE Trans. Veh. Technol.* **2025**. *early access*. [CrossRef]
26. Wang, X.; Fidge, C.; Nourbakhsh, G.; Foo, E.; Jadidi, Z.; Li, C. Anomaly detection for insider attacks from untrusted intelligent electronic devices in substation automation systems. *IEEE Access* **2022**, *10*, 6629–6649. [CrossRef]
27. Worden, K.; Farrar, C.R.; Manson, G.; Park, G. The fundamental axioms of structural health monitoring. *Proc. R. Soc. A* **2007**, *463*, 1639–1664. [CrossRef]
28. Wang, D.-Y.; Wang, Z.; Zhang, S.-W.; Cheng, D.-J. Diesel engine quality abnormal patterns recognition based on feature fusion and adaptive decision fusion. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2024**, *238*, 465–477. [CrossRef]
29. Wang, Q.; Yang, C.; Wan, H.; Deng, D.; Nandi, A.K. Bearing fault diagnosis based on optimized variational mode decomposition and 1D convolutional neural networks. *Meas. Sci. Technol.* **2021**, *32*, 104007. [CrossRef]
30. Bao, Z.; Liu, C.; Yang, H.; Zhang, J.; Li, Y. From theory to industry: A survey of deep learning-enabled bearing fault diagnosis in complex environments. *Eng. Appl. Artif. Intell.* **2026**, *163*, 113068. [CrossRef]
31. Qi, B.; Liang, J.; Tong, J. Fault diagnosis techniques for nuclear power plants: A review from the artificial intelligence perspective. *Energies* **2023**, *16*, 1850. [CrossRef]
32. Xu, Y.; Li, Z.; Wang, S.; Li, W.; Sarkodie-Gyan, T.; Feng, S. A hybrid deep-learning model for fault diagnosis of rolling bearings. *Measurement* **2021**, *169*, 108502. [CrossRef]
33. Guo, Y.; Liao, W.; Wang, Q.; Yu, L.; Ji, T.; Li, P. Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach. In Proceedings of the Asian Conference on Machine Learning, PMLR, Beijing, China, 14–16 November 2018; pp. 97–112.
34. Ren, L.; Wang, H.; Li, J.; Tang, Y.; Yang, C. AIGC for industrial time series: From deep-generative models to large-generative models. *IEEE Trans. Syst. Man Cybern. Syst.* **2025**, *55*, 7774–7791. [CrossRef]
35. Lei, Y.; He, Z.; Zi, Y. A new approach to intelligent fault diagnosis of rotating machinery. *Expert Syst. Appl.* **2008**, *35*, 1593–1600. [CrossRef]
36. Lubal, Y.S.; Srivastava, S.S.; Jakhete, M.D.; Babu, S.B.T.; Sheeba, A.; Mahender, K. IoT-enabled systems for smart monitoring of mining operations. *Int. J. Environ. Sci.* **2025**, 262–269.
37. Zhang, G.; Chen, C.-H.; Cao, X.; Zhong, R.Y.; Duan, X.; Li, P. Industrial Internet of Things-enabled monitoring and maintenance mechanism for fully mechanized mining equipment. *Adv. Eng. Inform.* **2022**, *54*, 101782. [CrossRef]
38. Dong, C. *Modern Reliability Theory of Structural Systems and Its Applications*; Science Press: Beijing, China, 2001.
39. Kato, M.; Shimada, S. Vibration of PC bridge during failure process. *J. Struct. Eng.* **1986**, *112*, 1692–1703. [CrossRef]
40. Sun, X.; Hardy, H. An investigation on applicability of modal analysis as nondestructive evaluation method in geotechnical engineering. In Proceedings of the International Modal Analysis Conference, SEM, San Diego, CA, USA, 3–7 February 1992; p. 9.
41. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]

42. Qamar, S.; Öberg, R.; Malyshev, D.; Andersson, M. A hybrid CNN–Random Forest algorithm for bacterial spore segmentation and classification in TEM images. *Sci. Rep.* **2023**, *13*, 18758. [CrossRef] [PubMed]
43. Cao, X.; Duan, Y.; Wang, G.; Zhao, J.; Ren, H.; Zhao, F.; Yang, X.; Zhang, X.; Fan, H.; Xue, X.; et al. Research review on life-cycle health management and intelligent maintenance of coal mining equipment. *J. China Coal Soc.* **2025**, *50*, 694–714.
44. Wu, M.; Yao, Z.; Verbeke, M.; Karsmakers, P.; Gorissen, B.; Reynaerts, D. Data-driven models with physical interpretability for real-time cavity profile prediction in electrochemical machining processes. *Eng. Appl. Artif. Intell.* **2025**, *160*, 111807. [CrossRef]
45. Wu, M.; Shukla, S.; Vrancken, B.; Verbeke, M.; Karsmakers, P. Data-driven approach to identify acoustic emission source motion and positioning effects in laser powder bed fusion with frequency analysis. *Procedia CIRP* **2025**, *133*, 531–536. [CrossRef]

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