electronics

MDPI

Article

Electric Compressor NVH Detection System Based on Acoustic-Vibration Signal Fusion

Xixiang Wu [1], Weijun Fan [1,2,*], Liping Chen [3], Xinxin Tan [3] and Haotong Fan [3]

[1] College of Metrology Measurement and Instrumentation, China Jiliang University, Hangzhou 310018, China; s24020804066@cjlu.edu.cn
[2] College of Modern Science and Technology, China Jiliang University, Jinhua 322002, China
[3] Hangzhou Wolei Intelligent Technology Co., Ltd., Hangzhou 310000, China
* Correspondence: 04a0202035@cjlu.edu.cn

Abstract

To address the vibration and noise generated during the pressure buildup process of an electric compressor at constant speed, a noise, vibration, and harshness (NVH) detection system was designed. Using cylinders and other actuating components, the product was clamped and sensors were mounted, while a soundproof enclosure was employed to reduce external noise interference. Based on the working principle of the electric compressor, a pneumatic system was designed to simulate real operating conditions. A real-time sound and vibration acquisition system was developed using an NI USB-4431 acquisition card, a triaxial accelerometer, and a microphone. Mel-frequency cepstral coefficient (MFCC) feature matrices were extracted from the sound signals, and short-time Fourier transform (STFT) was applied to the triaxial vibration signals to obtain feature matrices. A four-channel feature matrix was then constructed as input through acoustic-vibration fusion. Hyperparameter search was carried out using the Optuna framework, pruning, early stopping, and cross-validation to obtain the optimal parameter combination, and the results were compared with those of 1D-CNN, ResNet18, and MobileNetV3 models. Ablation experiments indicate that both multi-axis vibration fusion and acoustic-vibration fusion improve OK/NG classification accuracy over single-axis inputs. Based on the EfficientNet-B0 model, binary OK/NG classification under combined high-speed and low-speed conditions was performed, achieving a classification accuracy of 98.75%, with training time within 3 h and detection time within 9 s.

Keywords: electric compressor; NVH; Mel-frequency cepstral coefficients; short-time Fourier transform; acoustic-vibration fusion; EfficientNet-B0

Check for updates

Academic Editor: Young-Ho Seo

Received: 22 September 2026
Revised: 21 October 2026
Accepted: 23 October 2026
Published: 25 October 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

An electric scroll compressor is a positive-displacement compressor that integrates an electric motor and a scroll compression mechanism into one unit. It mainly consists of a fixed scroll, an orbiting scroll, a bearing housing, and a controller and is widely used in the air-conditioning systems of new energy vehicles and some conventional fuel vehicles. According to statistics, the global automotive electric scroll compressor market reached approximately RMB 8.9 billion in 2024 and is expected to grow at a compound annual growth rate (CAGR) of 16.9% from 2025 to 2031. Battery electric vehicles have no engine to mask noise, so the noise from the electric compressor becomes particularly prominent.

In compressor performance evaluation, vibration and noise are key indicators, and implementing early fault detection and classification is of great significance for improving the overall performance and reliability of electric scroll compressors [1].

Traditional detection methods for rotating machinery mainly rely on signal processing techniques, such as fast Fourier transform [2], Hilbert-Huang transform [3], variational mode decomposition (VMD) [4], empirical mode decomposition (EMD) [5], wavelet packet transform [6], and time-domain analysis [7]. These methods can extract fault features in most cases; however, in complex high-noise environments, fault features become blurred and detection accuracy drops substantially. In addition, because the noise of an electric scroll compressor usually originates from the compressor itself and is generated by the coupling of its components [8], this poses a major challenge to traditional fault detection and further reduces accuracy.

With the development of artificial intelligence, researchers have proposed many deep learning-based fault diagnosis models. Significant results have been achieved with convolutional neural networks (CNN) [9], residual networks (ResNet) [10], generative adversarial networks (GAN) [11], variational autoencoders (VAE) [12], Transformers [13], and long short-term memory networks (LSTM) [14]. The deployment of deep learning methods in industrial fault diagnosis is mainly limited by challenges at two levels. At the data and model level, these include imbalanced data samples, weak adaptability to dynamically changing operating conditions, and optimization difficulties in decoupling and identifying compound fault features. At the deployment and application level, they are reflected in the practical training difficulties of semi-supervised learning frameworks and the severe constraints on deploying models in real-time, resource-limited industrial environments. To address these challenges, existing studies have proposed many solutions. Jin, Y. et al. [15] proposed a bearing fault diagnosis method based on SMOTENC and deep transfer learning, which has significant advantages in handling imbalanced data. Zhang, X. et al. [16] proposed a multi-scale holographic spectrum CNN (MH-CNN) based on two-dimensional multi-scale feature fusion and decision-level feature fusion, improving fault diagnosis accuracy under variable operating conditions. Pietrzak, P. et al. [17] proposed a classification and diagnosis framework that integrates continuous wavelet transform and artificial intelligence, solving the problem of detecting and automatically classifying early inter-turn short-circuit faults in the stator windings of permanent magnet synchronous motors. Irfan, M. et al. [18] proposed a bearing fault diagnosis method based on complex wavelet transform and a lightweight fuzzy CNN (SOF-CNN).

With advances in sensor technology, sound-based detection has also developed. Nakamura, H. et al. [19] proposed identifying motion conditions based on the rotational sound of a motor during no-load testing and using LSTM to determine defects. Shirdel, S. et al. [20] proposed a non-contact motor fault detection method based on sound signals, fusing MFCC and wavelet scattering features, and combining ensemble learning with a CNN classifier; its efficiency was verified on standard datasets through an innovative fusion strategy. Lu, S. et al. [21] proposed a new method based on transient signal angular resampling (TSAAR), which uses sound signal analysis to achieve motor bearing fault diagnosis under variable-speed conditions; this method can adaptively eliminate frequency ambiguity and extract fault characteristic frequencies without an external speed sensor. Salazar-Villanueva, F. et al. [22] proposed a non-invasive induction motor fault monitoring method based on sound signals that combines intrinsic mode functions (IMFs) with fast Fourier transform, reliably identifying characteristic frequencies under a single fault state. Zastepa, M. [23] proposed a sound signal fault diagnosis method based on EMD and LSTM; by extracting statistical features of IMFs and feeding them into a customized LSTM network, rapid and highly accurate non-invasive diagnosis of five types of faults in single-phase

induction motors was achieved. Amarnath, M. et al. [24] used low-cost sound signals combined with statistical feature extraction and a decision tree classification algorithm, achieving effective bearing fault diagnosis through feature selection and dimensionality reduction. Ahvazi, P. et al. [25] proposed an improved MobileNetV2 model incorporating an attention-guided dynamic convolution (ADC) layer, which processes sound signals in noisy environments through Mel spectrograms, achieving high-accuracy, low-computational-cost non-contact diagnosis of five types of motor bearing faults. Sound-based detection has great advantages for early fault detection and is easy to install without affecting equipment operation, but it is highly susceptible to environmental influences and, as a single evaluation indicator, often lacks generalization and robustness.

In recent years, acoustic-vibration fusion technology has been widely studied in fault diagnosis of rolling bearings and gearboxes, and various fusion diagnosis methods combining signal processing and deep learning have been proposed. Fang, X. et al. [26] proposed a rolling bearing fault diagnosis method based on cross-fusion of vibration-acoustic data, interval division, periodic sampling, and fast Fourier transform combined with 1D-CNN. Wu, G. et al. [27] proposed a two-step vibration-acoustic signal fusion method for detecting weak fault features in rolling bearing systems. Li, C. et al. [28] proposed a gearbox fault diagnosis method using acoustic-vibration signals based on deep random forest fusion. Gu, X. et al. [29] proposed an improved multimodal fusion method for compound fault diagnosis of rolling bearings using acoustic-vibration fusion. Xu, Z. et al. [30] proposed a multi-sensor hybrid multimodal fusion diagnosis method based on PCA feature fusion and ResNet+SVM integration. Wan, H. et al. [31] proposed a fused multi-scale convolutional neural network for fusion diagnosis of sound and vibration signals under variable rotational speeds. Hartono, D. et al. [32] proposed an acoustic-vibration joint time-frequency analysis method based on the residual method and generalized linear chirplet transform, solving the problem of inaccurate assessment of gear fault severity when acoustic measurements are disturbed by noise. Dai, M. et al. [33] proposed a lightweight convolutional neural network based on frequency-domain feature fusion for acoustic-vibration fusion diagnosis. Li, X. et al. [34] proposed a lightweight multi-scale time-frequency fusion model, which, through dual-branch feature extraction, frequency attention adaptive filtering, and an improved cross-attention fusion module, achieves a balance between high accuracy and low parameter count in motor fault diagnosis, significantly improving diagnostic efficiency and robustness.

To address these challenges, this paper develops an industrial NVH inspection system for electric scroll compressors. The technical contribution lies in the system-level integration and engineering validation of acoustic-vibration NVH inspection under real production constraints. The main contributions are as follows:

(1) An industrial NVH inspection platform is designed and built, including a pneumatic system with proportional valves and air tanks, a soundproof enclosure, product clamping, and sensor mounting. The platform reproduces the pressure buildup process under two representative speed conditions.
(2) A four-channel acoustic-vibration representation is constructed by combining MFCC features of the acoustic signal with STFT time-frequency features of the triaxial vibration signals. This representation is adapted to EfficientNet-B0 by modifying the first convolutional layer from three to four input channels.
(3) The system is validated under real production constraints, where faulty units are scarce and must still be capable of running and building pressure. The dataset, augmentation rules, and product-level grouping are described in detail to ensure reproducibility.
(4) Diagnostic accuracy is 98.75%, and the single detection time is within 9 s.

2. Product Structure and Noise Generation Mechanism

2.1. Operating Principle of the Electric Scroll Compressor

The core working principle of an electric scroll compressor is based on a pair of intermeshing scrolls (a fixed scroll and an orbiting scroll) to compress gas [35]. The main structure is shown in Figure 1.

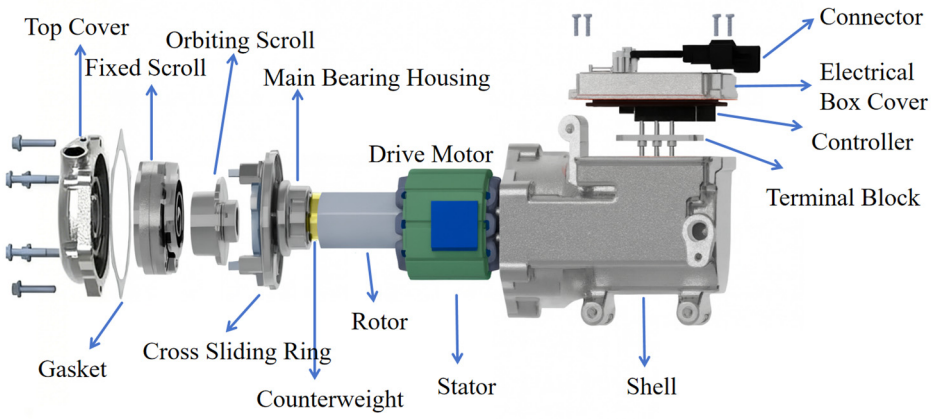


Figure 1. Electrically driven scroll compressor structure.

It is driven directly by a drive motor (comprising a stator and a rotor). An Oldham coupling is used to prevent the orbiting scroll from rotating about its own axis so that it undergoes orbital motion supported by the main bearing housing. The profiles of the two scrolls form a series of sealed chambers with gradually decreasing volumes, thereby enabling gas suction, compression, and discharge. The working process of the scrolls is shown in Figure 2, where the red scroll represents the orbiting scroll and the blue scroll represents the fixed scroll. As the orbiting scroll continues its orbiting motion, gas is drawn in, gradually compressed toward the center, and then discharged. To ensure smooth operation, a counterweight is mounted on the rotor to offset inertial forces. The entire assembly is sealed within the housing, and the controller provides electrical control through connectors and terminal blocks, while the exterior is enclosed by the top cover and the electrical box cover. This structure achieves efficient, low-vibration positive-displacement compression.

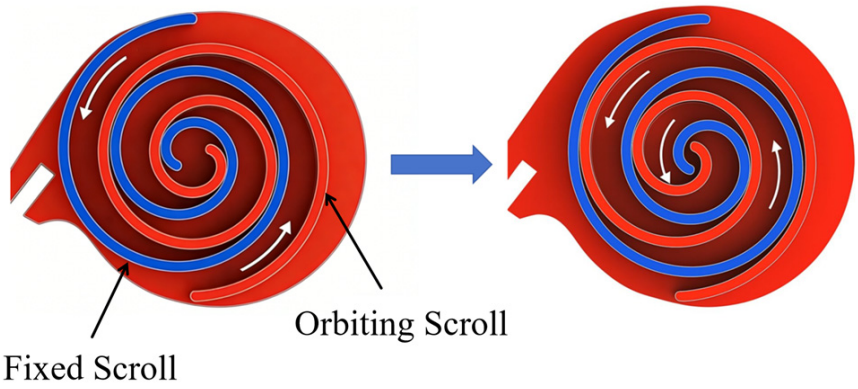


Figure 2. Scroll mechanism operation.

2.2. Noise Generation Mechanism

The noise of electric compressors mainly originates from three types: aerodynamic noise, mechanical noise, and electromagnetic noise.

(1) Aerodynamic noise: caused by airflow pressure pulsation, friction, and vortices. External motors are usually equipped with fans, and the noise sources include fan rotational noise, vortex noise, and whistling noise generated by interference between airflow and structures. Internal motors have no fan, and their noise mainly comes from disturbances when the cooling fluid flows through passages and the air gap.

(2) Mechanical noise: originates from friction, impact, unbalance, and structural resonance of moving parts. It mainly includes bearing noise (rolling element bearings are affected by geometric defects, lubrication, assembly, etc.; sliding bearings are affected by machining accuracy, load, etc.) and vibration noise caused by rotor unbalance. For motors that share a bearing with the compressor or have a cantilevered structure, load variations or rotor oscillation can intensify the noise.

(3) Electromagnetic noise: generated by stator vibration excited by electromagnetic force waves in the air gap. The radial electromagnetic force is the main cause. It arises from the interaction between the fundamental and harmonic magnetic fields and leads to periodic deformation and vibration of the stator core [36]. Electromagnetic noise is the main source of the overall noise of the electric compressor [37].

The vibration and noise generated during compressor operation are formed by the superposition of the vibration and noise from the compressor and the motor. The noise and vibration of the electric scroll compressor can be broadly associated with different frequency ranges. Low-frequency components are mainly related to mechanical and rotor-related sources, mid-frequency components to aerodynamic pulsation, and high-frequency components to electromagnetic excitation and structural resonance.

3. NVH Detection Platform

The testing method used in this paper is shown in Figure 3. It mainly consists of five stages: product clamping and sensor mounting, simulation of real operating conditions, raw signal acquisition, fault feature extraction, and binary OK/NG classification based on acoustic-vibration signal fusion and the EfficientNet-B0 network.

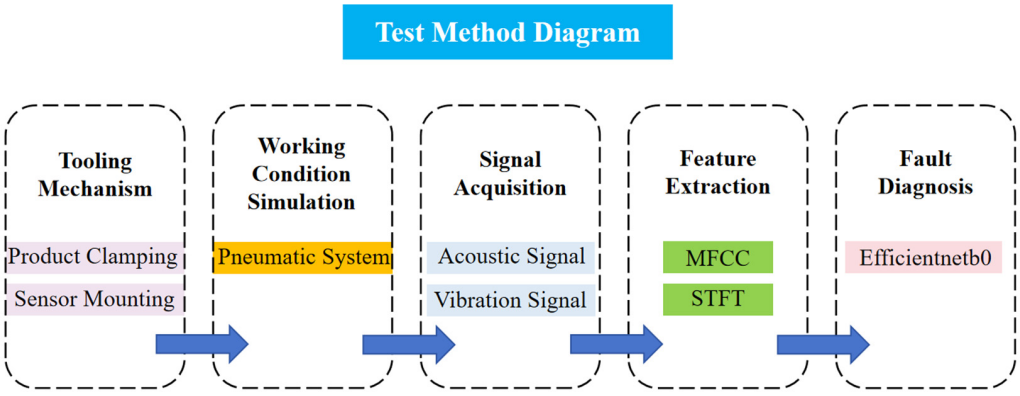


Figure 3. Test method diagram.

3.1. Product Clamping and Sensor Mounting

The NVH test platform is shown in Figure 4. The product is placed inside a soundproof enclosure to ensure that the ambient noise during testing remains below 50 dBA. The product is secured by a downward-pressing cylinder, the microphone is installed in position together with the downward-pressing cylinder, and a triaxial vibration accelerometer is loaded by an NVH cylinder.

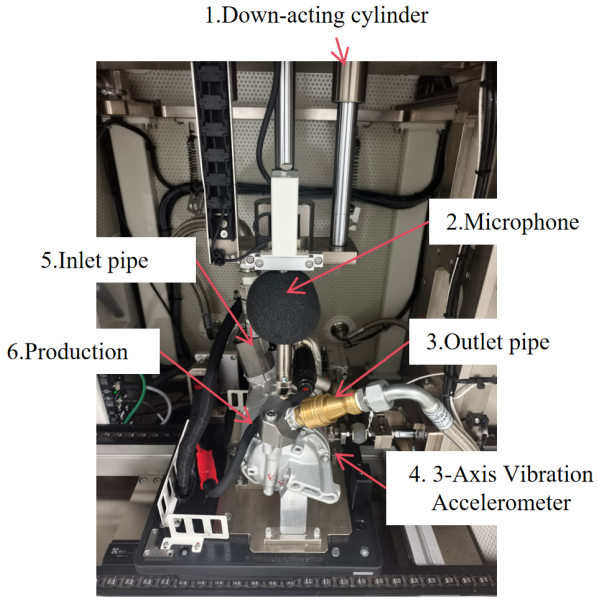


Figure 4. NVH test platform.

3.2. Pneumatic System for Simulating Real Operating Conditions

The pneumatic system was designed to simulate real operating conditions, as shown in Figure 5. The air source of the experimental system is provided by the main air tank. The gas first passes through a dryer for dehumidification so as to reduce the influence of moisture in the gas on test stability and sensor measurement accuracy. The gas then passes sequentially through the electro-pneumatic proportional valve SV2 and a pneumatic ball valve into the electric scroll compressor test system. The electro-pneumatic proportional valve SV2 is used to regulate the system flow rate and inlet pressure. In this experiment, the inlet pressure was set to 1 barG or 0.5 barG.

The system achieves multi-condition pressurization testing through different air circuit control modes. The pneumatic ball valve QV2 is opened for overall system pressurization control. When QV3 is opened, the low-speed condition test is carried out, during which the system pressure rises from 5 barG to 9 barG, with the 7–9 barG range serving as the low-speed test stage; the compressor motor speed was set to 1500 r/min.

After the low-speed test is completed, QV3 is closed and QV4 is opened for high-speed condition pressurization. The system likewise rises from 5 barG to 9 barG, with the 7–9 barG range serving as the high-speed test stage; the compressor motor speed is set to 6500 r/min. The electro-pneumatic proportional back pressure valve SV1 is initially set to 9.5 barG to prevent gas leakage during pressurization. After the test, it is adjusted to 5 barG, while QV1 and QV2 are closed and QV3, QV4, and QV5 are opened for pressure relief.

Because after the gas enters the compressor, its pressure and temperature increase significantly during compression, the compressed high-temperature and high-pressure gas carries some entrained oil. Oil-gas separation is then carried out through an oil mist separator, after which the oil mist is collected by exhaust purifier 1. Exhaust purifier 2 collects the oil mist in the oil storage tank. QV6 and QV7 discharge the residual oil in the system, and the manual ball valve PV1 drains the oil from the oil storage tank into the oil recovery system.

By setting different speed and pressure conditions, the complex operating states and load variation characteristics of the electric compressor during actual operation can be effectively simulated, thereby improving the adaptability of the fault detection model to complex operating conditions.

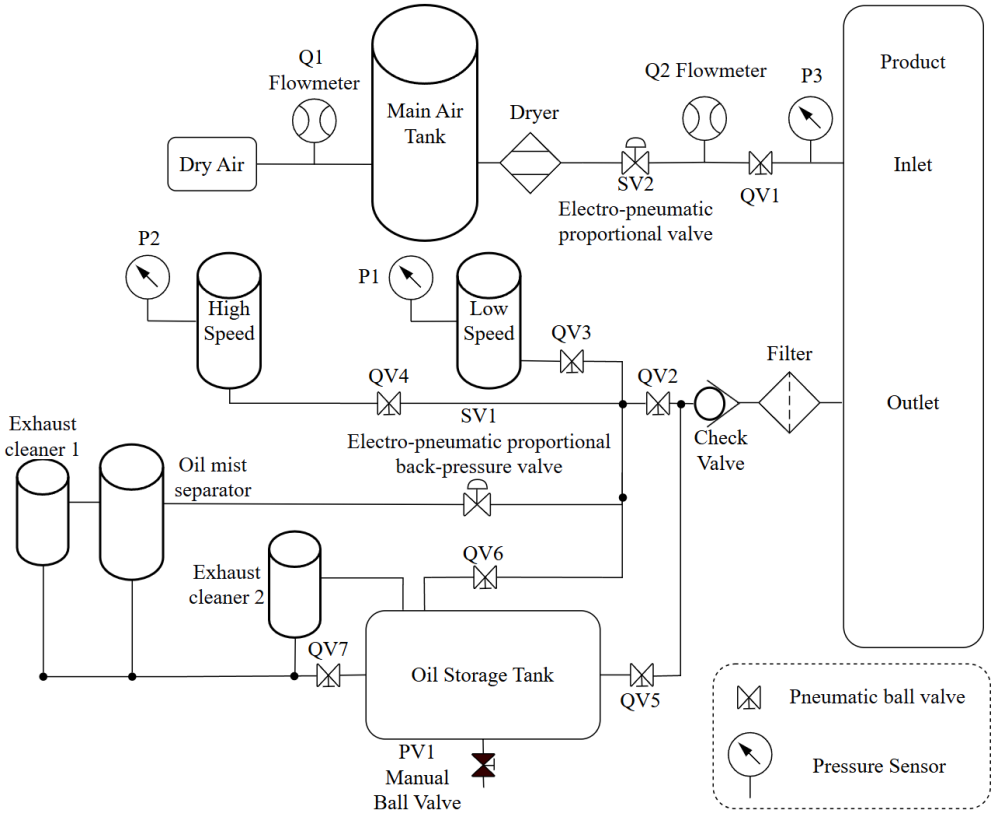


Figure 5. Pneumatic system diagram.

3.3. Raw Signal Acquisition

Vibration and acoustic data were acquired using an NI USB-4431 high-speed data acquisition device, a KISTLER 8763B050BB 3-axis vibration accelerometer, and a Hangzhou Aihua microphone, as shown in Figure 6.

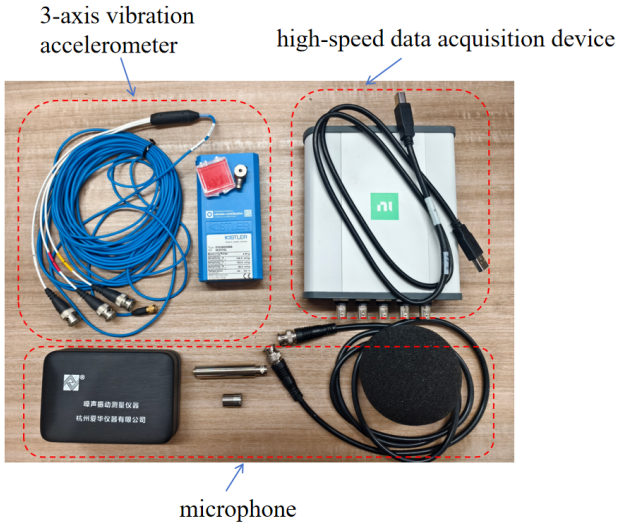


Figure 6. Data acquisition equipment.

When the compressor operates, the main noise energy is concentrated in the audible frequency range, generally below 20 kHz, while vibration is mainly concentrated below 8 kHz. According to the Nyquist sampling theorem, the sampling rate was therefore set to 51.2 kHz. Two operating conditions were considered: a low-speed condition and a high-speed condition, with low and high speeds of 1500 r/min and 6500 r/min, respectively.

3.4. Fault Feature Extraction

3.4.1. Data Preprocessing

First, DC removal was performed on the raw signals. The impulse response of a fault tends to excite high-frequency resonance; removing the DC component allows subsequent analysis to focus more on the impulsive components that carry fault information [38]. Then, the STFT spectrogram of the vibration signals and the Mel-frequency cepstral coefficients (MFCCs) of the acoustic signals were calculated separately.

3.4.2. STFT

Among common time-frequency methods, STFT is adopted here because it is well suited to non-stationary vibration signals. It divides the signal into short frames, applies a window, and computes the discrete Fourier transform for each frame, revealing how the spectrum evolves over time. For an input single-channel vibration signal $x[n]$, the STFT of the m-th frame is defined as

$$X(m,k) = \sum_{n=0}^{N-1} x[n+mH]w[n]e^{-j\frac{2\pi}{N}kn} \quad (1)$$

where $w[n]$ is the window function, N is the frame length (i.e., the number of FFT points), and H is the frame shift. To compress the dynamic range, highlight relative energy differences, and enhance the robustness of fault features, the complex spectrum $X(m,k)$ is further converted into a magnitude spectrum and mapped to the dB scale:

$$S_{\mathrm{dB}}(m,k) = 20\log_{10}\frac{|X(m,k)|}{\max\limits_{m,k}|X(m,k)|} \quad (2)$$

In this paper, the Hanning window was selected as the window function, which effectively suppresses spectral leakage while maintaining reasonable frequency resolution and avoiding strong harmonics masking adjacent weak fault sidebands. The frame length was set to 80 ms, i.e., 4096 points, giving a frequency resolution of 12.5 Hz, which can clearly resolve rotational harmonics of various orders. In addition, the window contains multiple fundamental periods, ensuring the stability of spectrum estimation, and this duration is sufficient to capture millisecond-level transient impulses. The frame shift was set to 40 ms, i.e., 2048 points, with 50% overlap, which is used to smooth amplitude transitions between frames, eliminate boundary discontinuities, and increase the observation density along the time axis.

The vibration frequency range is 0–8000 Hz. This frequency range was analyzed in this paper, and the resulting STFT spectrogram is shown in Figure 7.

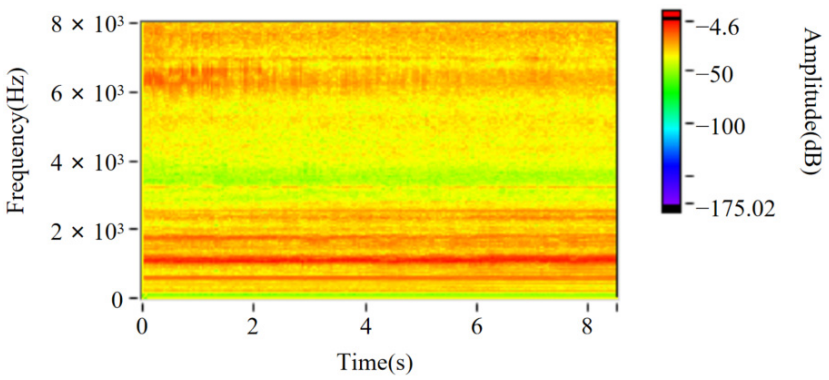


Figure 7. STFT result.

3.4.3. MFCC

MFCC simulates the nonlinear perceptual characteristics of the human ear, and its approximate relationship with frequency is given as follows [39]:

$$f_{mel} = 2595 \times \lg(1 + \frac{f}{700}) \tag{3}$$

where f_{mel} denotes the Mel scale and f denotes frequency.

MFCC is one of the most commonly used features in speech signal processing and audio analysis. It mainly involves five steps: pre-emphasis, signal framing and windowing, power spectrum computation, application of a Mel filter bank to the obtained power spectrum, and calculation of the logarithmic values of all filter bank outputs, as shown in Figure 8 [40].

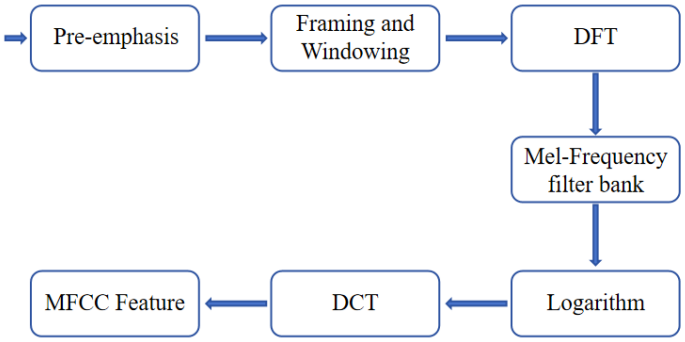


Figure 8. MFCC flowchart.

Step 1. Pre-Emphasis

Typically, a first-order high-pass filter is used, as given in Equation (4), to compensate for the attenuation of high-frequency components in the acoustic signal, boost the high-frequency components, and make the signal spectrum flatter.

$$y(n) = x(n) - \alpha \cdot x(n-1) \tag{4}$$

where n denotes the sample index, $y(\cdot)$ denotes the signal after pre-emphasis, $x(\cdot)$ denotes the original signal, and α typically ranges from 0.9 to 1.0 [40]; in this paper, it is taken as 0.97.

Step 2. Framing and Windowing

Because acoustic signals are non-stationary but approximately stationary over short intervals, and to mitigate spectral leakage, the signal is framed and windowed. The frame length is 80 ms and the frame shift is 40 ms, giving 40 ms overlap between adjacent frames. This overlap allows the temporal evolution of the signal to be tracked while each frame remains approximately stationary. Each frame is then multiplied by a window function to smooth the edges and reduce edge effects. The Hanning and Hamming windows are among the best-known candidate window functions [40], and the Hamming window is adopted in this paper.

Step 3. DFT

Since signal characteristics are usually difficult to discern from time-domain variations alone, a DFT is performed on each frame to calculate the power spectrum of each frame, as given by the following equation:

$$Y(k) = \sum_{n=0}^{N-1} y(n) e^{-j2\pi/N},\ 0 \le k \le N-1 \tag{5}$$

where N denotes the number of Fourier transform points, $Y(k)$ denotes the frequency-domain signal, and $y(n)$ denotes the time-domain signal.

Step 4. Mel-Frequency Filter Bank

The power spectrum is then passed through a set of Mel-scale triangular filters. A filter bank with M filters is defined, where triangular filters are adopted, M is taken as 26 [41], and the center frequency is $f(m)$. The frequency response of the triangular filter is defined as [40]:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \le k < f(m) \\ 1 & k = f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (6)$$

where $H_m(k)$ denotes the frequency response of the m-th filter.

Then, the logarithmic energy output of each filter is calculated:

$$s(m) = \ln\left(\sum_{k=0}^{N-1} |Y(k)|^2 H_m(k)\right), 0 \le m \le M-1 \quad (7)$$

Step 5. DCT

Finally, post-processing operations, including logarithmic computation and DCT are performed to convert the filtered signal into MFCC features, as given below [42]:

$$C(n) = \sum_{m=0}^{M-1} s(m) \cos\left(\frac{\pi n(m+0.5)}{M}\right) \quad (8)$$

where $n = 1, 2, \ldots, L$, with L denoting the order of the MFCCs, typically ranging from 12 to 16 and taken as 13 in this paper. $S(m)$ is the output energy of the m-th frequency band, and $C(n)$ is the output feature. The resulting MFCC feature map is shown in Figure 9.

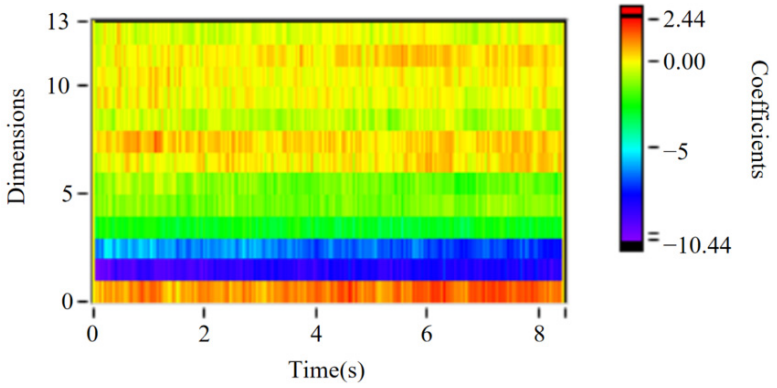


Figure 9. MFCC feature map.

3.5. Binary OK/NG Classification Based on EfficientNet-B0

EfficientNet-B0 is a lightweight convolutional neural network that uses a compound scaling strategy to balance depth, width, and input resolution. It consists of an initial convolutional stem, stacked MBConv modules with SE attention, a transition convolution, global average pooling, and a classification head. For the four-channel acoustic-vibration input in this study, the first convolutional layer of the native EfficientNet-B0 is adapted by changing the input channels from 3 to 4.

The binary OK/NG classification workflow based on EfficientNet-B0 is shown in Figure 10.

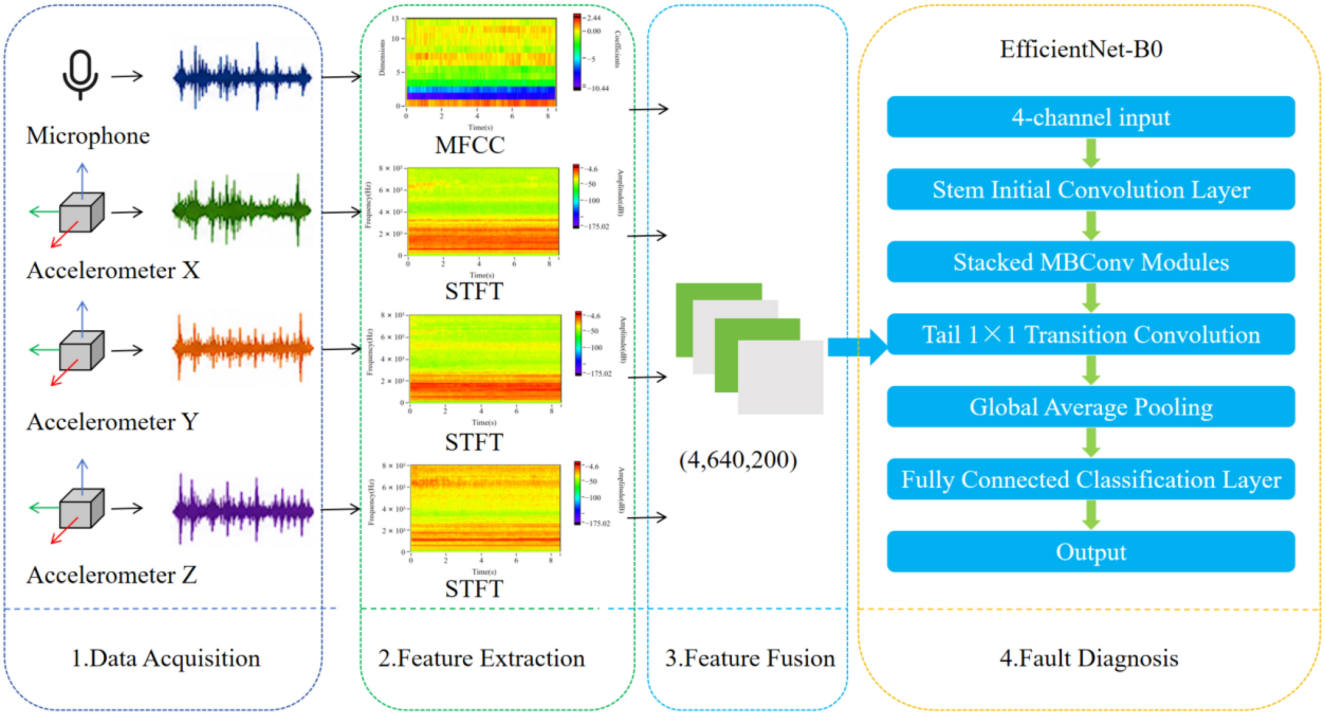


Figure 10. Binary OK/NG classification flowchart.

Step 1. Vibration and sound signals were acquired using a triaxial accelerometer and a microphone, respectively.

Step 2. MFCC was used to extract features from the sound signal, yielding a 13-dimensional feature matrix with variable time length. Meanwhile, STFT was used to extract features from the vibration signal, yielding a feature matrix with a fixed frequency dimension of 640 points and variable time length.

Step 3. Each of the 13 MFCC dimensions was repeated 50 times, resulting in 650 dimensions. In addition, since the duration of each test was not fixed and the test times under high-speed and low-speed conditions differed greatly, the four-channel feature matrices were respectively resized to (640, 200) using bilinear interpolation and then concatenated along the channel dimension into (4, 640, 200).

Step 4. The EfficientNet-B0 model was used to perform binary classification on the four-channel input and output the result.

3.6. Model Training and Results Analysis

3.6.1. Model Training

The product set consisted of 48 faulty units and 83 normal units, and each unit underwent both low-speed and high-speed tests. The samples were divided into four operating conditions: low-speed OK, low-speed NG, high-speed OK, and high-speed NG. In this paper, joint high-speed and low-speed training was adopted for binary classification. There were 83 low-speed OK units, 83 high-speed OK units, 48 low-speed NG units, and 48 high-speed NG units. The test set consisted of 20 OK products and 20 NG products, randomly selected at the product level. Since each product was tested at both speeds, this yielded 40 unique test products and 80 test recordings, ensuring that the high-speed and low-speed data of the same unit appeared only in either the test set or the training set. The remaining data were expanded to 100 samples through data augmentation, with one of three methods randomly selected: adding Gaussian white noise, amplitude scaling, and time stretching. The specific dataset composition is shown in Table 1. Representative fault

types are shown in Figure 11, from left to right, a missing orbiting-scroll positioning ring, a loose bolt, orbiting-scroll wear, and fixed-scroll wear.

Table 1. Dataset composition.

Operating Condition	Number of Original Training Samples	Number of Training Samples	Number of Test Samples
Low-speed OK	63	100	20
High-speed OK	63	100	20
Low-speed NG	28	100	20
High-speed NG	28	100	20

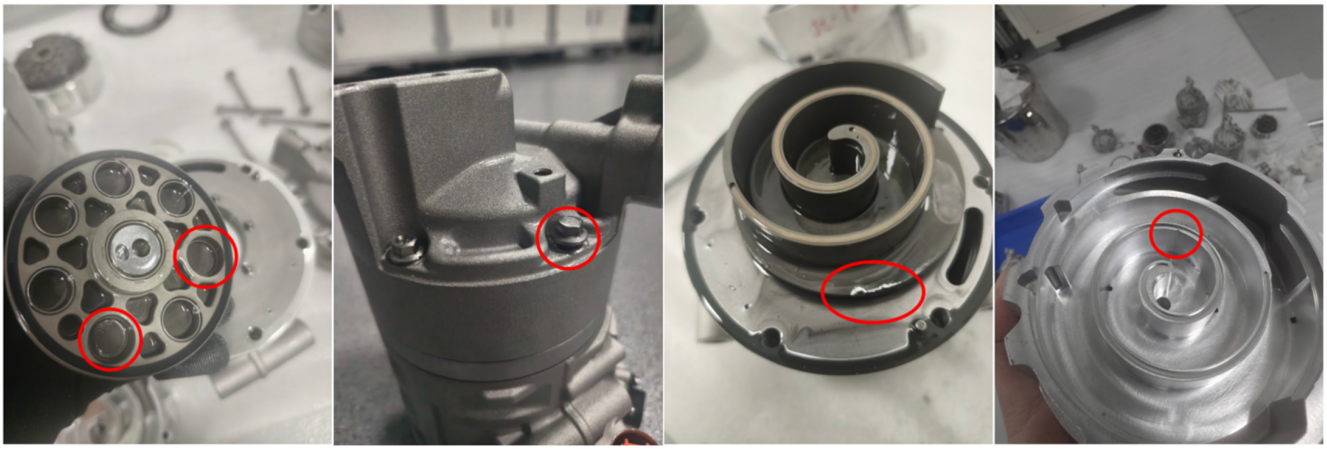


Figure 11. Fault type (partial).

During the data acquisition stage, the diversity of training samples was enhanced by setting different product inlet pressures, slightly changing the microphone position, and keeping the noise inside the soundproof enclosure between 45 dB and 50 dB. In the model evaluation and hyperparameter selection stage, random five-fold cross-validation was introduced. For each fold, 9 OK products and 9 NG products were randomly selected at the product level as the validation set. Since each product was tested at both speeds, this yielded 18 unique validation products and 36 validation recordings, ensuring that both the low-speed and high-speed data of the same product appeared only in the validation set, yielding a total of 36 data samples as the validation set. Based on the parent-product relationship, the training samples expanded from the validation set were removed to prevent data leakage. Because the number of parent products differed among operating conditions, the proportions of samples from different operating conditions in the training set of each fold fluctuated slightly but were roughly similar overall. Validation set accuracy was used as the performance metric. The TPE (Tree-structured Parzen Estimator) sampler in the Optuna framework was used to automatically search the continuous learning rate, batch size, and optimizer type, with a total of 20 trials. Median pruning and an early stopping strategy were adopted (patience set to 25, epochs set to 100, and random seed set to 42). The hyperparameter search ranges are shown in Table 2. Finally, the hyperparameter combination with the highest average cross-validation accuracy was selected for model retraining on the full training set and evaluation on the independent test set. No augmented samples were included in the test set, and augmented descendants of validation products were excluded from the training set to prevent data leakage.

Table 2. Hyperparameter search space.

Parameter	Definition	Range
lr	Learning rate	1×10^{-5}~1×10^{-2}
batch_size	Batch size	8, 16
optimizer_name	Optimizer	"Adam", "SGD"

3.6.2. Evaluation Metrics

The evaluation metrics are defined as follows. This paper deals with a binary classification problem, in which NG items are regarded as the positive class and OK items as the negative class. In the formulas, *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives, and *FN* is the number of false negatives.

In binary classification, the most straightforward metric is overall accuracy, which represents the proportion of all samples that are correctly classified. Its expression is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

The error rate represents the proportion of all samples that are misclassified, and its expression is:

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \tag{10}$$

Sensitivity represents the proportion of all actually positive samples that are correctly predicted as positive by the model, and its expression is:

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

In industrial inspection systems, real-time performance and model training time are important indicators for assessing model quality. To this end, this paper introduces training time and detection time as evaluation metrics to compare different methods.

3.6.3. Model Results and Comparison

To demonstrate the advantages of the proposed method, comparison experiments were conducted using 1D CNN, MobileNetV3, EfficientNet-B0, and ResNet18. The results are presented in Table 3, the specific detection times are listed in Table 4, and the corresponding hyperparameters are listed in Table 5. The 1D CNN uses the full signal as input, while the other models use the same input as the proposed method. Since the model node initialization time (~1500 ms) is incurred only at first startup, the effective per-inspection time in continuous operation is the complete inspection cycle time minus this initialization time. In industrial fault detection, the primary task is to identify faulty parts, and the most critical requirement is to prevent faulty parts from being misclassified as normal ones. Sensitivity reflects the classification accuracy of faulty parts and is a crucial metric for industrial application. The experimental results show that the EfficientNet-B0 model not only achieves the highest overall accuracy but also yields a Sensitivity significantly higher than other models. The confusion matrix of the EfficientNet-B0 model (where NG is the positive class) is as follows: TP = 39, FN = 1, TN = 40, FP = 0. The Wilson 95% confidence interval for its overall accuracy is [93.3%, 99.8%]. Overall, the preprocessing, inference, and effective per-inspection times of EfficientNet-B0 remain within an acceptable range, and it effectively avoids missed detections, demonstrating high potential for industrial deployment.

Table 3. Binary OK/NG classification results.

Model	Training Time (min)	Detection Time (ms)	Accuracy (%)	Error Rate (%)	Sensitivity (%)
1D-CNN	163	5319	83.75	16.25	67.50
ResNet18	79	8299	87.50	12.50	80.00
MobileNetV3	68	8211	93.75	6.25	90.00
EfficientNet-B0	138	8205	98.75	1.25	97.50

Table 4. Detailed statistics of detection time (ms).

Model	Data Preprocessing Time	Model Node Initialization Time	Inference Time	Complete Inspection Cycle Time
1D-CNN	0	1478	3841	5319
ResNet18	824	1516	5959	8299
MobileNetV3	845	1499	5867	8211
EfficientNet-B0	820	1491	5894	8205

Table 5. Optimal hyperparameters of the model.

Model	lr	batch_size	optimizer_name
1D-CNN	0.002661901888489057	8	“Adam”
ResNet18	0.002661901888489057	8	“Adam”
MobileNetV3	0.002661901888489057	8	“Adam”
EfficientNet-B0	0.00017277890583771574	8	“Adam”

3.7. Ablation Study

An ablation study was conducted based on the EfficientNet-B0 model. The number of channels was changed to 1 and 3, respectively, and hyperparameter search was re-performed for models with different channel numbers using the training method described in this paper to obtain the optimal hyperparameter combinations. These were then compared with the 4-channel input, and the results are shown in Table 6. Using triaxial vibration data and microphone data as the 4-channel input provides a relatively good improvement in accuracy.

Table 6. Ablation study results.

Channel	Test Set Accuracy (%)	lr	batch_size	optimizer_name
[mic,X,Y,Z]	98.75	0.00017277890583771574	8	“Adam”
[X]	80.00	0.0001329291894316216	8	“Adam”
[Y]	83.75	0.00011494747628961817	16	“Adam”
[Z]	81.25	0.002661901888489057	8	“Adam”
[mic]	90.00	0.0001329291894316216	8	“Adam”
[X,Y,Z]	97.50	0.00017277890583771574	8	“Adam”

4. Discussion

The results show that the proposed acoustic-vibration fusion framework, combined with EfficientNet-B0, provides an effective solution for NVH-based fault detection of electric scroll compressors. Under joint low- and high-speed training, the model achieved 98.75% accuracy on the independent test set, while training was completed within 3 h and a single online detection took less than 9 s. These findings support the central hypothesis of this work: because fault-induced information in electric compressors is distributed across both acoustic and vibrational modalities, a multimodal model can learn more discriminative

representations than a single-modality model. The ablation results indicate that sound and vibration are complementary rather than redundant. Acoustic signals are sensitive to high-frequency aerodynamic or electromagnetic anomalies, while vibration signals provide mechanical impact, resonance, and directional information. Their fusion therefore improves robustness under complex operating conditions.

From the perspective of previous studies, traditional signal-processing methods such as FFT, HHT, VMD, EMD, and wavelet packet transform can extract fault features under controlled conditions, but their performance degrades in high-noise environments and when multiple sources are coupled. Deep learning models such as CNN, ResNet, GAN, VAE, Transformer, and LSTM have addressed some of these limitations, but many studies still focus on bearing or gearbox datasets rather than electric scroll compressors. Acoustic-vibration fusion has been extensively investigated for rolling bearings and gearboxes; however, its application to online NVH inspection of electric scroll compressors under combined high- and low-speed conditions remains limited. The present study extends this line of research by integrating MFCC-based acoustic features and STFT-based vibration features into a four-channel representation and by using EfficientNet-B0 for binary OK/NG classification. The comparison with 1D-CNN, ResNet18, and MobileNetV3 further supports this design. EfficientNet-B0 achieved the highest accuracy (98.75%), while MobileNetV3 achieved 93.75%, ResNet18 87.50%, and 1D-CNN 83.75%. Although MobileNetV3 had a shorter training time and 1D-CNN had a shorter detection time, their lower accuracy would increase the risk of missed or false rejections in industrial inspection. The compound scaling strategy and SE attention in EfficientNet-B0 appear to be beneficial for learning discriminative time-frequency patterns from the fused input. From a diagnostic-interpretation perspective, preserving fault-induced transient patterns is important for linking model decisions to physical failure mechanisms. Recent work on burst-informed acoustic emission diagnosis has shown that explicitly separating burst-type transients from background noise and extracting burst-energy features can connect diagnostic decisions to crack initiation and propagation in milling [43]. While the present study relies on airborne microphones and accelerometers rather than acoustic emission sensors—and thus captures different physical quantities and frequency ranges—the principle that transient preservation supports interpretable diagnosis is relevant. Future work could adapt similar burst-aware or attribution-based methods to identify which time-frequency patterns in the fused representation drive the OK/NG decision.

The practical implications of this work are also noteworthy. The proposed system does not merely evaluate a model on a static dataset; it integrates product clamping, sensor mounting, real-condition simulation, raw signal acquisition, feature extraction, and online diagnosis into a complete NVH inspection platform. The air circuit system simulates pressure buildup under low- and high-speed conditions, while the soundproof enclosure reduces external interference. This system-level design is important because industrial deployment requires repeatable measurement conditions, low detection latency, and robustness to product-to-product variation. The use of product-level splitting in cross-validation and the removal of augmented validation samples also reduce the risk of data leakage, making the reported accuracy more credible than a random sample-level split.

Nevertheless, several limitations should be acknowledged. First, the dataset contains 48 faulty and 83 normal units, which is moderate in size and may not cover the full range of manufacturing and operational variability. Second, the study focuses on binary classification (OK/NG) rather than identifying specific fault types, severities, or root causes. Third, the experiments were conducted in a controlled soundproof enclosure, so the acoustic environment differs from that of a real vehicle or production line. Fourth, the detection time is acceptable for the current inspection process but may need further optimization for

higher-throughput lines. Fifth, the model was validated on one electric scroll compressor configuration, and its transferability to other compressor models or operating conditions remains to be tested. In addition, bilinear interpolation may smooth transient components and potentially introduce condition-dependent patterns. The present system is designed for process-level screening, and the model operates on the overall time-frequency pattern rather than on precise transient durations.

Future work can proceed in several directions. First, multi-class fault diagnosis should be investigated to distinguish different fault modes and severity levels. Second, transfer learning and domain adaptation could be used to improve generalization across compressor models, sensor placements, and noise environments. Third, lightweight and edge-deployable versions of the model should be developed to further reduce detection latency and hardware cost. Fourth, semi-supervised or self-supervised learning may help leverage large amounts of unlabeled operational data. Fifth, interpretability methods such as attention visualization or time-frequency attribution could provide insight into which acoustic and vibration patterns drive the diagnosis. Sixth, validation on real vehicle air-conditioning systems and long-term online monitoring would help move the proposed method from a laboratory prototype toward industrial deployment. Seventh, the 1D-CNN, ResNet18, MobileNetV3, and EfficientNet-B0 models used in this study are general-purpose classifiers. They were selected because industrial deployment requires lightweight models that can run in real time at the inspection station. Future work will consider a direct comparison with recent acoustic-vibration fusion architectures. Eighth, targeted ablations will be conducted, including MFCC-only versus STFT-only versus fused input, late fusion versus early fusion, common-frequency-axis representation versus channel stacking, different resizing strategies versus fixed-length windowing, different backbones under identical inputs, and with and without Optuna-based tuning. Ninth, If the scope is extended from screening to longitudinal degradation monitoring and prognosis, recent work on physics-guided degradation state-space modeling [44], trustworthy multistep-ahead RUL prediction [45], and physics-guided degradation trajectory modeling [46] provides relevant methodological references for physically consistent and uncertainty-aware health assessment. These studies explicitly constrain degradation evolution to be monotonic and irreversible, quantify uncertainty to support risk-aware decisions, and embed physical knowledge into model construction to avoid black-box predictions. However, such frameworks address degradation trajectory modeling and remaining-useful-life prediction, whereas the present work focuses on binary discrimination; the two tasks differ in output form, time scale, and validation requirements. Tenth, future work will address the statistical limitations of the current evaluation by reporting product-level accuracy with bootstrap confidence intervals clustered by product, the mean and standard deviation of test accuracy over repeated training runs, and a formal rule for combining the two speed-specific predictions into a single inspection decision. Expanding the test set will require accumulating additional faulty units over a longer production period. In addition, we plan to compare the proposed early-fusion construction against independent modality encoders with late fusion and against representations that map both modalities onto a common frequency axis so that the contribution of each fusion level can be isolated more rigorously.

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

Funding: This research was funded by the Central Government Guidance Fund for Local Science and Technology Development, grant number 2025ZY02004.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank all members of the research team for their assistance with the experiments and data collection, and the reviewers for their constructive comments. We are also grateful to China Jiliang University and Hangzhou Wolei Intelligent Co., Ltd. for providing the experimental conditions.

Conflicts of Interest: Author Liping Chen, Xinxin Tan and Haotong Fan was employed by the company Hangzhou Wolei Intelligent Technology 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:

NVH	Noise, Vibration, and Harshness
MFCC	Mel-frequency cepstral coefficients
STFT	Short-time Fourier transform
CNN	Convolutional neural network
1D-CNN	One-dimensional convolutional neural network
ResNet	Residual network
GAN	Generative adversarial network
VAE	Variational autoencoder
LSTM	Long short-term memory
VMD	Variational mode decomposition
EMD	Empirical mode decomposition
FFT	Fast Fourier transform
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
PCA	Principal component analysis
SVM	Support vector machine

References

1. Solichin, M.; Young-Jin, K.; Yoojeong, N.; Sunhwa, P.; Byeongha, A. Stable Hybrid Feature Selection Method for Compressor Fault Diagnosis. *IEEE Access* **2021**, *9*, 97415–97429. [CrossRef]
2. Tang, J.; Cheng, X.; Sun, J.; Qing, J.; Luo, P.; Hu, S. A novel method for untrained detection of compound fault in rolling bearing via fast Fourier Transform-Transformer model. *Measurement* **2025**, *253*, 117755. [CrossRef]
3. Feng, T.; Wang, Z.; Qiu, L.; Li, H.; Wang, Z. Rolling Based on Multi-Source Time-Frequency Feature Fusion with a Wavelet-Convolution, Channel-Attention-Residual Network-Bearing Fault Diagnosis Method. *Sensors* **2025**, *25*, 4091. [CrossRef] [PubMed]
4. Yan, Y.; Li, Z.; Li, S.; Wu, Y.; Liu, H.; Tao, Y. VAF-IBFD: A VMD-Assisted approach to resolve FFT limitations in intelligent bearing fault diagnosis. *Digit. Signal Process.* **2026**, *169*, 105713. [CrossRef]
5. Li, J.; Zhan, S.; Guo, H.; Wu, Z. A Multi-sensor Cross scale Time-Frequency Network for intelligent gear fault diagnosis in noisy environment. *Eng. Appl. Artif. Intell.* **2026**, *164*, 113292. [CrossRef]
6. Liang, Y.; Zhang, J.; Shi, Z.; Zhao, H.; Wang, Y.; Xing, Y.; Zhang, X.; Wang, Y.; Zhu, H. A Fault Identification Method of Hybrid HVDC System Based on Wavelet Packet Energy Spectrum and CNN. *Electronics* **2024**, *13*, 2788. [CrossRef]
7. Dhakar, A.; Singh, B.; Gupta, P. Diagnosing Faults in Rolling Bearings of an Air Compressor Set Up Using Local Mean Decomposition and Support Vector Machine Algorithm. *J. Vib. Eng. Technol.* **2024**, *12*, 6635–6648. [CrossRef]
8. Cheng, Z.; Lu, Y. Diagnosis and analysis of abnormal noise in the pure electric vehicle's air condition compressor at idle. *J. Low Freq. Noise Vib. Act. Control* **2018**, *37*, 711–724. [CrossRef]
9. Hou, J.; Lu, X.; Zhong, Y.; He, W.; Zhao, D.; Zhou, F. A comprehensive review of mechanical fault diagnosis methods based on convolutional neural network. *J. Vibroeng.* **2023**, *26*, 44–65. [CrossRef]

10. Wen, L.; Li, X.; Gao, L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Comput. Appl.* **2020**, *32*, 6111–6124. [CrossRef]
11. Chen, K.; Xu, G.; Zhang, Y.; Wang, Y. Small-Sample Thermal Fault Diagnosis Using Knowledge Graph and Generative Adversarial Networks. *Electronics* **2026**, *15*, 355. [CrossRef]
12. Ibrahim, R.; Zemouri, R.; Tahan, A.; Kedjar, B.; Merkhouf, A.; Al-Haddad, K. Fault Detection Based on Vibration Measurements and Variational Autoencoder-Desirability Function. *IEEE Open J. Ind. Appl.* **2024**, *5*, 106–116. [CrossRef]
13. Fu, Z.; Liu, Z.; Ping, S.; Li, W.; Liu, J. TRA-ACGAN: A motor bearing fault diagnosis model based on an auxiliary classifier generative adversarial network and transformer network. *ISA Trans.* **2024**, *149*, 381–393. [CrossRef] [PubMed]
14. Kibrete, F.; Woldemichael, D.; Gebremedhen, H. Fault Diagnosis of Rotating Machines Based on Combination of One-Dimensional Convolutional Neural Network and Long Short-Term Memory in Variable Working Conditions. *J. Eng.* **2025**, *2025*, 1670810. [CrossRef]
15. Jin, Y.; Yang, J.; Yang, X.; Liu, Z. Cross-domain bearing fault diagnosis method based on SMOTENC and deep transfer learning under imbalanced data. *Meas. Sci. Technol.* **2024**, *35*, 015121. [CrossRef]
16. Zhang, X.; Liu, S.; Li, L.; Lei, J.; Chang, G. Multiscale holospectrum convolutional neural network-based fault diagnosis of rolling bearings with variable operating conditions. *Meas. Sci. Technol.* **2021**, *32*, 105027. [CrossRef]
17. Pietrzak, P.; Wolkiewicz, M. Fault Diagnosis of PMSM Stator Winding Based on Continuous Wavelet Transform Analysis of Stator Phase Current Signal and Selected Artificial Intelligence Techniques. *Electronics* **2023**, *12*, 1543. [CrossRef]
18. Irfan, M.; Khan, N.; Mushtaq, Z.; Kareri, T.; Abbas, G.; Mursal, S. A light-weight factorized convolutions based dual-input fuzzy-CNN for efficient motor bearing fault diagnosis. *Nondestruct. Test. Eval.* **2025**, *40*, 5418–5454. [CrossRef]
19. Nakamura, H.; Asano, K.; Usuda, S.; Mizuno, Y. A Diagnosis Method of Bearing and Stator Fault in Motor Using Rotating Sound Based on Deep Learning. *Energies* **2021**, *14*, 1319. [CrossRef]
20. Shirdel, S.; Teimoortashloo, M.; Mohammadiun, M.; Gharahbagh, A. A hybrid method based on deep learning and ensemble learning for induction motor fault detection using sound signals. *Multimed. Tools Appl.* **2024**, *83*, 54311–54329. [CrossRef]
21. Lu, S.; Wang, X.; He, Q.; Liu, F.; Liu, Y. Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals. *J. Sound Vib.* **2016**, *385*, 16–32. [CrossRef]
22. Salazar-Villanueva, F.; Ibarra-Manzano, O. Spectral Analysis for Identifying Faults in Induction Motors by Means of Sound. In Proceedings of the CONIELECOMP 2013, 23rd International Conference on Electronics, Communications and Computing, Cholula, Mexico, 11–13 March 2013.
23. Zastępa, M. The statistical-based diagnosis with usage of acoustic sound decomposition and projected LSTM network of induction motors. *Eksploat. Niezawodn.–Maint. Reliab.* **2025**, *27*, 205651. [CrossRef]
24. Amarnath, M.; Sugumaran, V.; Kumar, H. Exploiting sound signals for fault diagnosis of bearings using decision tree. *Measurement* **2013**, *46*, 1250–1256. [CrossRef]
25. Ahvazi, P.; Mohammadi, H. A High-Precision voiceprint recognition for fault diagnosis of motor bearings through lightweight dynamic convolutional neural network. *Results Eng.* **2026**, *29*, 108463. [CrossRef]
26. Fang, X.; Zheng, J.; Jiang, B. A rolling bearing fault diagnosis method based on vibro-acoustic data fusion and fast Fourier transform (FFT). *Int. J. Data Sci. Anal.* **2025**, *20*, 2377–2386. [CrossRef]
27. Wu, G.; Yan, N.; Choi, K.; Jung, H.; Cao, K. A two-step vibration-sound signal fusion method for weak fault feature detection in rolling bearing systems. *Adv. Mech. Eng.* **2021**, *13*, 1–15. [CrossRef]
28. Li, C.; Sanchez, R.; Zurita, G.; Cerrada, M.; Cabrera, D.; Vásquez, R. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mech. Syst. Signal Process.* **2016**, *76–77*, 283–293. [CrossRef]
29. Gu, X.; Tian, Y.; Li, C.; Wei, Y.; Li, D. Improved SE-ResNet Acoustic-Vibration Fusion for Rolling Bearing Composite Fault Diagnosis. *Appl. Sci.* **2024**, *14*, 2182. [CrossRef]
30. Xu, Z.; Chen, X.; Li, Y.; Xu, J. Hybrid Multimodal Feature Fusion with Multi-Sensor for Bearing Fault Diagnosis. *Sensors* **2024**, *24*, 1792. [CrossRef] [PubMed]
31. Wan, H.; Gu, X.; Yang, S.; Fu, Y. A Sound and Vibration Fusion Method for Fault Diagnosis of Rolling Bearings under Speed-Varying Conditions. *Sensors* **2023**, *23*, 3130. [CrossRef] [PubMed]
32. Hartono, D.; Halim, D.; Roberts, G. Gear fault diagnosis using the general linear chirplet transform with vibration and acoustic measurements. *J. Low Freq. Noise Vib. Act. Control* **2019**, *38*, 36–52. [CrossRef]
33. Dai, M.; Jo, H.; Kim, M.; Ban, S. MSFF-Net: Multi-Sensor Frequency-Domain Feature Fusion Network with Lightweight 1D CNN for Bearing Fault Diagnosis. *Sensors* **2025**, *25*, 4348. [CrossRef] [PubMed]
34. Li, X.; Wang, Y.; Wang, Z.; Li, Y. A lightweight multi-scale time-frequency fusion model for motor fault diagnosis. *Eng. Res. Express* **2025**, *7*, 045245. [CrossRef]
35. Zhang, Y.; Peng, B.; Zhang, P.; Sun, J.; Liao, Z. Key Technologies and Application of Electric Scroll Compressors: A Review. *Energies* **2024**, *17*, 1790. [CrossRef]

36. Shin, H.; Choi, J.; Cho, H.; Jang, S. Analysis on Electromagnetic Vibration Source Permanent Magnet Synchronous Motor for Compressor of Electric Vehicles. In Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC), Seoul, Republic of Korea, 9–12 October 2012.
37. Khanchoul, M.; Elumalai, I.; Gurudevan, V.; Guitari, I.; Krebs, G. High efficiency permanent magnet synchronous motor for an automotive electrical air conditioning compressor. In Proceedings of the 20th IEEE Vehicle Power and Propulsion Conference (IEEE VPPC), Milan, Italy, 24–27 October 2023. [CrossRef]
38. Smith, W.; Randall, R. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–131. [CrossRef]
39. Wang, Y.; Sheng, D.; Hu, H.; Han, K.; Zhou, J.; Hou, L. A Novel Series Arc Fault Detection Method Based on Mel-Frequency Cepstral Coefficients and Fully Connected Neural Network. *IEEE Access* **2022**, *10*, 97983–97994. [CrossRef]
40. Abdul, Z.; Al-Talabani, A. Mel Frequency Cepstral Coefficient and its Applications: A Review. *IEEE Access* **2022**, *10*, 122136–122158. [CrossRef]
41. Ayvaz, U.; Gürüler, H.; Khan, F.; Ahmed, N.; Whangbo, T.; Bobomirzaevich, A. Automatic Speaker Recognition Using Mel-Frequency Cepstral Coefficients Through Machine Learning. *CMC-Comput. Mater. Contin.* **2022**, *71*, 5511–5521. [CrossRef]
42. Yao, Q.; Wang, Y.; Yang, Y.; Shi, Y. Seal call recognition based on general regression neural network using Mel-frequency cepstrum coefficient features. *Eurasip J. Adv. Signal Process.* **2023**, *2023*, 1–20. [CrossRef]
43. Umar, M.; Siddique, M.; Kim, J. Burst-informed acoustic emission framework for explainable failure diagnosis in milling machines. *Eng. Fail. Anal.* **2026**, *185*, 110373. [CrossRef]
44. Umar, M.; Saleem, F.; Kim, J. Failure-consistent degradation state-space modeling of tool wear evolution and remaining useful life in machining systems. *Eng. Fail. Anal.* **2026**, *190*, 110751. [CrossRef]
45. Yin, C.; Sun, T.; Wu, H.; Dong, Y. Trustworthy multistep-ahead remaining useful life prediction for rolling bearings with limited data. *Reliab. Eng. Syst. Saf.* **2026**, *267*, 111902. [CrossRef]
46. Yin, C.; Li, Y.; Wang, Y.; Dong, Y. Physics-guided degradation trajectory modeling for remaining useful life prediction of rolling bearings. *Mech. Syst. Signal Process.* **2025**, *224*, 112192. [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.