

Article

Research on a Fault-Diagnosis Method for Heavy-Duty Bearings of Thin Coal-Seam Shearers

Minghao Li ^{1,2,*}, Shuting Wang ^{1,*}, Xiao-Guang Zhang ^{3,4}, Xiaoxu Yi ¹ and Dongsheng Wu ^{1,2} 

¹ School of Mechanical Engineering, Shenyang Ligong University, Shenyang 110159, China; yiyiyixiaoxu@163.com (X.Y.); wuds@sylu.edu.cn (D.W.)

² Science and Technology Development Corporation, Shenyang Ligong University, Shenyang 110159, China

³ School of Intelligent Manufacturing and Future Technologies, Fuyao University of Science and Technology, Fuzhou 350109, China; zhangxiaoguang@ustc.edu.cn

⁴ Key Laboratory of Future Intelligent Manufacturing Technologies for High-End Equipment, Fuyao University of Science and Technology, Fuzhou 350109, China

* Correspondence: lmh0315@sylu.edu.cn (M.L.); 2401610041@stu.sylu.edu.cn (S.W.);

Tel.: +86-159-4020-7840 (M.L.)

Abstract

Aiming at the problems of fault samples being difficult to obtain in fault diagnosis research on heavy-duty bearings of thin-seam shearers and the insufficient diagnostic performance of existing methods under small-sample conditions, this study focuses on dataset construction, sample augmentation and fault diagnosis. First, based on the virtual prototype model of the cutting-unit transmission system of a thin-seam shearer, three-dimensional models of healthy bearings and four typical fault types (inner ring fault, outer ring fault, rolling-element fault and cage fault) of heavy-duty bearings were built using SolidWorks 2024. Vibration signals were collected through ADAMS dynamic simulation and converted into time-frequency images via Continuous Wavelet Transform (CWT), thereby constructing an original fault dataset with five states. Furthermore, a conditional generative adversarial network incorporating VGG perceptual loss (VGG-CGAN) was proposed to achieve targeted sample augmentation for the five bearing states, effectively alleviating the class-imbalance problem. On this basis, an improved ResNet50 fault-diagnosis model was constructed, and Bayesian optimization was used to automatically tune key hyperparameters. Experimental results show that the improved ResNet50 model achieved an accuracy of 87.14% on the self-built thin-seam shearer heavy-duty bearing dataset and 98.28% on the public CWRU dataset. The proposed method exhibits strong diagnostic performance and generalization ability under small-sample and imbalanced data conditions. This study can provide new ideas and useful references for fault diagnosis of heavy-duty bearings in thin-seam shearers.

Keywords: thin coal seam; cylindrical roller bearing; heavy-duty bearing; fault diagnosis; small sample



Academic Editor: Shiyuan Wang

Received: 20 May 2026

Revised: 9 July 2026

Accepted: 16 July 2026

Published: 19 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In intelligent, fully-mechanized mining systems for thin coal seams, the shearer is the core equipment for coal cutting and loading. The transmission system of the shearer cutting unit is the most fundamental power transmission component. Compared with other components, the heavy-duty bearings in the cutting unit bear larger loads and more complex stresses, and the front cutting unit operates under particularly harsh conditions.

As thin-seam shearers operate under heavy load for long periods in narrow, high-dust, and high-humidity environments, the heavy-duty bearings in the front cutting unit, as key rotating components, are subjected to complex, alternating impact loads and extremely high contact stresses. Under the coupled effects of continuously strong vibration, frequent start-stop cycles, and coal-rock media intrusion, bearings are prone to typical faults such as fatigue spalling, crack propagation, and aggravated wear, directly affecting the reliability and mining efficiency of the shearer. Heavy-duty bearings have high maintenance costs, and once a fault occurs, it rapidly spreads through the transmission chain, causing equipment shutdowns, production interruptions, or even transmission-system collapse and serious safety accidents. Therefore, fault diagnosis of heavy-duty bearings in the transmission system of thin-seam shearers directly affects production efficiency; reduces energy consumption during operation; extends service life; enhances the level of mine intelligence; and ensures miner safety, which is of great significance.

Vibration signals are the most important information source for fault diagnosis of shearer bearings and gearboxes because they are sensitive to mechanical state changes and contain abundant fault information. Liu Xunan et al. established common fault models of broken teeth, cracked gears, and defective bearings using PRO/E software and successfully identified faulty components and fault types in the cutting-unit transmission system through dynamic simulation analysis [1]. Deng Yuxu proposed a fault-diagnosis method combining wavelet analysis and an improved BP neural network. By decomposing and reconstructing vibration signals using wavelet packets and optimizing the BP network with an additional momentum method and adaptive learning rate, diagnostic accuracy was significantly improved, and computation time was notably reduced [2]. Xu et al. combined Ensemble Empirical Mode Decomposition (EEMD) with an improved Harris Hawks Optimization (HHO) algorithm for the denoising of shearer vibration signals. After decomposing signals into intrinsic mode functions via EEMD, the improved HHO algorithm adaptively selected effective modes for reconstruction, effectively enhancing the signal-to-noise ratio of vibration signals under strong noise and improving the identifiability of fault features [3]. Zhang Xuhui et al. pre-trained a deep residual network using a large amount of labeled data from the source domain, then transferred feature extraction capability to the target domain via model-parameter migration and finally fine-tuned the network with only a small number of target-domain fault samples to achieve high-precision fault diagnosis of rocker-arm bearings under small-sample conditions [4]. Feng Yinhui et al. focused on the problem of extremely few fault samples and severe class imbalance in shearer anomaly detection, designing an anomaly detection method based on an autoencoding Siamese neural network, which significantly improved the accuracy and robustness of time-series anomaly detection [5].

Most faults in thin-seam shearers occur in the gear-transmission part of the cutting unit and the hydraulic-transmission part of the traction unit [6]. As key components of the cutting unit, fault diagnosis and identification of these bearings are among the key points for real-time online diagnosis. Sun Mingbo et al. proposed a fault-diagnosis model based on HGWO-MSVM for feature-vector extraction of shearer rolling bearings, which significantly improved the accuracy and efficiency of shearer bearing fault diagnosis [7]. Quan Baochao et al. deeply studied the applicability of sparse decomposition methods for rolling bearing fault diagnosis of shearer rocker arms [8]. Cao Xiangang et al. proposed a fault-diagnosis model based on vibration images combined with dynamic convolutional neural networks, which can efficiently and accurately classify different fault forms of shearer rolling bearings [9]. Cheng Liang et al. proposed a novel fault-diagnosis method for rolling bearings in shearer cutting units based on CEEMD-RSVPSO-SVM, achieving high accuracy and convergence speed [10]. Yang Feiwen proposed a method applying momen-

tum contrast tuning to shearer bearing fault diagnosis, which was tested on six publicly available rotating machinery fault datasets, providing strong support for the improvement of the accuracy of shearer bearing fault prediction [11]. Qin F Y et al. proposed a method for extracting shearer bearing faults through ADAMS simulation and HHT feature extraction, providing valuable theoretical support for field applications of shearer fault diagnosis [12]. Liu et al. further extended the research from fault diagnosis to remaining-useful-life (RUL) prediction, proposing a graph learning method based on variational graph structural information bottleneck (GL-VGSIB), which adaptively learns the graph structural dependencies of multi-sensor signals, thereby effectively improving the accuracy of equipment RUL prediction [13]. Guo Jinhui proposed a bearing fault diagnosis method for shearer cutting units based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) and an improved Support Vector Machine (SVM), achieving high classification accuracy in practical engineering applications [14]. In addition, in the field of transfer learning, Liu et al. proposed an interpretable domain adaptation Transformer (IDAT) model, which leverages the global information-capturing capability of the Transformer architecture to address the insufficiency of feature representation in traditional methods for cross-domain fault diagnosis, and enhances model interpretability through an attention-weight mechanism, achieving excellent performance in various cross-condition bearing fault-diagnosis tasks [15].

In summary, scholars worldwide have conducted extensive research on fault diagnosis of shearer transmission components, covering various technical routes, such as signal processing, machine learning, deep learning, and transfer learning, and have achieved good results in laboratory environments and specific industrial scenarios. However, existing research still has the following shortcomings: First, most studies focus on general rolling bearings or rocker-arm bearings, while research on fault diagnosis of dedicated heavy-duty bearings for thin-seam shearer cutting units, especially four-row cylindrical roller bearings, is rare. Second, most methods rely on sufficient and balanced fault samples, making them difficult to directly apply to actual working conditions of thin-seam shearers, where fault samples are scarce and class distribution is severely imbalanced.

In view of this, this paper focuses on heavy-duty bearings of thin-seam shearer cutting units, conducting research on simulation data construction, small-sample augmentation, and imbalanced diagnosis. A conditional generative adversarial network incorporating VGG perceptual loss (VGG-CGAN) is proposed to achieve high-quality, targeted sample augmentation of the original time-frequency image dataset of five states constructed via dynamic simulation. For small-sample and imbalanced tasks, an improved ResNet50 model is designed, combined with Bayesian optimization for automatic hyperparameter tuning. This study can provide a theoretical basis and valuable reference for fault diagnosis of heavy-duty bearings in thin-seam shearers.

The research content of this paper is summarized as follows. Section 2 constructs an original fault dataset of heavy-duty thin-seam shearer bearings based on dynamic simulation and continuous wavelet transform. Section 3 proposes a conditional generative adversarial network incorporating VGG perceptual loss (VGG-CGAN) for targeted dataset augmentation, alleviating the sample-imbalance problem. Section 4 builds an improved ResNet50 fault-diagnosis model, performs hyperparameter tuning via Bayesian optimization, and verifies the diagnostic performance and generalization ability under small-sample conditions. Finally, the research achievements and main conclusions are summarized.

2. Construction of Original Dataset for Heavy-Duty Bearings of a Thin-Seam Shearer Cutting Unit

2.1. Establishment of Virtual Prototype Model of Thin-Seam Shearer Cutting Unit

To achieve accurate dynamic simulation of the thin-seam shearer cutting unit (transmission system), this study adopts contact modeling theory based on the impact function method to construct a dynamic interaction model between transmission-system components. In this model, the contact stiffness coefficient is a key physical parameter determining simulation accuracy, and its theoretical calculation relies on the Hertz contact theory framework. The specific derivation process is outlined as follows:

According to the geometric relationship at the contact point of the gear pair, the equivalent radius of curvature is calculated as follows:

$$\frac{1}{R} = \frac{1}{R_1} \pm \frac{1}{R_2} \quad (1)$$

The equivalent elastic modulus is calculated by integrating the elastic properties of the gear materials:

$$\frac{1}{E} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \quad (2)$$

Based on Hertz contact theory, the analytical expression for the base stiffness (k) is derived as follows [16]:

$$K = \frac{4}{3} R^{\frac{1}{2}} E \quad (3)$$

where R_1 and R_2 are radius values from the gear involute contact point to the gear center in mm, E_1 and E_2 are the elastic moduli of the two gear materials in MPa, and ν_1 and ν_2 are the Poisson ratios of the corresponding materials.

The plus sign (+) corresponds to external gear meshing, while the minus sign (−) applies to internal gear meshing. By adding constraints and driving with a motor, a multi-coupled virtual prototype model of the thin-seam shearer cutting unit is established, as shown in Figure 1.

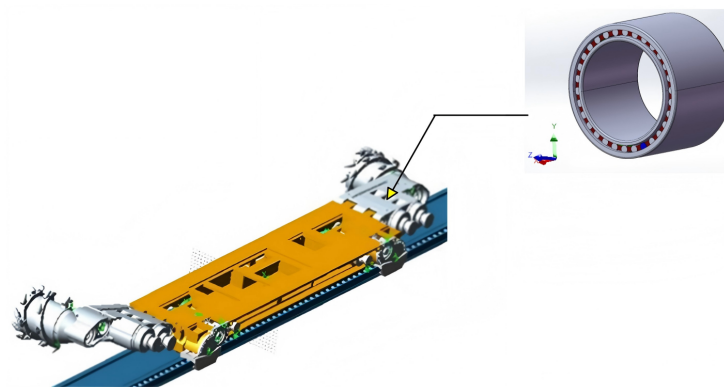


Figure 1. Multi-coupled virtual prototype model of the shearer.

2.2. Establishment of Simulation Model for Heavy-Duty Bearings of a Cutting Unit

Thin-seam shearers are key equipment for the mining of coal seams with a thickness below 1.3 m. Due to the extreme narrowness of the working space, the overall structure must be highly compact with high power density. The cutting unit, as the core component directly breaking coal and rock, bears huge impact loads, vibrations, and bending moments. The performance and reliability of its main shaft bearing directly determine the coal-mining efficiency and service life of the entire machine. The four-row cylindrical roller bearing is a classic solution designed to cope with such extreme working conditions. Its multi-row

parallel load-carrying structure distributes the total load into four independent load-bearing units, achieving increased load capacity and structural rigidity. The bearing structure is shown in Figure 2.

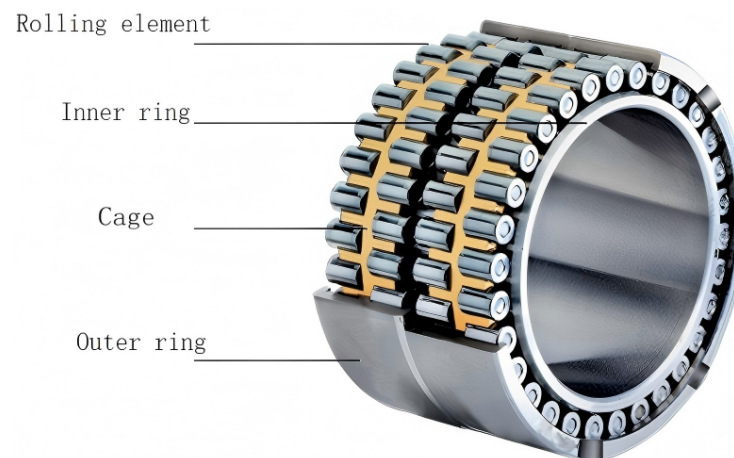


Figure 2. Structure of the heavy-duty bearing in the cutting unit.

This study focuses on the model 250RY1681/250 × 340 × 230 FC-type four-row cylindrical roller bearing, with an inner diameter of 250 mm, outer diameter of 340 mm, and width of 230 mm, belonging to the wide series. The specific structural parameters are listed in Table 1. SolidWorks 3D modeling software was used to establish the 3D model of heavy-duty bearing of the thin-seam shearer. During modeling, the Z-axis is the axial direction, and the X and Y axes are radial directions, ensuring accurate dimensions and correct assembly relationships. The final 3D model of the healthy heavy-duty bearing of the thin-seam shearer was established.

Table 1. Specific parameters of the 250RY1681 four-row cylindrical roller bearing.

Parameter	Value	Parameter	Value
Bore diameter	250 mm	Bearing width	230 mm
Outer diameter	340 mm	Inner raceway diameter	276 mm
Roller diameter	18.9 mm	Outer raceway diameter	314 mm
Radial clearance	0.1 mm	Axial clearance	0.2 mm
Pitch diameter	292.5 mm	Contact angle	0°
Number of rollers	30 × 4	Raceway width	42.5 mm

According to relevant research, four-row bearings often fail under harsh conditions due to lubrication failure, foreign-object intrusion, misalignment, etc. [17]. The main failure modes include fatigue spalling, wear, plastic deformation, corrosion, gluing, fracture, and misalignment. Spalling is one of the most common and serious failure modes for four-row cylindrical roller bearings. Therefore, for inner ring, outer ring, and rolling element faults, spalling simulation is used for 3D modeling; cage faults are simulated by introducing local fractures on the cage structure.

The vibration-response characteristics of a bearing are governed by the relative proportion of the fault size to the key geometric parameters of the bearing. Accordingly, the fault dimensions in this study are determined by referring to the standard simulated fault sizes provided by the Case Western Reserve University (CWRU) bearing dataset [18]. Specifically, spalling faults are set as spherical pits with a diameter of 0.54 mm (0.021 in) and a depth of 0.26 mm on the corresponding component surface to simulate local spalling. The cage fault is defined as a fracture oriented at 120° relative to the axial bearing direction, with a fracture depth of 4.5 mm and a penetration of 5 mm.

In this study, four fault types are considered: outer race fault, inner race fault, roller fault and cage fault. The models and fault locations for each fault type are shown in Figure 3.

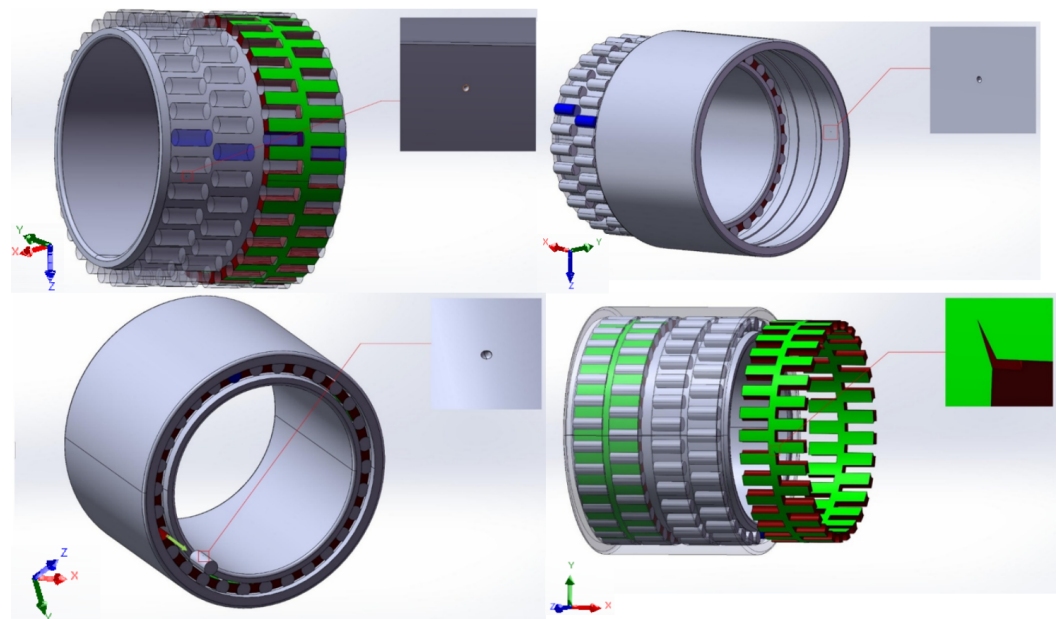


Figure 3. Demonstration of fault-location settings for bearing components.

2.3. Dynamic Analysis of a Heavy-Duty Cutting-Unit Bearing

Dynamic analysis of the heavy-duty bearing in the cutting unit is a key step to obtain its vibration signals. First, based on the actual working-condition data of the MG400/951 ((Jiamusi Coal Mine Machinery Co., Ltd., Jiamusi, Heilongjiang, China)) thin-seam shearer cutting unit, the working load of the target heavy-duty bearing is calculated. Then, the obtained load curve is used as input for dynamic simulation analysis to collect vibration signals under healthy and faulty conditions.

2.3.1. Load Calculation of the Heavy-Duty Bearing of a Cutting Unit

The main operating parameters of the MG400/951 thin-seam shearer cutting unit are a rotational speed (n) of 58 r/min and motor power of 400 KW. Based on the actual distribution of cutting-unit components, the dimensional parameters of each part were measured. Using the bending–torsional composite strength condition of the shaft, the load borne by the target bearing in actual operation was calculated. The load-distribution diagram of the shaft where the heavy-duty bearing is located was generated. All gears are spur gears, and the right spline is connected to the drum, as shown in Figure 4.

According to the specific parameters of the MG400/951 thin-seam shearer, the torque of gear 1 T_1 is calculated using Formula (4):

$$T_1 = 9550 \frac{P}{n} = \frac{400 \times 9550}{58} \approx 65862 \text{ N} \cdot \text{m} \quad (4)$$

where subscripts A and D correspond to Gear 1 and Gear 2 (as shown in Figure 4), respectively; F_{tA} and F_{tD} denote the tangential forces acting on Gear A and Gear D; and F_{rA} and F_{rD} denote the corresponding radial forces. The tangential force acts along the tangent to the pitch circle, while the radial force acts along the line connecting the gear centers.

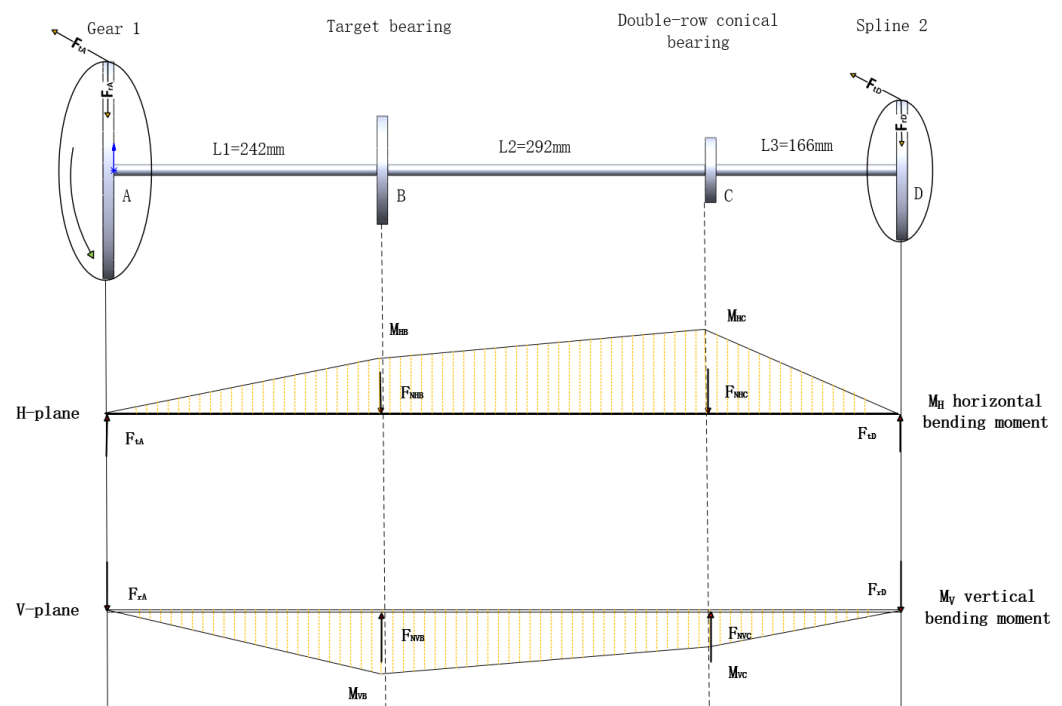


Figure 4. Load analysis diagram of the shaft with the target bearing.

According to the force formulas for spur gears, the radial force F_r and tangential force (F_t) of Gear 1 and Gear 2 are calculated:

$$F_t = \frac{2T_1}{d_1} \quad (5)$$

$$F_r = F_t \tan \alpha \quad (6)$$

where d_1 is the pitch-circle diameter of Gear 1 (406 mm) and α is the pressure angle (20°).

According to the above equations, the calculated results are $F_{tA} \approx 324,443$ N and $F_{rA} \approx 118,081$ N.

On the horizontal plane (H-plane), the bearing reaction forces are opposite to the direction of F_t . By taking moments about points B and C and writing equilibrium equations, the target bearing reaction force at point B (F_{NHB}) and the bearing reaction force at point C (F_{NHC}) can be obtained:

$$\begin{cases} \sum M_{HB} = 0 \Rightarrow F_{NHC} \cdot L_2 + F_{tA} L_1 = F_{tD} (L_2 + L_3) \\ \sum M_{HC} = 0 \Rightarrow F_{NHB} \cdot L_2 + F_{tD} L_2 = F_{tA} (L_1 + L_2) \end{cases} \quad (7)$$

where L_1 , L_2 , and L_3 are the lengths of lever arms for segments AB, BC, and CD (mm), the specific values of which are shown in Figure 4.

According to the above equations, the calculated result is expressed as follows:

$$\begin{cases} F_{NHC} = 1.568 F_{tD} - 2688887 \\ F_{NHB} = 5933330 - 0.3568 F_{tD} \end{cases}$$

The horizontal bending moments (M_{HB} and M_{HC}) are calculated as follows:

$$\begin{cases} M_{HB} = F_{tA} \cdot L_1 = 78515.206 \text{ N} \cdot \text{m} \\ M_{HC} = F_{tD} \cdot L_3 = 0.166 F_{tD} \end{cases} \quad (8)$$

On the vertical plane (V-plane), the bearing reaction forces (F_{NVB} and F_{NVC}) are opposite to the direction of F_r . By taking moments about points B and C and setting counterclockwise moments equal to clockwise moments, the target bearing reaction force at point B (F_{NVB}) and the bearing reaction force at point C (F_{NVC}) on the vertical plane are obtained:

$$\begin{cases} \sum M_{VB} = 0 \Rightarrow F_{NVC} \cdot L_2 + F_{rA}L_1 = F_{rD}(L_2 + L_3) \\ \sum M_{VC} = 0 \Rightarrow F_{NVB} \cdot L_2 + F_{rD}L_2 = F_{rA}(L_1 + L_2) \end{cases} \quad (9)$$

The calculated result is
$$\begin{cases} F_{NVC} = 1.568F_{rD} - 97866 \\ F_{NVB} = 215953 - 0.568F_{rD} \end{cases}.$$

The vertical bending moments (M_{VB} and M_{VC}) are calculated as follows:

$$\begin{cases} M_{VB} = F_{rA} \cdot L_1 = 28577.054 \text{ N} \cdot \text{m} \\ M_{VC} = F_{rD} \cdot L_3 = 0.166F_{rD} \end{cases} \quad (10)$$

Finally, the resultant bending moments (M_B and M_C) at points B and C are expressed as follows:

$$\begin{cases} M_B = \sqrt{M_{VB}^2 + M_{HB}^2} \approx 83554.087 \text{ N} \cdot \text{m} \\ M_C = \sqrt{M_{VC}^2 + M_{HC}^2} \end{cases} \quad (11)$$

2.3.2. Vibration Signal Acquisition Through Dynamic Simulation of a Heavy-Duty Bearing

The healthy heavy-duty bearing model and four fault-state models established in Section 2.2 were imported into ADAMS dynamic simulation software. According to the sampling theorem, the simulation step size was set to 0.001 s, and the sampling frequency was set to 1000 Hz to ensure alias-free signals. A vibration signal acquisition point was set on the bearing's inner ring, and the working load calculated in Section 2.3.1 was applied for dynamic simulation. During the simulation, radial vibration acceleration signals of the bearing under healthy and various fault states were recorded, with a simulation duration of 3 s. Partial time-domain waveforms of vibration signals are shown in Figures 5 and 6.

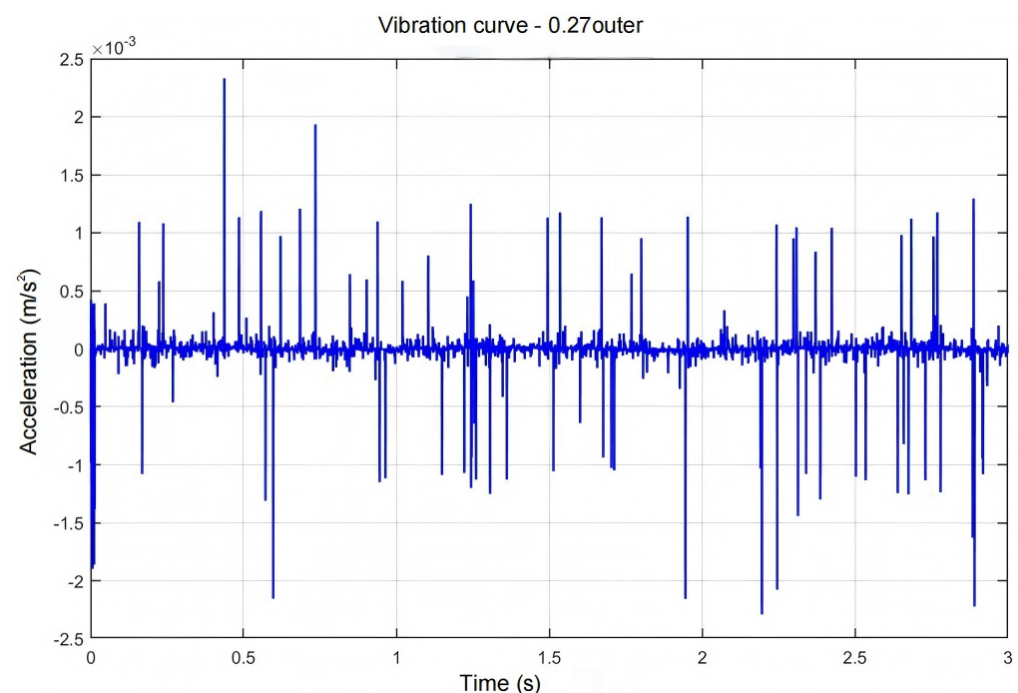


Figure 5. Vibration curve of a bearing with outer race fault.

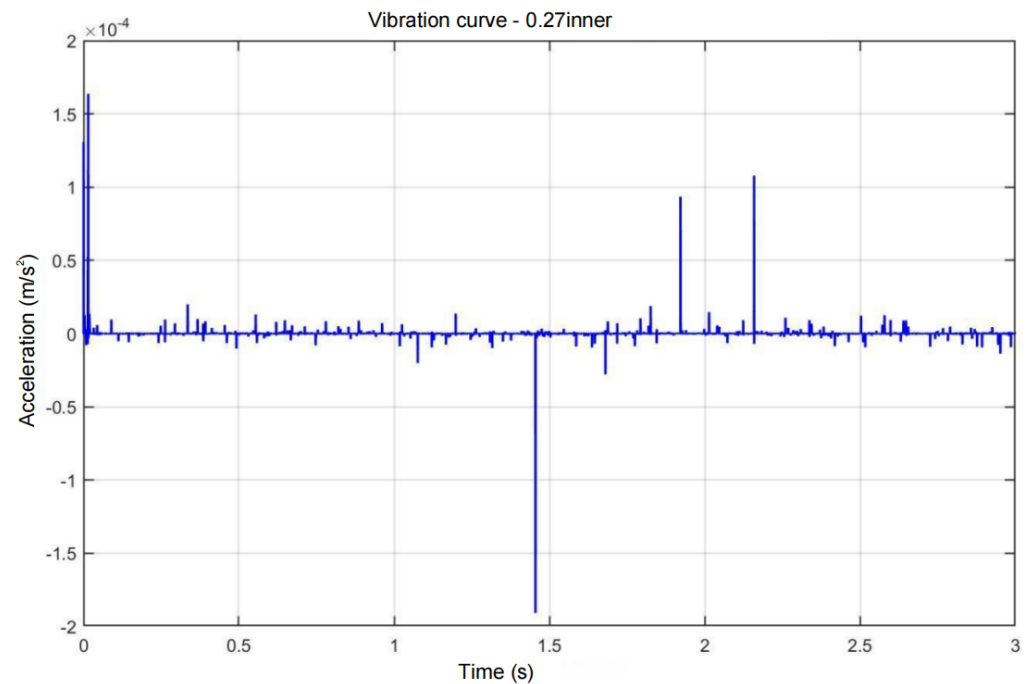


Figure 6. Vibration curve of a bearing with inner race fault.

2.4. Vibration Feature Extraction Based on Wavelet Transform

The vibration impact signals caused by shearer bearing faults are related to the fault location, and different fault locations exhibit different spectral characteristics of vibration signals. The key to shearer bearing fault diagnosis is the extraction of useful signals from complex signals and the identification of the characteristic fault frequencies of shearer bearings [19]. When a bearing fault occurs, periodic transient signals generate low-frequency vibrations, typically below 1 kHz. The fault frequencies of each component are related to the structural parameters and kinematic parameters of the bearing, which are summarized as follows:

Natural frequency of roller:

$$f_m = 0.212 \frac{Eg}{r\rho} \quad (12)$$

where R is the elastic modulus of the roller material (MPa), ρ is the density of the roller material (kg/m^3), and r is the roller radius (mm).

Shaft frequency of shearer bearing:

$$f_r = \frac{n}{60} \quad (13)$$

Outer race fault frequency of shearer bearing:

$$f_o = \frac{Z}{2} f_r \left(1 - \frac{d}{D} \cos \alpha \right) \quad (14)$$

Inner race fault frequency of shearer bearing:

$$f_i = \frac{Z}{2} f_r \left(1 + \frac{d}{D} \cos \alpha \right) \quad (15)$$

Roller fault frequency of shearer bearing:

$$f_b = \frac{1}{2} f_r \frac{D}{d} \left[1 - \left(\frac{d}{D} \right)^2 \cos^2 \alpha \right] \quad (16)$$

Cage fault frequency of shearer bearing:

$$f_c = \frac{Z}{2} f_r \left(1 - \frac{d}{D} \cos \alpha \right) \quad (17)$$

where n is the rotational speed of the shaft cooperating with the bearing (r/min), Z is the number of rollers, α is the contact angle, d is the roller diameter (mm), and D is the pitch-circle diameter of the rollers (mm).

To achieve accurate extraction and visualization of fault features of heavy-duty bearings of thin-seam shearers, this study uses the Continuous Wavelet Transform (CWT) method for time-frequency analysis of the vibration signals collected in Section 2.3.2. For the mother wavelet, the complex Morlet wavelet is adopted as the analysis basis function, with the following mathematical expression:

$$\varphi(t) = \pi^{-1/4} e^{i\omega_0 t} e^{-t^2/2} \quad (18)$$

where ω_0 is the wavelet center frequency, usually taken as 2π (rad/s).

For the target object of this study, the 250RY1681/250 × 340 × 230 four-row cylindrical roller bearing, substituting the specific parameters (Table 1) into the characteristic vibration Formulas (12)–(17) for faulty bearings yields the following fault frequencies:

Target bearing shaft frequency: $f_r \approx 0.9667$ Hz

Target bearing outer race fault frequency: $f_o \approx 13.562$ Hz

Target bearing inner race fault frequency: $f_i \approx 15.434$ Hz

Target bearing roller fault frequency: $f_b \approx 7.445$ Hz

Target bearing cage fault frequency: $f_c \approx 0.452$ Hz

In practical applications, the number of data points for wavelet transform is often set to an integer power of 2 to facilitate FFT calculation. Meanwhile, to ensure that each time-frequency image displays three to five complete impact cycles for effective fault-mode confirmation and avoidance of random noise misjudgment, three sets of parameters (window size of 1024 and step size of 256; window size of 512 and step size of 128; window size of 256 and step size of 64) were compared. Based on the generated original time-frequency image dataset, the dataset augmented by the generative adversarial network, and the final fault-diagnosis classification results, the final parameter settings are determined as shown in Table 2.

Table 2. Parameter selection for CWT.

Parameter Name	Set Value	Parameter Name	Set Value
Sampling frequency	1000 Hz	Step size	64
Window size	256	Wavelet center frequency	6 rad/s

Through the above processing, a complete original fault-diagnosis dataset for heavy-duty bearings of thin-seam shearer was constructed: 59 time-frequency images for the healthy state, 58 for inner race fault, 135 for outer race fault, 32 for rolling element fault, and 58 for cage fault. Some generated time-frequency images are shown in Figure 7.

Preliminary analysis of the generated time-frequency images reveals that the time-frequency energy distribution of healthy bearings is uniform, without obvious impact features; faulty bearings exhibit significant energy concentration at specific frequencies; and different fault types display unique “fingerprint” features on the time-frequency plane.

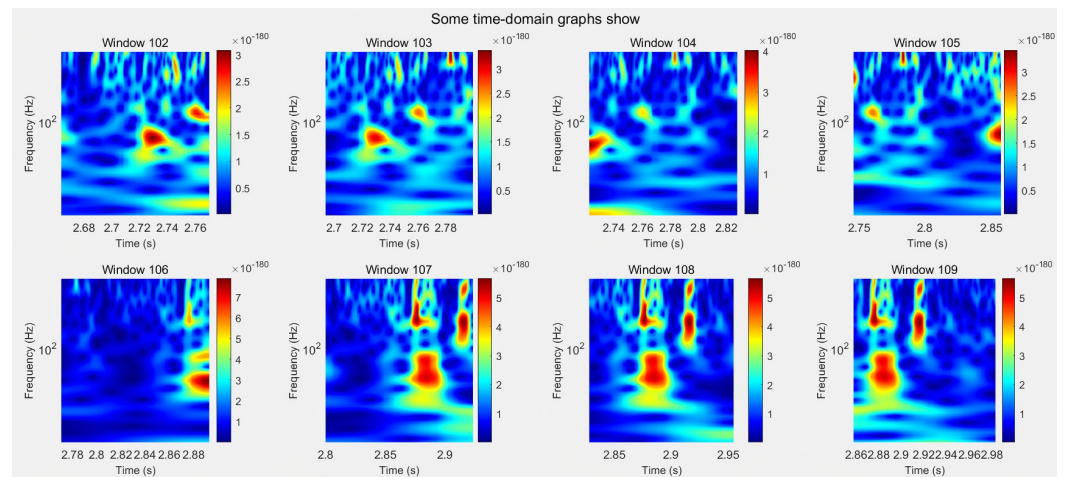


Figure 7. Time–frequency representation obtained via CWT.

3. Dataset Augmentation for Heavy-Duty Bearings of a Thin-Seam Shearer

3.1. Stratified Partitioning Strategy for Heavy-Duty Bearing Dataset

Following machine learning practices [20], this study adopts a stratified sampling strategy to partition the dataset: 60% for the training set, 20% for the validation set, and 20% for the test set, ensuring consistent class distribution across subsets. Data augmentation (including online augmentation) is applied only to the training set, while the validation and test sets remain unchanged to prevent data leakage and ensure objective evaluation and generalization. This strategy effectively alleviates the small-sample and class-imbalance issues of the original dataset, avoiding distribution bias caused by random partitioning.

3.2. VGG-CGAN Generative Adversarial Network

A dataset with balanced classes, diverse samples, and clear features is fundamental for the training of high-performance deep learning models. The achievement of controllable, targeted sample generation can effectively mitigate the impact of imbalanced samples in the original thin-seam shearer heavy-duty bearing dataset on fault diagnosis [21]. CGAN was chosen as the dataset augmentation network mainly because CGAN precisely controls the generated image class through conditional labels, ensuring sample realism while maintaining intra-class diversity, effectively enriching the training set and enhancing the feature robustness of subsequent diagnostic models; it converges faster and more stably under small-sample conditions, and the discriminator simultaneously performs authenticity discrimination and label matching, strengthening the semantic association between generated features and fault types.

CGAN consists of a generator and a discriminator: the generator input includes a random noise vector and class label, progressively upsampling through fully connected and transposed convolutional layers, using an improved residual structure to avoid gradient vanishing; the discriminator simultaneously judges the authenticity and class label of the input image, outputting the image-authenticity probability and class-probability distribution. The structures are shown in Figures 8 and 9.

To address the small-sample challenge in fault diagnosis of heavy-duty bearings of thin-seam shearers, this study adopts online data augmentation techniques, applying random transformations to input images during model training, providing diverse training samples for each epoch and effectively improving model generalization. The online data augmentation pipeline includes: a random horizontal flip ($p = 0.3$), random rotation (range $\pm 5^\circ$), color jitter (minor adjustments to brightness, contrast, saturation, and range $\pm 10\%$).

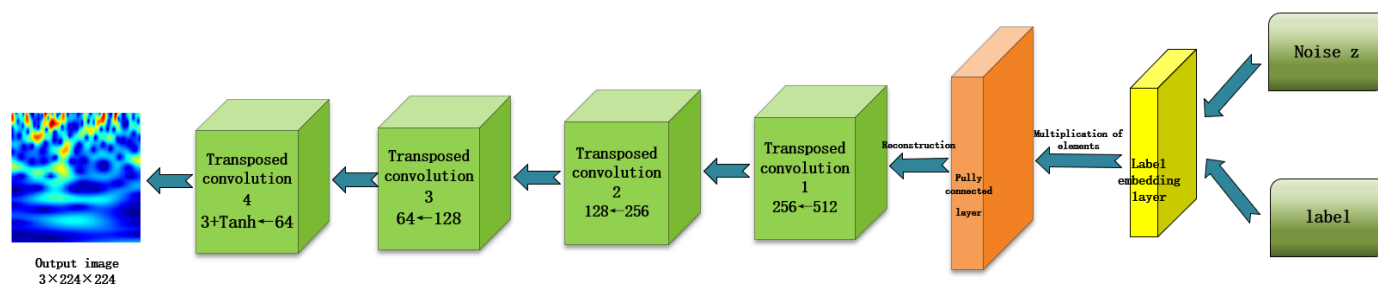


Figure 8. Architecture of the CGAN generator.

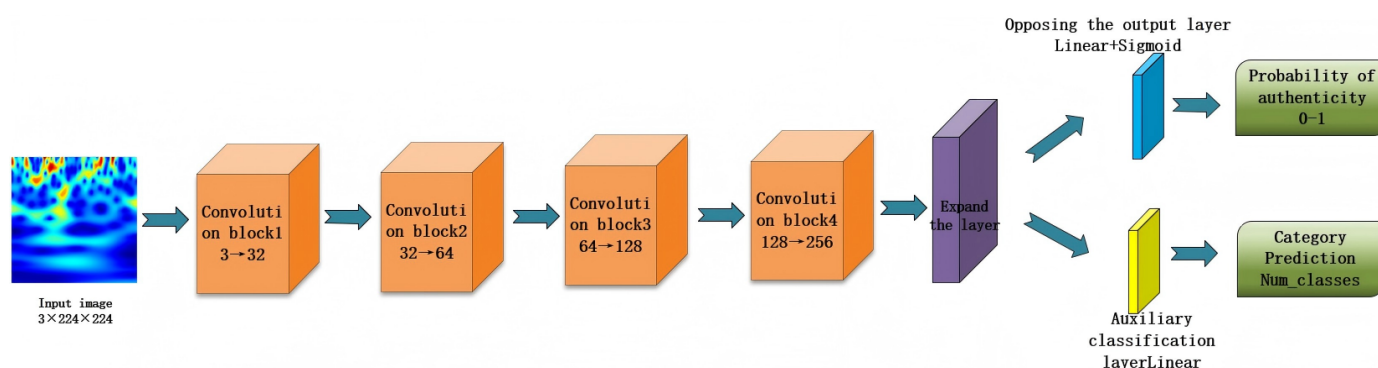


Figure 9. Architecture of the CGAN discriminator.

Traditional GANs adopt pixel-level loss functions, which tend to produce blurry time-frequency images and result in the loss of transient fault features. To address this limitation, this paper proposes a deep convolutional generative adversarial network integrated with Visual Geometry Group (VGG) perceptual loss (VGG-CGAN) to generate high-quality and diverse fault samples for further expansion of the training dataset. The introduction of VGG perceptual loss can effectively alleviate the above problem, which has been validated in bearing fault diagnosis [22]. The VGG network pre-trained on ImageNet has powerful feature extraction capabilities, capturing essential features of time-frequency images. Perceptual loss extracts high-level semantic features of images using a pre-trained deep neural network, computing the difference between generated and real images in feature space, guiding the generator to produce visually more realistic samples. This paper uses the first five convolutional layers of the VGG16 network as the feature extractor. The configuration of the first five layers of VGG16 is shown in Table 3.

Table 3. Parameter configuration for the first five layers of VGG16.

Layer No.	Module Type	Input Channels	Output Channels	Kernel Size	Output Size
1	Conv2d	3	64	3×3	$64 \times 64 \times 64$
2	ReLU	-	-	-	$64 \times 64 \times 64$
3	Conv2d	64	64	3×3	$64 \times 64 \times 64$
4	ReLU	-	-	-	$64 \times 64 \times 64$
5	Maxpool2d	-	-	2×2	$64 \times 32 \times 32$
6	Conv2d	64	128	3×3	$128 \times 32 \times 32$
7	ReLU	-	-	-	$128 \times 32 \times 32$

The overall structure of the VGG-CGAN proposed in this paper is shown in Figure 10, integrating the adversarial training mechanism of CGAN and VGG perceptual loss to form a multi-objective optimization framework. The generator and discriminator co-evolve through adversarial training, while the perceptual loss ensures that generated images are consistent with real images in visual features.

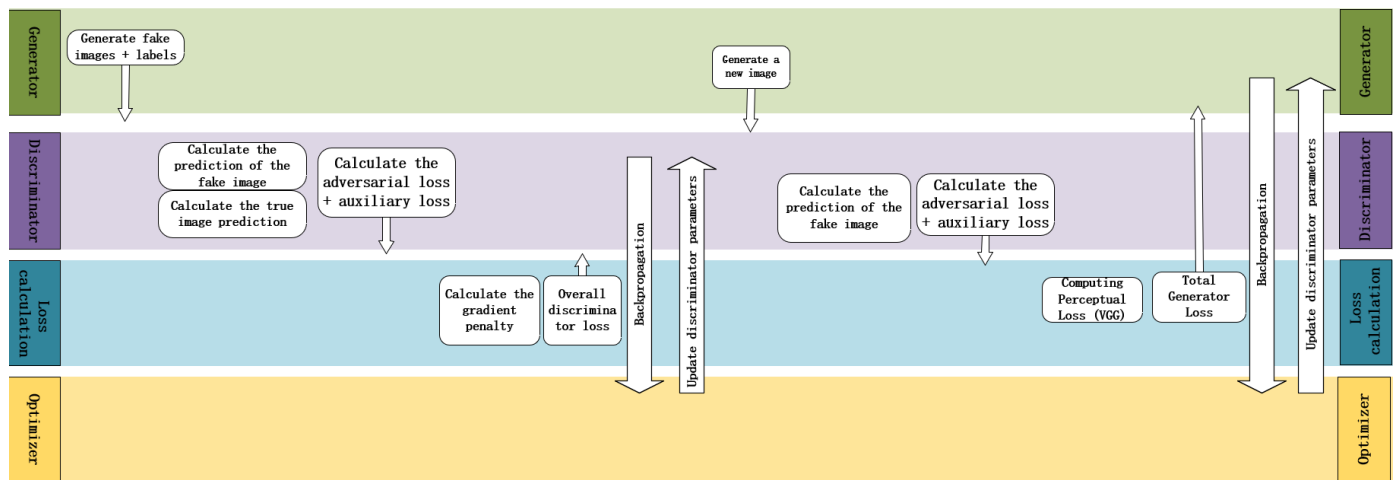


Figure 10. Schematic diagram of the integrated VGG-CGAN generative adversarial network.

The generator adopts a stable upsampling structure: the input is a random noise vector (z) and fault-class label), the latter of which is mapped to a dense vector via an embedding layer and multiplied element-wise by z for deep fusion of conditional information; the fused vector is progressively upsampled through fully connected and transposed convolutional layers, outputting a synthetic time-frequency image $G(z|y)$ with dimensions of $224 \times 224 \times 3$, whose distribution approximates the real image distribution for the corresponding class.

The discriminator is based on the WGAN-GP architecture, stabilizing training via gradient penalty. Its backbone network extracts hierarchical image features, outputting authenticity probability through a Sigmoid layer on one hand and class probability through a parallel branch with a fully connected layer on the other, achieving dual tasks of authenticity discrimination and auxiliary classification, forcing the discriminator to deeply understand fault types and providing more accurate gradient feedback to the generator.

To improve the visual quality and feature fidelity of generated images, perceptual loss based on VGG16 is introduced as an independent supervisory signal. This loss computes the L1/L2 distance between generated and real images in the high-level feature space of the pre-trained network, measuring semantic differences such as texture and structure rather than pixel-level comparison. By minimizing this loss, the generator is guided to synthesize time-frequency images that are visually realistic and have clear fault features.

During training, the generator aims to minimize adversarial loss and perceptual loss, while the discriminator aims to maximize its discriminative ability between real and generated images [23,24]. The final objective function is expressed as follows:

$$L_{\text{total}} = L_{\text{CGAN}} + \gamma L_{\text{perceptual}} \quad (19)$$

where $L_{\text{CGAN}} = L_{\text{adv}} + L_{\text{aux}}$ consists of the adversarial loss (L_{adv}) and the auxiliary classification loss (L_{aux}), $L_{\text{perceptual}}$ is the VGG-based perceptual loss, and $\gamma = 1 \times 10^{-4}$ is the corresponding weight coefficient.

In terms of training strategy, a cosine annealing learning-rate scheduler was chosen, and gradient clipping was added to prevent gradient explosion. Label smoothing (0.9, 0.1) was applied to the real and generated labels of the discriminator to improve generalization, along with an adaptive learning-rate adjustment mechanism that dynamically adjusts the learning rate based on image quality and the KID value.

3.3. Heavy-Duty Bearing Data Augmentation Based on VGG-CGAN

3.3.1. Hyperparameter Settings

To ensure the stability and generation quality of VGG-CGAN, key hyperparameters were determined through multiple experiments, as shown in Table 4. An adaptive learning-rate adjustment strategy was adopted: when the quality score of generated images did not improve for 15 consecutive epochs, the learning rates of both the generator and discriminator were multiplied by a decay factor of 0.7. This method helps the model escape local optima in later training stages and promotes finer convergence.

Table 4. Hyperparameter configuration of VGG-CGAN.

Hyperparameter	Set Value	Hyperparameter	Set Value
Latent space dimension	100	Batch size	8
Number of epochs	600	Optimizer	Adam
Generator learning rate	1×10^{-4}	Discriminator learning rate	8×10^{-5}
Gradient-clipping max norm	1.0	Perceptual loss weight	1×10^{-4}
Adam parameters (β_1 and β_2)	(0.5, 0.9)	Gradient penalty weight	1.0

3.3.2. Analysis of Generated Results

During training, qualitative and quantitative comprehensive evaluations of generated samples were conducted every 10 epochs, and loss anomaly detection was introduced: when NaN or loss >10.0 was encountered, the current batch was skipped, and gradients were cleared; when the epoch-average discriminator or generator loss exceeded 4.0, the learning rate was automatically decayed by 50%. These mechanisms ensured training stability. After training, performance was evaluated via loss curves, discriminator output curves, and KID curves, and the model with the best KID value was selected to generate 100 time-frequency images per class for addition to the training set, avoiding overfitting while augmenting samples. Finally, based on the original dataset, the performance-metric curves of dataset augmentation using VGG-CGAN are shown in Figure 11.

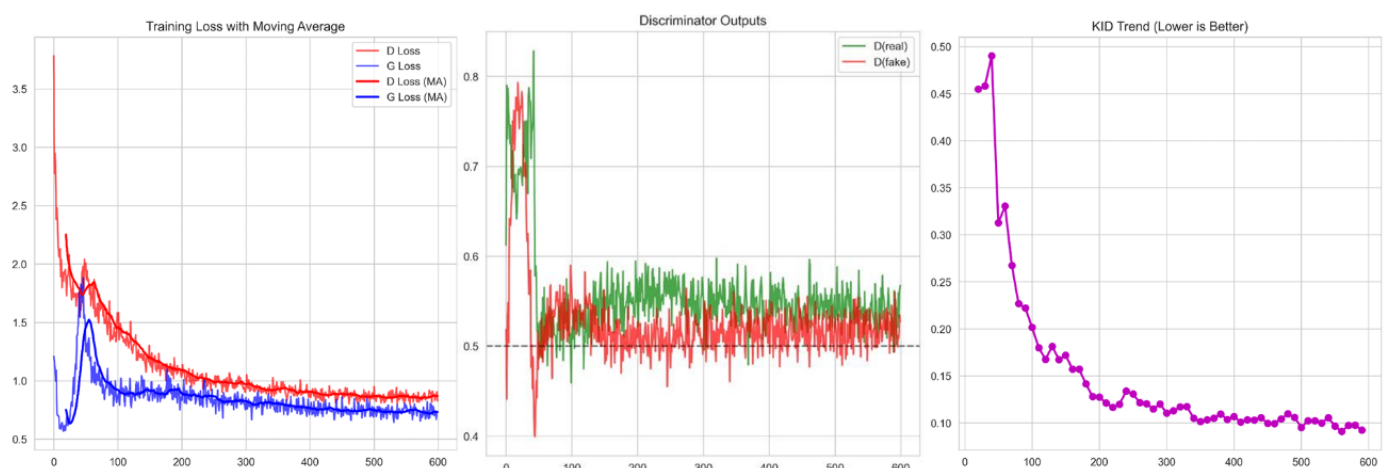


Figure 11. VGG-CGAN performance-metric curves.

The left training-loss curve in Figure 11 shows the loss curves of the generator and discriminator and their moving averages. It can be observed that after approximately 400 initial training epochs, both generator loss (G Loss) and discriminator loss (D Loss) enter a relatively stable oscillation range with similar values. This sustained, small-amplitude oscillation indicates that the GAN training reaches a dynamic equilibrium, ensuring the diversity of generated samples.

The middle image reflects the discriminator output for real and generated images. The discriminator output for real images ($D(\text{real})$) converges to the range of 0.55–0.60 after 50 epochs, while the output for generated images ($D(\text{fake})$) oscillates around the reference line of 0.5. The generator and discriminator basically reach Nash equilibrium, indicating that generated samples are in an ideal state [25]. The right KID trend graph shows that as training epochs increase, KID gradually decreases from 0.15 and ultimately converges to a trend between 0.02 and 0.04. The best KID value of 0.02 at epoch 570 indicates that the quality of time-frequency images generated by VGG-CGAN is high, reaching a level suitable for use as a training set for the next step of fault diagnosis of heavy-duty bearings of thin-seam shearers.

As shown in Figure 12, visual inspection further confirms that the generated images completely retain the typical time-frequency features of various faults, with rich texture details and intra-class diversity.

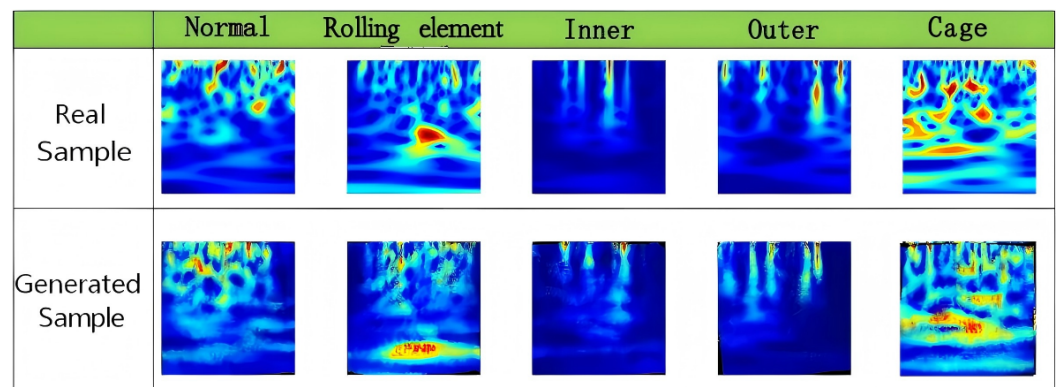


Figure 12. Comparison of real images and generated images from VGG-CGAN.

In summary, the VGG-CGAN training process is stable and effective, successfully generating training-set sample images with quality and diversity meeting the basic requirements for fault diagnosis of thin-seam shearers. The numbers of each bearing category after stratified sampling and VGG-CGAN augmentation of the training set are shown in Table 5.

Table 5. Final dataset class distribution.

Bearing Class	Training-Set Size	Test-Set Size	Validation-Set Size
Healthy	135	12	12
Rolling element fault	119	7	6
Inner race fault	134	12	12
Outer race fault	181	27	27
Cage fault	134	12	12

4. Research on a Fault-Diagnosis Method for Heavy-Duty Bearings of Thin-Seam Shearers Based on Improved ResNet50

4.1. Improved ResNet50 Fault-Diagnosis Model

4.1.1. Improved ResNet50 Neural Network Structure

Based on the specific requirements of the thin-seam shearer heavy-duty bearing fault diagnosis task (small-sample dataset and class imbalance), this paper makes the following targeted improvements to the pre-trained ResNet50 model provided in the MATLAB 2023b Deep Learning Toolbox as the base network structure. Main optimizations include the following:

Network-structure adjustment: Remove the fully connected layer (fc1000), softmax layer, and classification layer and replace them with a custom classification head more

suitable for the five-classification task (capable of N-classification, where N is the number of subfolders in the dataset).

Feature extraction enhancement: After the original global average pooling layer, two additional fully connected layers (512 and 256 dimensions) are added to improve feature expression capability.

Regularization optimization: Dropout layers and batch normalization layers are introduced to prevent overfitting and improve model generalization [26,27].

Class-imbalance handling: Class weights based on sample counts are introduced in the final classification layer, using a square-root smoothing strategy to avoid excessive bias towards minority-class samples, with the weight for class i calculated as follows:

$$w_i = \frac{\sqrt{1/N_i}}{\text{mean}_j(\sqrt{1/N_j})} \quad (20)$$

where N_i is the number of training samples for class i .

Activation-function optimization: The ReLU activation function is used after the newly added fully connected layers to maintain non-linear expression capability while alleviating gradient vanishing [28].

The overall structure of the improved ResNet50 is shown in Figure 13. The fault features of heavy-duty bearings of thin-seam shearers appear as texture patterns and energy distribution patterns at different scales in the time-frequency images. Local features such as edges and spots are captured by the shallow convolutional layers of ResNet50; periodic impacts, resonance bands, etc., are extracted by the intermediate residual blocks; the global distribution features of fault modes are integrated by the deep network and newly added fully connected layers.

4.1.2. Hyperparameter Settings Based on Bayesian Optimization

Based on multiple experimental validations and the specific requirements of the thin-seam shearer heavy-duty bearing fault diagnosis task, this paper determines the optimal hyperparameter configuration through four-parameter Bayesian optimization. The core idea of Bayesian optimization is to achieve efficient parameter search by constructing a surrogate model and an acquisition function [29]. The four-parameter Bayesian optimization framework designed in this study optimizes the parameters shown in Table 6. The optimal solution is found in the four-parameter space through 20 evaluations. Each evaluation selects the next evaluation point based on the current surrogate model and acquisition function, constructs the corresponding improved ResNet50 network, performs eight epochs of fast training, calculates classification accuracy on the validation set, and updates the surrogate Gaussian process model.

Table 6. Parameter settings of the Bayesian optimization framework.

Optimization Parameter	Search Range	Transform	Physical Meaning
Learning rate	$10^{-5} \sim 10^{-2}$	Logarithmic	Controls the model-parameter update step and affects training stability
L2 regularization coefficient	$10^{-6} \sim 10^{-2}$	Logarithmic	Controls model complexity and prevents overfitting
Dropout1	0.3~0.7	-	Dropout probability for first fully connected layer
Dropout2	0.2~0.6	-	Dropout probability for second fully connected layer

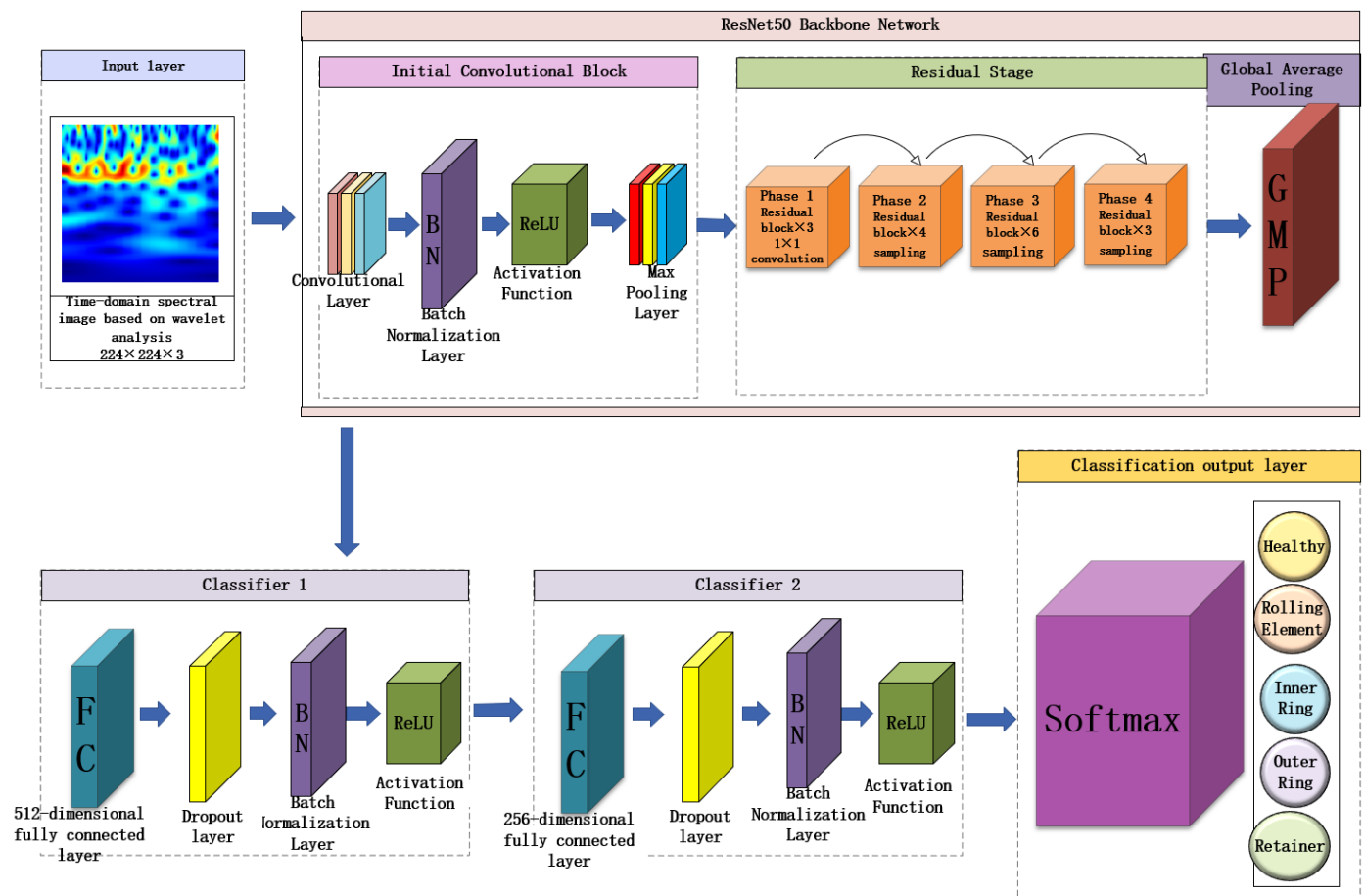


Figure 13. Architecture of the modified ResNet50 neural network.

Analysis results show that the learning rate has the greatest impact on model performance, with a correlation coefficient of about 0.65, and L2 regularization has a moderate impact, with a correlation coefficient of about 0.45, while dropout parameters have a relatively smaller impact, with a correlation coefficient of about 0.3. Based on the above Bayesian theory, the optimal parameter combination suitable for the improved ResNet50 fault-diagnosis network was obtained and applied to the complete training process.

The remaining hyperparameter settings for the improved ResNet50 network are shown in Table 7, also serving as the base surrogate model parameters for Bayesian optimization.

Table 7. Hyperparameter settings for the improved ResNet50 model.

Parameter Name	Set Value	Parameter Name	Parameter Value
Optimizer	Adam	Batch size	16
Initial learning rate	Optimized optimal solution	Early stopping patience	15 epochs
Learning-rate decay	Piecewise decay factor of 0.5	L2 regularization coefficient	Optimized optimal solution
Decay period	15 epochs	Dropout rate	Optimized optimal solution
Maximum training epochs	50	Gradient threshold	1.0

4.2. Verification Results and Analysis of Fault-Diagnosis Model for Heavy-Duty Bearings of Thin-Seam Shearers

During the operation of the improved ResNet50 network, the optimization process and results of automatic tuning of key hyperparameters by Bayesian optimization were comprehensively analyzed. After 20 iterations, the visualization results generated by the Bayesian optimization experiment provide a basis for hyperparameter selection and model performance evaluation for the target test set.

The convergence curve of Bayesian optimization (Figure 14) for the four hyperparameters (learning rate, L2 regularization coefficient, and dropout rates of the two fully connected layers) shows that as the number of trials increases, the validation accuracy shows an upward trend, increasing from an initial 65.22% to a best value of 92.75%, proving the effectiveness of the optimization strategy.

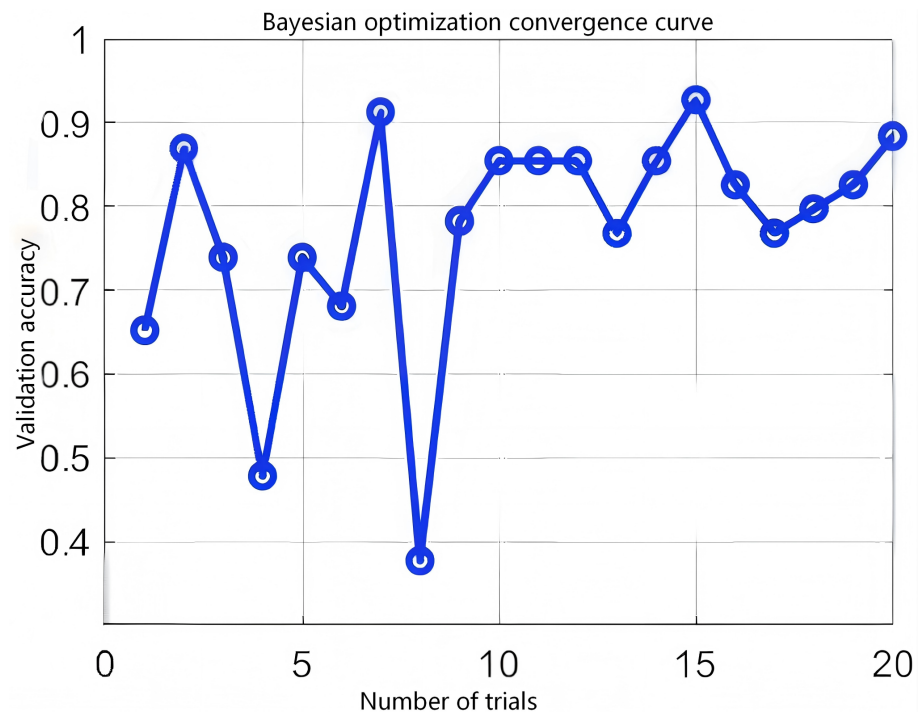


Figure 14. Bayesian optimization convergence curve.

On this basis, complete training was carried out using the optimal parameters. The model exhibited relatively stable convergence characteristics. During training, the validation accuracy reached a peak of 89.86% at epoch 36, triggering the early-stopping mechanism to control overfitting risk. The training loss curve shows that the mini-batch accuracy steadily increased from an initial 18.75% to 100%, and the validation loss decreased from 1.5421 to 0.3073.

Using the optimal hyperparameter combination identified by Bayesian optimization, we conducted multiple independent training runs and selected the model with the best generalization performance on the test set. Evaluation results on the independent test set (Figure 15) show that the improved ResNet50 model achieved an overall accuracy of 87.14% and a macro-average F1 score of 0.8536, verifying the effectiveness of the optimization strategy. Per-class performance analysis: rolling element fault (rolling element): precision of 1.000, recall of 0.714, and F1 score of 0.909—balanced and reliable; inner race fault (inner): precision of 1.000, recall of 0.818, F1 score of 0.889—good performance; healthy state (normal): precision of 1.000, recall of 0.833, F1 score of 0.909—balanced and reliable; outer race fault (outer): recall as high as 1.000 but precision of 0.818, and few false positives yet an F1 score of 0.900; cage fault (cage): recall of 1.000, precision of 0.800, and F1 score of 0.889—good performance. It should be pointed out that the recall of rolling element faults is relatively low, mainly restricted by the inherent statistical properties of small-sample datasets. As this type of fault has the fewest original samples, the generative model is unable to fully capture and fit its data distribution.

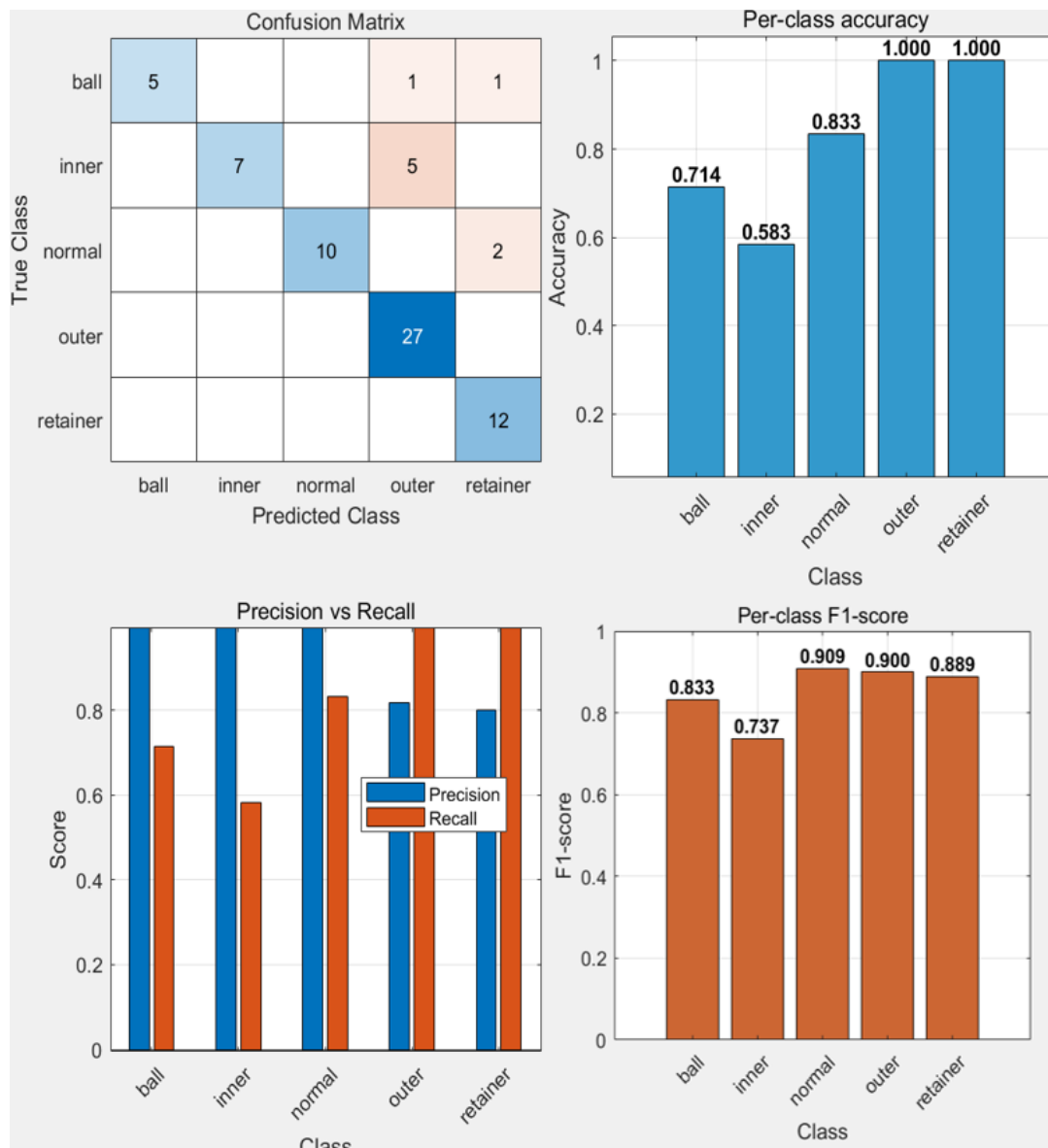


Figure 15. Performance analysis of the improved ResNet50 on the custom dataset.

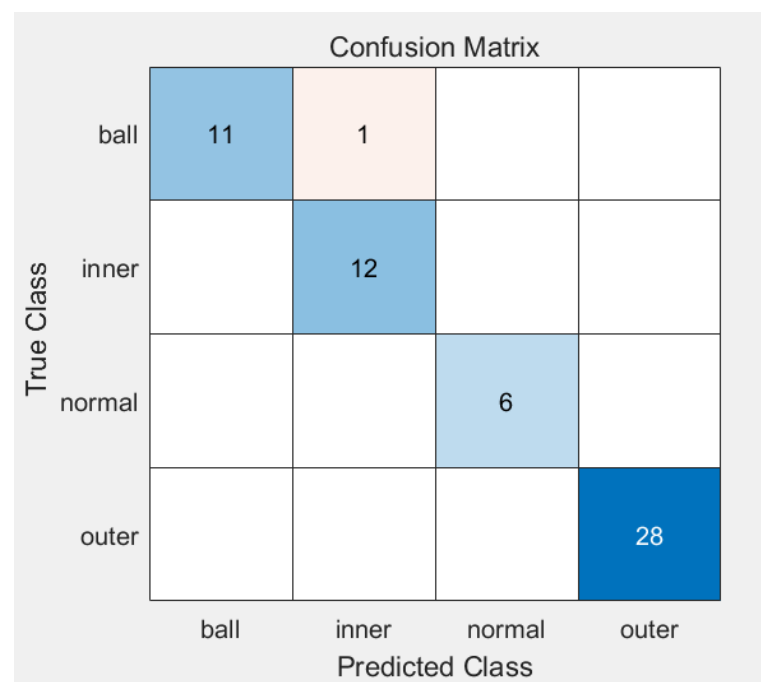
4.3. Performance Test of the Improved ResNet50 Fault-Diagnosis Model

To comprehensively verify the performance of the improved ResNet50 fault-diagnosis model proposed in this paper, comparative analysis was conducted using the public bearing fault dataset provided by the Bearing Data Center of Case Western Reserve University (CWRU) [30]. Vibration data of the drive-end bearing from the CWRU dataset were selected, with a motor load of 1 horsepower and a sampling frequency of 12 kHz. Four states (healthy, inner race fault, outer race fault, and rolling element fault) were selected (each fault type with a fault diameter of 0.021 inches, consistent with the self-built simulation dataset). One time-frequency image after wavelet transform was generated per 1024 data points. The image size was 224×224 , consistent with the self-built simulation dataset. The number of samples per bearing state was also similar to the self-built dataset, maintaining small-sample diagnosis and class-imbalance characteristics, as detailed in Table 8. Performance testing of the improved ResNet50 fault-diagnosis model achieved a test accuracy of 98.28% and a macro-average F1 score of 0.9791.

Table 8. Class distribution of the public dataset.

Bearing Class	Trainin-Set Size	Test-Set Size	Validation-Set Size
Healthy	36	6	6
Rolling element fault	134	12	12
Inner race fault	134	12	12
Outer race fault	181	28	28

Analysis of the visualization charts from the test results: Figure 16 shows the confusion matrix of the improved ResNet50 fault-diagnosis model on the CWRU test set. The values on the diagonal indicate the model's strong ability to correctly classify samples of each category, with healthy and outer race fault achieving perfect classification.

**Figure 16.** Confusion matrix for test results.

In the performance-analysis charts shown in Figure 17, the per-class accuracy distribution reflects the model's consistency in identifying different fault modes, with healthy and outer race fault accuracy reaching 1.000; F1-score comparison shows all classes have F1 scores above 0.95, indicating a good balance between precision and recall; the multi-class ROC curve shows all classes have AUC values above 0.98, confirming the model's excellent class-discrimination ability; the prediction confidence distribution shows correctly classified samples concentrated in the high-confidence interval, while misclassified samples are in the medium–low-confidence interval, reflecting the reliability of model predictions.

Comprehensive analysis of various metrics and visualization charts shows that the improved ResNet50 model exhibits balanced and excellent fault-diagnosis capability on the CWRU dataset, particularly demonstrating good adaptability and stability when handling fault modes with different characteristics.

Comparing the test results on the public dataset and the self-built thin-seam shearer heavy-duty bearing dataset, a comprehensive evaluation of the improved ResNet50 network's performance showed that overall accuracy reached 98.28% on the CWRU public dataset, which is 11.14 percentage points higher than on the self-built dataset. This difference is not due to the deficiency of the model itself but is mainly attributed to the inherent quality gap between the simulation data and the real data. Specifically, the self-built data

comes from the dynamic ADAMS simulation. The vibration signals are too monotonous and lack the inevitable background noise and random interference encountered in real working conditions, which makes the model prone to overfitting on ideal signals. Moreover, the simulation data only contains a single rotational speed and a constant load, lacking the characteristic diversity brought about by multiple working conditions and multiple fault sizes in the CWRU data, resulting in a high degree of overlap of the time-frequency patterns of different faults in the feature space, making it difficult for the model to effectively distinguish them; the macro-average F1 score was 0.9791 on CWRU and 0.8536 on the self-built dataset, showing a significant difference. Rolling element faults exhibited excellent classification performance on both datasets (F1 scores of 0.957 and 0.909 respectively), while inner race fault achieved relatively worse performance on the self-built dataset (F1 score of 0.889). The high accuracy of the improved ResNet50 on the standard CWRU dataset verifies its feature extraction capability based on architecture and Bayesian-optimized hyperparameters; the 87.14% accuracy on the self-built thin-seam shearer-bearing dataset indicates its generalizability to complex industrial vibration signals. After introducing class weights and a square-root smoothing strategy, the model performs reasonably on minority classes, demonstrating a good ability to handle class imbalance.

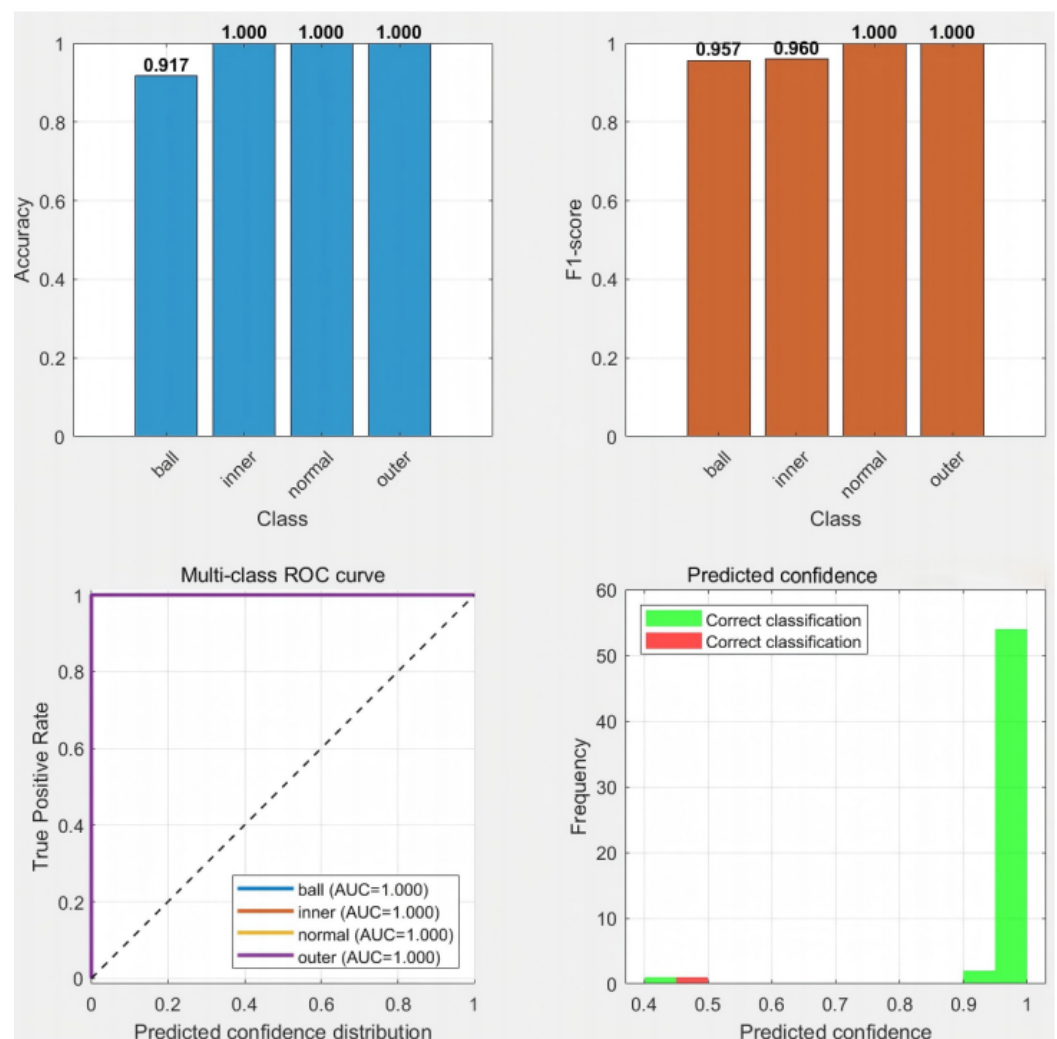


Figure 17. Performance analysis of the improved ResNet50 on the public dataset.

4.4. Comparative Evaluation of Improved ResNet50 Network on the Same Dataset

To evaluate the performance of the improved ResNet50 fault-diagnosis model, training results of three ResNet50 networks of varying complexity were compared based on the

previously constructed thin-seam shearer heavy-duty bearing dataset. For all three models, the reported results were selected from multiple independent training runs using the same configuration rather than relying on a single run in order to ensure the reliability of the comparison. One of them is the improved ResNet50 fault-diagnosis model constructed in Section 4.1.1 (Model-3).

The other two structures are described as follows:

Model-1: The pre-trained original ResNet50 network from MATLAB, with the original classification head (fc1000, etc.) removed, directly adding a fully connected layer for five-class output and applying class weights to mitigate sample imbalance. The learning rate is fixed at 0.0001, and the L2 regularization coefficient is fixed at 0.0002.

Model-2: The same network structure as Model-1 but using Bayesian optimization to automatically adjust the learning rate and L2 regularization coefficient.

The network structures of Model-1 and Model-2 are shown in Figure 18. Other parameter settings are the same as Model-3.

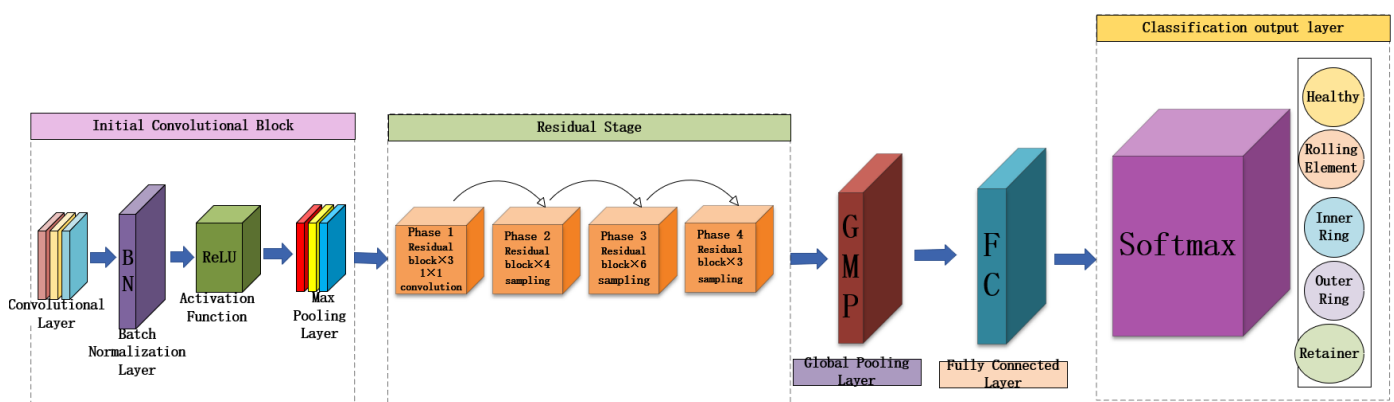


Figure 18. Architecture of the ResNet50 network.

To quantify the stability of each model after multiple rounds of training, the test-accuracy indicators of five independent repeated training sessions for different models were separately counted. The results are shown in Table 9.

Table 9. Statistical results of multiple repeated training experiments.

Test Accuracy	1	2	3	4	5
Model-1	0.7681	0.7826	0.7971	0.7714	0.7681
Model-2	0.7286	0.7162	0.7325	0.7371	0.7281
Model-3	0.8551	0.8986	0.8996	0.8714	0.8696

To further visually compare the fine-grained diagnostic performance of each model in various categories, five experiments from each model were selected from Table 9, and the run with the test accuracy closest to the mean value within the group was chosen (i.e., the median-performance experiment). The overall performance indicators were compared, and the following comparison table was drawn (Table 10).

To further verify the effectiveness of the VGG-CGAN data augmentation method proposed in this paper, based on the optimal diagnostic model (Model-3), three different data processing schemes were set up for comparative experiments. Results without data augmentation (A), results obtained using traditional online data augmentation (B), results obtained using standard CGAN enhancement (C), and the diagnostic accuracy results of the VGG-CGAN expansion method (Model-3) are presented as follows (Table 11):

Table 10. Performance comparison of three models on the thin-seam shearer heavy-duty bearing dataset.

Model	Test Accuracy	Macro-Average Precision	Macro-Average Recall	Macro-Average F1 Score
Model-1	0.7714	0.8485	0.7333	0.6938
Model-2	0.7286	0.8041	0.7270	0.7038
Model-3	0.8714	0.9236	0.8262	0.8536

Table 11. Ablation comparison of diagnosis performance with different data augmentation schemes.

Group	A	B	C	Model-3
Test Accuracy	/	0.5252	0.7159	0.8714

Based on the results of the three sets of control experiments, it can be concluded that the VGG-CGAN method proposed in this paper effectively enhances the authenticity and feature integrity of the augmented samples by introducing the pre-trained VGG16 perceptual loss to constrain the features of the generated samples. This method overcomes the shortcomings of traditional augmentation methods and achieves an average accuracy of 87.14% for fault diagnosis, which is significantly superior to the other three data processing schemes. This fully validates the effectiveness and advancement of this data augmentation method in the scenario of a small sample, category imbalance, and heavy-load bearing fault diagnosis.

The analysis based on the results of the previous model-comparison experiments indicates that the improved ResNet50 fault-diagnosis model outperforms the other two models in all performance metrics, with a test accuracy 10.00 percentage points higher than that of Model-1 and 14.28 percentage points higher than that of Model-2. In terms of macro-average F1 score, the improved model shows an improvement of about 16 percentage points over Model-1 and about 15 percentage points over Model-2. For the self-built thin-seam shearer heavy-duty bearing dataset, the improved ResNet50 network has stronger fault-diagnosis capability than the original ResNet50 neural network, making it more suitable for small-sample imbalanced datasets.

5. Conclusions

Fault diagnosis of heavy-duty bearings in thin-seam shearers is of great significance. This paper constructed a bearing fault dataset through virtual prototyping, dynamic simulation, vibration signal acquisition, and time-frequency transformation and proposed a small-sample diagnosis method combining data augmentation and an improved residual network. The main conclusions are outlined as follows:

(1) Using SolidWorks, 3D models of healthy bearings and four fault types (inner ring, outer ring, rolling element, and cage) were established. Vibration signals were obtained via ADAMS simulation, and 342 time-frequency images were generated using Continuous Wavelet Transform (CWT) constructing an original dataset of five states. The time-frequency images clearly present unique “fingerprint” features of different faults, providing high-quality input for subsequent deep learning.

(2) To address the challenges of an insufficient number of original samples and class imbalance, stratified sampling was used to ensure consistent training-/validation-/test-set distributions. A conditional generative adversarial network incorporating VGG perceptual loss (VGG-CGAN) was proposed, using high-level semantic features extracted by VGG as perceptual loss to guide the generator to synthesize realistic time-frequency images with prominent fault features. The model generated five classes of samples directionally

achieving a minimum KID value of 0.02 and generating 100 images per class augmentation, alleviating data imbalance to a certain extent and supporting subsequent diagnostic model training. To verify the superiority of the proposed amplification method, comparative multi-scheme ablation experiments were conducted based on the subsequent optimal diagnostic model, which fully demonstrated the effectiveness of the perception loss constraint mechanism in amplifying small-sample fault datasets.

(3) An improved ResNet50 fault-diagnosis model was proposed, reconstructing the classification head, enhancing feature extraction, and introducing square-root-smoothed class weights and dropout/BN regularization to improve adaptability to small-sample and imbalanced data. Bayesian optimization was used to automatically tune the learning rate, L2 coefficient, and dropout rates. Experiments showed that the improved model achieved accuracies of 87.14% on the self-built dataset and 98.28% on the public CWRU dataset, with AUC values above 0.85 for all classes and reliable prediction confidence. Compared with the original ResNet50 and its variants, the proposed improvement strategy has obvious advantages in classification performance and application value.

Author Contributions: Conceptualization, M.L., S.W. and X.Y.; formal analysis, M.L., X.-G.Z. and D.W.; funding acquisition, M.L.; investigation, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 51674134), the Fundamental Research Funds for the Universities of Liaoning Province (Grant No. LJJKQZ20222448), the Fundamental Research Funds for the Universities of Liaoning Province (Grant Nos. LJ212410144076 and LJ232410144074), and the Liaoning Province Science and Technology Plan Joint Program Project 2025 (2025-MSLH-589).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: We ensured that all individuals listed in this section have agreed to this acknowledgment. We thank the editors and the anonymous reviewers for their valuable comments and suggestions. Additionally, we confirm that no generative artificial intelligence (GenAI) tools or services were utilized throughout the preparation of this manuscript.

Conflicts of Interest: Authors Minghao Li and Dongsheng Wu were employed by Science and Technology Development Corporation, Shenyang Ligong University, Shenyang 110159, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Liu, X.; Zhao, L.; Fu, D.; Zhang, F. Research on wavelet packet decomposition method for fault signals of shearer cutting unit transmission system. *J. Vib. Shock* **2019**, *38*, 169–175. <https://doi.org/10.13465/j.cnki.jvs.2019.14.024>.
2. Deng, Y. Fault diagnosis of shearer cutting unit transmission system based on wavelet analysis and improved neural network. *Coal Mine Mach.* **2021**, *42*, 180–183. <https://doi.org/10.13436/j.mkjx.202106058>.
3. Xu, J.; Ren, C.; Liu, Y.; Chang, X. Noise Elimination for Coalcutting Vibration Signal Based on Ensemble Empirical Mode Decomposition and an Improved Harris Hawks Optimization Algorithm. *Symmetry* **2022**, *14*, 1978. <https://doi.org/10.3390/sym14101978>.
4. Zhang, X.; Pan, G.; Guo, H.; Mao, Q.; Fan, H.; Wang, X. Fault diagnosis method for rolling bearings in shearer rocker arm based on deep transfer learning. *Coal Sci. Technol.* **2022**, *50*, 256–263. <https://doi.org/10.13199/j.cnki.cst.2019-1425>.
5. Feng, Y.; Song, Y.; Li, W.; Wu, Y.; Qin, Z. Anomaly detection of shearer based on autoencoding siamese neural network. *Coal Eng.* **2022**, *54*, 136–141.
6. Li, X.; Li, S.H.; Chen, H.; Si, L.; Wei, D.; Zou, X.Y. Fault diagnosis of shearer cutting unit gearbox based on improved cascaded broad learning. *Ind. Mine Autom.* **2025**, *51*, 86–95. <https://doi.org/10.13272/j.issn.1671-251x.2024110057>.
7. Sun, M.; Ma, Q.; Zhang, Y.; Lei, J. Fault diagnosis method for shearer rolling bearing based on HGWO-MSVM. *Ind. Mine Autom.* **2018**, *44*, 81–86. <https://doi.org/10.13272/j.issn.1671-251x.2017110006>.

8. Quan, B.; Zhou, C.; Wang, C. Research on fault diagnosis method for shearer rocker arm bearing based on sparse representation. *Coal Mine Mach.* **2020**, *41*, 167–170. <https://doi.org/10.13436/j.mkjx.202001056>.
9. Cao, X.; Zhang, G.; Zhang, X.; Zhang, S. Fault diagnosis of shearer rolling bearing based on vibration image and DCNN. *Coal Mine Mach.* **2020**, *41*, 149–152. <https://doi.org/10.13436/j.mkjx.202007052>.
10. Cheng, L.; Zhang, B.; Zhang, J. Fault diagnosis of rolling bearings in shearer cutting unit based on CEEMD-RSVPSO-SVM. *Coal Technol.* **2022**, *41*, 179–182. <https://doi.org/10.13301/j.cnki.ct.2022.05.044>.
11. Yang, F. Application of momentum contrast tuning method in shearer bearing fault diagnosis. *Mach. Tool Hydraul.* **2024**, *52*, 213–220. <https://doi.org/10.3969/j.issn.1001-3881.2024.08.034>.
12. Qin, F.Y.; Fu, X.; Li, K.X.; Li, H.-J. ADAMS Simulation and HHT Feature Extraction Method for Bearing Faults of Coal Shearer. *Processes* **2024**, *12*, 164. <https://doi.org/10.3390/pr12010164>.
13. Liu, D.; Li, Y.; Xiao, Y.; Zhang, J.; Cui, L. Graph learning with variational graph structural information bottleneck for multi-sensor machinery remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2026**, *272*, 112465. <https://doi.org/10.1016/j.res.2026.112465>.
14. Guo, J. Fault diagnosis technology for shearer cutting unit bearings based on ICEEMDAN-NOA-SVM. *Coal Eng.* **2025**, *57*, 156–162. <https://doi.org/10.11799/ce202502022>.
15. Liu, D.; Cui, L.; Wang, G.; Cheng, W. Interpretable domain adaptation transformer: A transfer learning method for fault diagnosis of rotating machinery. *Struct. Health Monit.* **2025**, *24*, 1187–1200. <https://doi.org/10.1177/14759217241249656>.
16. Li, F.F.; Li, H.K.; Zhai, J.Y.; Yuan, M.Y.; Sun, B.; Dong, Z. Improved analytical model of gear mesh stiffness considering free end effect and contact gap variation. *Appl. Math. Model.* **2025**, *151*, 116460. <https://doi.org/10.1016/j.apm.2025.116460>.
17. Sharma, S. Comprehensive Analysis of Bearing Failure Modes: Mechanisms, Diagnostics, and Maintenance Strategies. *Int. J. Res. Appl. Sci. Eng. Technol.* **2025**, *13*, 90. <https://doi.org/10.22214/ijraset.2025.74468>.
18. Li, C.; Xiang, L.; Liu, Y.; Liao, Y. Simulation Analysis of Typical Faults of Double Row Tapered Roller Bearing for High Speed Locomotive Based on ADAMS. *Bearing* **2018**, *6*, 55–59. <https://doi.org/10.19533/j.issn1000-3762.2018.06.015>.
19. Wang, L.; Wu, M. Research on bearing fault diagnosis based on machine learning and SHAP interpretability analysis. *Sci. Rep.* **2025**, *15*, 41242. <https://doi.org/10.1038/s41598-025-25083-4>.
20. Miao, N.; Yang, M.; Han, P.; Qiao, J.; Che, Z.; Xu, F.; Dai, X.; Zhu, M. A new ensemble learning method stratified sampling blending optimizes conventional blending and improves prediction performance. *Bioinform. Adv.* **2025**, *5*, vbaf002. <https://doi.org/10.1093/bioadv/vbaf002>.
21. Migaybil, H.H.; Gopaluni, B. Advancing predictive modeling in conventional solar stills: A deep learning approach leveraging data augmentation and convolutional neural networks. *Energy Convers. Manag.* **2025**, *346*, 120565. <https://doi.org/10.1016/j.enconman.2025.120565>.
22. Li, Y.; Li, A.; Wang, X.; Yin, J. PerDCGAN: A Perceptual Generative Framework for High-Fidelity Bearing Fault Diagnosis. *Appl. Sci.* **2026**, *16*, 4054. <https://doi.org/10.3390/app16084054>.
23. Lin, B.; Zhao, G.; Wang, X.; Li, H. Facial expression generation based on group residual block generative adversarial network. *Comput. Eng. Appl.* **2024**, *60*, 240–249. <https://doi.org/10.3778/j.issn.1002-8331.2210-0234>.
24. Deng, F.; Wang, S.; Yang, J. Region-aware discriminative learning GAN for super-resolution reconstruction of infrared imagery. *Neurocomputing* **2025**, *639*, 130202. <https://doi.org/10.1016/j.neucom.2025.130202>.
25. Kim, S.; Song, Q.; Liang, F. A New Paradigm for Generative Adversarial Networks based on Randomized Decision Rules. *Stat. Sin.* **2025**, *35*, 897–918. <https://doi.org/10.5705/ss.202022.0404>.
26. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
27. Kung, J.; Chung, H.; Gozalens, R.; Hoo, C.; Cheng, I. Improving the Generalization of Deep Neural Networks Through Regularization Techniques. *EasyChair Prepr.* **2025**, 15810. Available online: <https://easychair.org/publications/preprint/B7s8> (accessed on 22 March 2026).
28. Chen, H.; Sun, Y.; Ma, Y. A two-stage multi-fidelity Bayesian optimization framework with landscape-aware initialization and dual-adaptive sampling. *Appl. Soft Comput.* **2026**, *201*, 115490. <https://doi.org/10.1016/j.asoc.2026.115490>.
29. Rao, P.S.; Sivaprasad, S.; Bar, H.; Dey, P.P. Constitutive modeling and experimental analysis of asynchronous multiaxial fatigue in 304LN stainless steel using Bayesian optimization. *Int. J. Fatigue* **2026**, *202*, 109214. <https://doi.org/10.1016/j.ijfatigue.2025.109214>.
30. He, X.; Ge, X.; Wu, Z.; Zhang, Q.; Yang, Y.; Cao, Y. A Hybrid Intelligent Fault Diagnosis Framework for Rolling Bearings and Gears Based on BAYES-ICEEMDAN-SNR Feature Enhancement and ITOC-LSSVM. *Sensors* **2026**, *26*, 1543. <https://doi.org/10.3390/s26051543>.

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.