

Article

Multi-Phase Adaptive Snow Goose Algorithm-Driven Frequency-Enhanced Residual Networks for Bearing Fault Diagnosis

Zhongcheng Lei ^{1,*} , Bo Ren ^{2,3}, Junqing Wang ^{2,3}, Zhou Li ¹, Chengxiang Liu ^{2,3}  and Zhuoyan Zhong ^{2,3}

¹ School of Robotics, Wuhan University, Wuhan 430072, China

² China Yangtze Power Co., Ltd. (CYPC), Wuhan 430014, China

³ National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Wuhan 430019, China

* Correspondence: zhongcheng.lei@whu.edu.cn

Abstract

To address the problems of low diagnostic accuracy and weak robustness in bearing fault diagnosis, this paper establishes a theoretical framework for frequency-domain-enhanced residual network diagnosis driven by a multi-stage adaptive snow goose algorithm (SGA), breaking through the fundamental bottleneck of structural rigidity in traditional methods. The dual-flexible residual network achieves adaptive matching between network topology and fault characteristics through variable convolutional layers and dynamic skip connections. The frequency-domain attention module, based on Fourier transform and magnitude-phase separation techniques, provides in-depth analysis of the frequency-domain discrimination mechanism in vibration signals. The improved snow goose algorithm integrates multi-stage adaptation, opposition-based learning, elite memory management strategies, and cooperative foraging mechanisms, enhancing the optimization performance of SGA. Validation on the Jiangnan University and Case Western Reserve University (CWRU) bearing fault datasets demonstrates that the proposed method achieves diagnostic accuracies of 99.81% and 99.78%, respectively, representing improvements of 4.62%, 6.41%, and 8.76% over mainstream methods such as ResNet, WDCNN, and Transformer. Furthermore, when faced with changes in data structure, the proposed method exhibits a performance degradation of only 0.43%, which is significantly better than other methods, demonstrating excellent robustness and cross-domain generalization capability, and it has certain practical significance.

Keywords: bearing fault diagnosis; snow goose algorithm; frequency-domain attention; flexible residual network; robustness; intelligent optimization



Academic Editor: Dionisis Kandris

Received: 10 June 2026

Revised: 4 July 2026

Accepted: 19 July 2026

Published: 20 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

Bearings, as core components in rotating machinery equipment, play a crucial role in the safe and stable operation of entire mechanical systems [1]. In industrial production processes, bearings operate under high-speed rotation and heavy-load conditions for extended periods, making them susceptible to faults such as inner race cracks, outer race wear, and rolling element spalling. These faults not only reduce equipment operational efficiency but may also cause equipment damage and safety accidents. Therefore, there is an urgent need for an efficient intelligent diagnostic model to provide technical support for system safe operation [2,3].

Currently, deep-learning-based fault diagnosis methods are widely applied in rotating machinery fields such as bearings. Convolutional Neural Networks (CNNs) [4] have achieved significant success in vibration signal analysis due to their excellent local feature extraction capabilities. Chen et al. [5] applied CNNs to bearing fault diagnosis and extracted various types of frequency and sequential features, achieving good diagnostic accuracy. Kiakojouri et al. [6] optimized CNNs to address the problem of training data scarcity, improving diagnostic performance. However, CNNs have a fixed network structure, making it difficult to adaptively adjust to different fault types or operating conditions, and they also struggle to capture complex patterns over larger ranges. Residual Networks (ResNets) [7] effectively solve the gradient vanishing problem in deep networks through skip connection mechanisms, demonstrating superior performance in fault diagnosis. Hou et al. [8] combined ResNet with Transformer for analysis, validating effectiveness on two bearing fault datasets. Fu et al. [9] utilized ResNet for fault diagnosis analysis, improving fault diagnostic capability. Nevertheless, existing research predominantly employs fixed network structures, making it difficult to adaptively adjust according to different data characteristics and diagnostic tasks, thereby limiting model generalization capability and diagnostic accuracy.

Furthermore, traditional methods primarily focus on time-domain features, with insufficient utilization of frequency-domain information, resulting in deficiencies when processing complex spectral features. To overcome the limitations of fixed network structures, Neural Architecture Search (NAS) [10] technology has gradually become a research hotspot. Researchers have begun applying automated network design concepts to fault diagnosis fields. However, traditional NAS methods incur high computational costs, and existing approaches often remain at simple parameter searches, lacking deep optimization of network intrinsic mechanisms, unable to achieve synergistic enhancement of structural flexibility and feature representation capability.

With the rapid development of attention mechanisms, frequency-domain attention technology has gained widespread attention in signal processing fields. Frequency-domain analysis can reveal spectral information contained in vibration signals, exhibiting unique sensitivity to different fault patterns [11]. Although frequency-domain analysis is sensitive to fault patterns, existing methods have failed to effectively couple frequency-domain feature extraction with attention mechanisms, and the discriminative potential of magnitude and phase information has not been fully exploited.

The snow goose algorithm (SGA) [12,13], as an emerging swarm intelligence algorithm, primarily achieves global optimization by simulating the migration and foraging behaviors of snow geese. However, traditional SGA suffers from issues such as a tendency to fall into local optima and an imbalance between local exploitation and global exploration, making it difficult to meet the demands of complex network structure optimization. Therefore, how to enhance SGA performance in neural network optimization still requires in-depth research.

In summary, this paper proposes a multi-phase adaptive snow goose algorithm-driven frequency-enhanced residual network bearing fault diagnosis method. This method achieves efficient bearing fault diagnosis through the synergistic effects of dual-flexible residual structures, frequency-domain attention mechanisms, and improved optimization algorithms. The main contributions of this paper are as follows:

- (1) To address the problem of fixed network structures in ResNet-based methods, a dual-flexible residual network (DF-ResNet) is proposed, which achieves adaptive adjustment of the network structure through variable convolutional layers and dynamic skip connections, thereby enhancing the model's adaptability to different fault patterns.

- (2) To address the insufficient coupling between frequency-domain feature extraction and attention mechanisms, a frequency-domain attention module (FDAM) is designed to fully exploit the frequency-domain discriminative features of vibration signals.
- (3) To address the issue that the original SGA tends to fall into local optima, an improved snow goose algorithm (ISGA) is proposed, which synergistically integrates multi-stage adaptive transition, opposition-based learning, elite memory management, and cooperative foraging mechanisms to improve the efficiency and accuracy of network structure optimization.
- (4) An ISGA-DF-ResNet-FDAM integrated framework is constructed to achieve joint optimization of structural parameters and model weights, and its effectiveness is validated on different datasets.

The remainder of this paper is organized as follows: Section 2 introduces the design principles of dual-flexible residual networks and frequency-domain attention modules; Section 3 presents the core mechanisms of the ISGA algorithm; Section 4 elaborates on the overall architecture of the ISGA-FResNet-FDAM framework; Section 5 provides experimental results analysis; and Section 6 concludes the paper.

2. Dual-Flexible Residual Network and Frequency-Domain Attention Module

2.1. ResNet

Residual networks effectively solve the gradient vanishing problem in deep neural network training by introducing skip connection mechanisms [14]. Figure 1 shows the basic structure of ResNet.

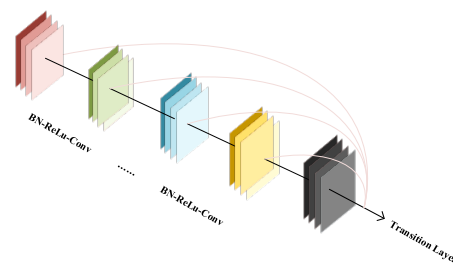


Figure 1. Basic structure of ResNet.

The core idea of standard residual blocks is to learn residual mapping rather than direct mapping, with its mathematical expression represented by the following equation:

$$H(x) = F(x, \{W_i\}) + x \quad (1)$$

where x and $H(x)$ are the input and output of the residual block, W_i is the weight matrix of the i -th layer, and $F(-)$ is the residual function.

For standard residual blocks, the residual function typically contains two convolutional layers, as shown in the following equation:

$$F(x) = \text{ReLU}(\text{Bn}(W_2 * \text{ReLU}(\text{Bn}(W_1 * x)))) \quad (2)$$

where $\text{ReLU}(-)$ is the ReLU activation function, $\text{BN}(-)$ is the batch normalization operation, and $*$ denotes the convolution operation.

2.2. DF-ResNet

Standard ResNet employs fixed numbers of convolutional layers and skip connection patterns, limiting the network's ability to adapt to different data characteristics. Therefore,

this paper proposes a flexible residual block [15] that achieves adaptive structural adjustment by introducing variable parameters. Figure 2 demonstrates the basic principle of DF-ResNet.

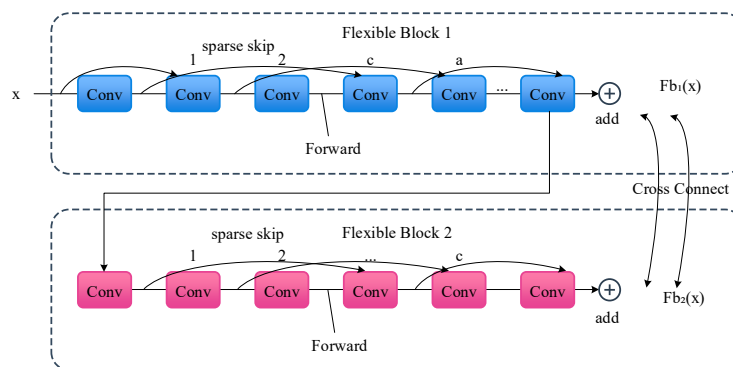


Figure 2. Basic structure of DF-ResNet.

The output h_i of the i -th convolutional layer can be expressed as follows:

$$h_i = \begin{cases} \text{ReLU}(\text{BN}(W_i * h_i - 1)) + \text{Dp}(h_i - 1), & i < N_c \\ \text{BN}(W_i * h_i - 1), & i = N_c \end{cases} \quad (3)$$

where N_c denotes the number of convolutional layers in the flexible residual block, and $\text{Dp}(-)$ denotes the Dropout function, which is used to mitigate overfitting caused by the strong periodicity of vibration signals and the correlation between samples. The dropout rate is set to 0.2, providing effective regularization while retaining sufficient feature information.

The position C_t of skip connections can be dynamically determined by the following equation:

$$C_t = \left(\frac{N_s}{N_c} \times t + 0.5 \right), t = 1, 2, \dots, c \quad (4)$$

where N_s is the number of skip connections in the flexible residual block, t is the t -th skip connection, and c is the number of skip connections.

Since the skip connection position C_t calculated by (4) may be a non-integer while the layer index must be an integer, we apply the rounding operation to round it down to the nearest integer, ensuring that the skip connection is connected to a valid convolutional layer.

To further enhance the network's expressive capability, this paper designs dual-flexible residual blocks containing two cascaded flexible residual paths. This process can be represented by the following equation:

$$\begin{cases} h_1 = Fb_1(x; N_c^{(1)}, N_s^{(1)}) \\ h_2 = Fb_2(x; N_c^{(2)}, N_s^{(2)}) \end{cases} \quad (5)$$

where Fb_1 and Fb_2 are the first and second flexible residual blocks.

The two cascaded flexible residual paths perform complementary feature extraction functions: the first path, Fb_1 , has fewer convolutional layers and focuses on extracting shallow, local temporal features; the second path, Fb_2 , is deeper and is used to capture higher-level fault representations. This division of labor enables the dual paths to simultaneously model both local and global fault characteristics, and the cross-connection and attention fusion in Equation (6) further integrate features from these two levels, the cross-connection and attention fusion mechanism can then be expressed as follows:

$$\begin{cases} h_c = \text{BN}(W_c * h_1) \\ A = \text{sigmoid}(G(\text{Concat}(h_2, h_c))) \\ y = A \odot h_2 + (1 - A) \odot h_c \end{cases} \quad (6)$$

where W_c is the cross-connection weight, $G(-)$ is the global average pooling operation, A is the attention weight, and $\odot$ denotes element-wise multiplication.

2.3. Frequency-Domain Attention Module

2.3.1. Frequency-Domain Transformation and Decomposition

Traditional CNN networks primarily perform feature extraction in the time domain, with insufficient utilization of frequency-domain information. Therefore, this paper proposes a frequency-domain attention module (FDAM) that exploits frequency-domain discriminative features through Fourier transform (FFT).

For input signal $x \in \mathbb{R}^{B \times C \times L}$, FFT transformation is first performed and decomposed into magnitude and phase components. This process can be represented by the following equation:

$$\begin{cases} x_f = \text{FFT}(x) \\ x_m = |x_f| \\ x_p = \text{angle}(x_f) \end{cases} \quad (7)$$

where $\text{FFT}(-)$ is the Fourier function, and B , C , and L are the batch size, number of channels, and signal length, respectively.

2.3.2. Frequency-Domain Attention Mechanism

The frequency-domain channel attention W_{ch} learns the importance weights of different frequency components through the following process:

$$W_{ch} = \text{sigmoid}(w_2 \cdot \text{ReLU}(w_1 \cdot G(x_m))) \quad (8)$$

where w_1 and w_2 are channel attention weights.

The frequency-domain spatial attention W_s focuses on the importance of different frequency positions through the following process:

$$W_s = \text{sigmoid}(\text{BN}(W_{cn} * \text{Concat}(\text{mean}(x_m), \text{max}(x_m)))) \quad (9)$$

where W_{cn} is the spatial attention convolution weight.

To ensure numerical stability, this paper processes the magnitude and phase separately. A small constant is added before computing the phase to avoid division by zero, and the signal is restored to the real domain through the inverse transform via a gating mechanism, thereby preventing the accumulation of complex numerical errors during backpropagation.

2.3.3. Frequency-Domain Feature Enhancement and Reconstruction

The frequency-domain features weighted by attention mechanisms require magnitude enhancement and phase correction. This process can be represented by the following equation:

$$\begin{cases} x'_m = W_s \odot W_{ch} \odot x_m \odot F_a \\ x'_p = x_p + 0.1 \times \tanh(\text{BN}(W_p * x_m)) \end{cases} \quad (10)$$

where F_a is the learnable adaptive filtering parameter, and W_p is the phase correction weight. Meanwhile, the phase correction coefficient is set to 0.1, which keeps the corrected phase close to the original spectral structure, thereby avoiding numerical instability and feature distortion caused by excessive correction.

Finally, enhanced features are output through inverse FFT transformation and gating mechanisms. This process can be represented by the following equation:

$$\begin{cases} x' = \text{Real}(\text{IFFT}(x'_m \odot \exp(j \cdot x'_p))) \\ y_{FDAM} = \text{sigmoid}(W_g * x') \odot x' + x \end{cases} \quad (11)$$

where $\text{IFFT}(-)$ is the inverse Fourier operation, and W_g is the gating weight.

3. ISGA Algorithm

3.1. SGA

3.1.1. Population Initialization

The snow goose algorithm (SGA) is a novel swarm intelligence optimization algorithm proposed by Tian et al. in 2024, which achieves global optimization by simulating the migration and foraging behaviors of snow geese [16].

In the snow goose population, each snow goose represents a candidate solution. The population initialization process can be expressed as:

$$X_{i,j} = l_j + \text{rand} \times (u_j - l_j) \quad (12)$$

where $X_{i,j}$ is the position of the i -th snow goose in the j -th dimension, l_j and u_j are the lower and upper bounds of the problem, and rand is a random number between 0 and 1.

The snow goose population matrix X and fitness matrix F can be represented by the following equations:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,D} \\ X_{2,1} & X_{2,2} & \dots & X_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N,1} & X_{N,1} & \dots & X_{N,D} \end{bmatrix} \quad (13)$$

$$F = [F_1, F_2, \dots, F_N]^T \quad (14)$$

where N is the snow goose population size and D is the dimensionality.

3.1.2. V-Shaped Formation Flight (Exploration Phase)

During long-distance migration, snow geese form a special V-shaped formation to reduce flight resistance and energy consumption. This formation structure can effectively utilize the updraft generated by individuals ahead, thereby improving overall flight efficiency. Based on this phenomenon, the velocity update mechanism of the snow goose algorithm can be represented by the following equation:

$$V_{t+1} = q \cdot V_t + p \quad (15)$$

where V_t is the velocity vector, V_{t+1} is the updated velocity vector, p is the random perturbation term, and q is the inertia weight coefficient, which can be expressed as:

$$q = \frac{4t}{T \cdot e^{\frac{4t}{T}}} \quad (16)$$

where T is the maximum number of iterations and t is the current iteration number.

In actual group migration, snow geese form different behavioral patterns based on flight experience and physical condition. Dominant individuals typically occupy front positions in the formation, while other individuals adopt following strategies. The SGA

algorithm defines the top one-fifth of individuals in the population as leaders, with their position update formula expressed as:

$$X_{t+1}^i = X_t^i + B(X_t^g - X_t^i) - D(X_t^i - X_t^j) + V_{t+1}^i \quad (17)$$

where X_t^i is the position of the i -th snow goose at the t -th generation, X_{t+1}^i is the updated position of the snow goose, X_t^g is the current global optimal position, X_t^j is the position of a randomly selected other individual, and B and D are attraction and repulsion factors.

For other following individuals in the formation, their position update strategy can be expressed as:

$$X_{t+1}^i = X_t^i + B(X_t^g - X_t^i) - D(X_t^k - X_t^i) - E(X_t^n + X_t^i) + V_{t+1}^i \quad (18)$$

where X_t^k is the position of a neighboring excellent individual, X_t^n is the position of an adjacent individual, and E is the coordination factor.

3.1.3. Line Shape (Development Stage)

After completing formation adjustment, the group enters a relatively stable linear flight phase. During this phase, individuals primarily maintain flight direction consistency through local information exchange. The SGA algorithm employs a probabilistic selection mechanism to simulate this behavioral characteristic, which can be represented by the following equation:

$$X_{t+1}^i = \begin{cases} X_t^i + (X_t^i - X_t^g) \cdot R, & r > 0.5 \\ X_t^g + (X_t^i - X_t^g) \cdot R \oplus B(D), & r \leq 0.5 \end{cases} \quad (19)$$

where R and r are both random numbers between 0 and 1, $\oplus$ denotes vector addition operation, and $B(D)$ is a random vector generated by d-dimensional Brownian motion to simulate minor disturbances during flight.

3.2. ISGA

Due to problems such as easily falling into local optima and slow convergence speed in the standard SGA algorithm, this paper proposes an improved snow goose algorithm (ISGA) that integrates multiple enhancement strategies to improve optimization performance.

3.2.1. Multi-Phase Adaptive Mechanism

The SGA algorithm employs fixed search strategies, making it difficult to adapt to the demands of different optimization phases. Here, a multi-phase adaptive mechanism is designed that can dynamically adjust search behavior according to population states. The phase discrimination function $\phi(t)$ can be defined as follows:

$$\phi(t) = \begin{cases} 1, & \text{if } \sigma(t) > \theta_e \text{ or } \Delta f(t) < \epsilon_s \\ 2, & \text{else} \end{cases} \quad (20)$$

where $\sigma(t)$ represents population diversity, θ_e is the exploration threshold, $\Delta f(t)$ is the fitness improvement amount, and ϵ_s is the stagnation discrimination threshold.

The population diversity and fitness improvement amount can be represented by the following equations:

$$\begin{cases} \sigma(t) = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \|X_t^i - X_t^j\|_2 \\ \Delta f(t) = \frac{|f_{t-1}^b - f_t^b|}{|f_{t-1}^b| + \epsilon} \end{cases} \quad (21)$$

where f_t^b is the optimal fitness value at the t -th generation, and ϵ is a positive number to prevent zero values.

3.2.2. Opposition-Based Learning Strategy

Opposition-based learning enhances the algorithm's search capability by simultaneously considering both solutions and their opposite solutions. This process can be represented by the following equation:

$$P_o(t) = P_b \times \left(1 + \frac{\sigma(t)}{\bar{\sigma}}\right) \times \left(1 - \frac{t}{T}\right)^\alpha \quad (22)$$

where $P_o(t)$ is the opposition-based learning probability at the t -th iteration, P_b is the basic opposition probability and α is the decay parameter.

The opposite solution generation formula can be represented by the following equation:

$$\hat{X}_t^i = l_j + u_j - X_t^i + \beta \times (X_t^g - X_t^i) \quad (23)$$

where β is the elite guidance coefficient.

3.2.3. Elite Memory Management

To improve the optimization performance of the SGA algorithm, an elite memory bank is designed to store excellent solutions and guide the search process. This process can be represented by the following equation:

$$E = \{e_1, e_2, \dots, e_i, \dots, e_M\} \quad (24)$$

where M is the elite bank capacity and e_i is the i -th elite solution.

The elite-guided position update can be represented by the following equation:

$$X_{t+1}^i = X_t^i + \eta \times \text{rand} \times (e_s - X_t^i) \quad (25)$$

where η is the learning rate and e_s is the elite solution selected based on fitness and diversity weighting.

3.2.4. Collaborative Foraging Mechanism

To enhance coordination among populations, a sub-population collaboration mechanism is designed to strengthen information exchange. This process can be represented by the following equation:

$$C_{ij} = \exp\left(-\frac{\|X_i - X_j\|_2^2}{2\sigma_c^2}\right) \times \exp\left(-\frac{|f_i - f_j|}{f_a + \epsilon}\right) \quad (26)$$

where C_{ij} is the collaboration intensity between individual i and individual j , σ_c is the spatial distance parameter, and f_a is the average fitness.

The collaboration-guided velocity update formula can be expressed as follows:

$$V_{t+1}^i = w_v \times V_t^i + c_1 \times r_1 \times (X_t^b - X_t^i) + c_2 \times \sum_{k=1}^N C_{i,k} \times V_t^k \quad (27)$$

where w_v is the inertia weight, c_1 and c_2 are acceleration coefficients, and r_1 is a random number.

To enable the continuous swarm-intelligence optimizer (ISGA) to search over discrete network hyperparameters, an encoding–decoding scheme is designed. In the encoding stage, each candidate network is represented by a three-dimensional continuous vector $x = [x_1, x_2, x_3]$, where x_1 , x_2 , and x_3 correspond to the number of convolutional layers, the number of convolutional kernels, and the number of skip connections, respectively. The search ranges are set to $[1, 8]$, $[1, 64]$, and $[1, 6]$ for the three dimensions. During optimization, ISGA updates these vectors in the continuous space according to its migration and foraging mechanisms.

4. ISGA-DFResNet-FDAM Integration Framework

The ISGA-DFResNet-FDAM integration framework organically combines the improved snow goose algorithm, dual-flexible residual network, and frequency-domain attention module to achieve end-to-end optimization for bearing fault diagnosis. Structural parameters such as the number of convolutional layers, number of convolution kernels, and number of skip connections directly affect the network's feature extraction capability and diagnostic performance. Therefore, the ISGA algorithm is utilized to optimize these key parameters. The constructed ISGA-DFResNet-FDAM fault diagnosis process is shown in Figure 3.

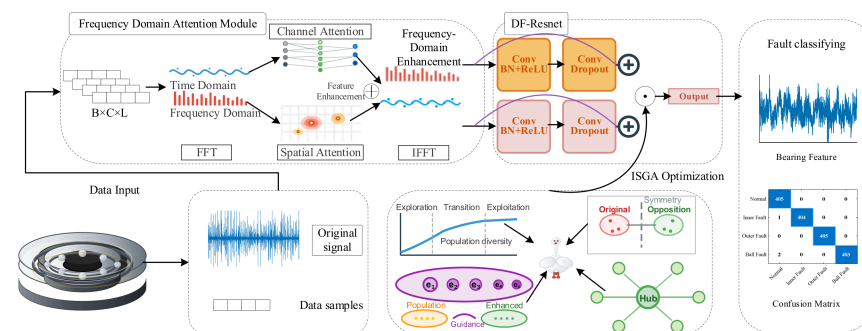


Figure 3. ISGA-DFResNet-FDAM diagnosis process.

The specific implementation steps are as follows:

Step 1: Collect bearing fault data, perform standardization processing and sliding window sampling on the original signals, and divide the training and testing sets in a 7:3 ratio, ensuring balanced distribution of samples across all fault categories.

Step 2: Set the population size, set the maximum iteration number to 40, and initialize the elite memory bank capacity to 20, opposition-based learning probability to 0.3, and collaborative foraging weight to 0.1.

Step 3: Execute the optimization loop: construct the DF-ResNet-FDAM network according to current population individual encoding, perform rapid training using training data and evaluate fitness on the validation set; update population positions using the multi-phase adaptive mechanism, apply opposition-based learning, elite memory, and collaborative foraging strategies to improve the SGA algorithm; and use the improved SGA algorithm to optimize network hyperparameters to obtain the current optimal solution.

Step 4: Termination condition judgment: Check whether the maximum iteration number is reached or fitness converges. If not satisfied, return to Step 3 to continue optimization; if satisfied, terminate the search and output optimal network structural parameters.

Step 5: Construct the final DF-ResNet-FDAM network using the optimal structural parameters obtained from ISGA search and perform comprehensive training on the complete training set.

Step 6: Evaluate the optimal model's fault diagnosis performance on the test set and output the final results.

5. Experimental Verification

5.1. Case 1: Jiangnan University Dataset

5.1.1. Dataset Introduction

The experiments employ the Jiangnan University bearing fault dataset for validation analysis; a detailed data description can be found in reference [17]. This dataset creates bearing defects through artificial cutting methods, with bearing types including outer race fault, inner race fault, rolling element fault, and normal conditions. The inner race fault bearing model is NU205, while other healthy condition bearings are model N205. Both N205 and NU205 bearings have outer and inner diameters of 25 mm and 15 mm, respectively. N205 has 10 rolling elements with a 7 mm diameter, while NU205 has 11 rolling elements with a 7 mm diameter. The damage dimensions (width $\times$ depth) for inner and outer races are both 0.3 mm $\times$ 0.25 mm, and the rolling element damage dimension is 0.5 mm $\times$ 0.15 mm.

Vibration data acquisition includes three operating conditions: sampling frequency of 50 kHz, 0 horsepower load, and speed of 600 r/min; sampling frequency of 50 kHz, 0 horsepower load, and speed of 800 r/min; and sampling frequency of 50 kHz, 0 horsepower load, and speed of 1000 r/min. Detailed dataset information is shown in Table 1.

Table 1. Detailed dataset information.

Running State	Speed (rpm)	Frequency (kHz)
Inner Fault	600	50
Outer Fault	600	50
Ball Fault	600	50
Normal	600	50
Inner Fault	800	50
Outer Fault	800	50
Ball Fault	800	50
Normal	800	50
Inner Fault	1000	50
Outer Fault	1000	50
Ball Fault	1000	50
Normal	1000	50

First, data processing is performed using sliding windows for sample construction, where the sliding window step size is set to 1024 sampling points, the sliding window length is set to 2048 sampling points, and 450 samples are extracted from each file. To prevent data leakage, the raw vibration signals are first divided into training and test segments at a 7:3 ratio, and the sliding window sampling is then performed independently within each segment, ensuring that the 50% overlapping samples never span both sets. Then, the random seed is set to 24, the AdamW optimizer is used for training with a weight

decay of 5×10^{-3} , and the learning rate follows a cosine annealing schedule, and Z-score standardization is subsequently applied for data preprocessing:

$$z_i^* = \frac{z_i - \mu_i}{\delta_i} \quad (28)$$

where z_i is the original data, μ_i and δ_i are the mean and standard deviation of the original data.

5.1.2. Experimental Setup

First, the relevant parameters of the ISGA algorithm are set: population size is set to 5, maximum iteration number to 40, elite memory bank capacity to 20, opposition-based learning probability to 0.3, multi-phase adaptive threshold to 0.01, learning rate to 0.001, batch size to 64, and maximum training epochs to 50, and dropout layers are added to prevent model overfitting. Since each fitness evaluation in architecture search requires training a candidate network, a small population size and a moderate number of iterations are adopted to ensure convergence while controlling computational cost. The elite memory capacity is set to retain a sufficient number of high-quality solutions while avoiding excessive storage overhead. The opposition-based learning probability adopts a commonly used value. The multi-stage adaptive threshold is set to a small value to achieve a smooth transition from the exploration stage to the exploitation stage. The network training parameters are set to obtain stable convergence on the two datasets.

Second, all experiments are conducted in MATLAB 2022b. The experimental platform consists of a Windows 11 system, 13th Gen Intel (R) Core (TM) i5-13400F processor, and NVIDIA GeForce GTX 1660 Ti graphics card.

Finally, the experiments employ accuracy, precision, recall, and F1-score as main evaluation metrics for research.

5.1.3. Diagnostic Results Analysis

First, to verify the effectiveness of each module, the following four ablation experiments are designed: ISGA-DF-ResNet (without FDAM), ISGA-ResNet-FDAM (without flexible structure), SGA-DF-ResNet-FDAM (without ISGA improvement), and DF-ResNet-FDAM (without optimization algorithm). The experimental results are shown in Table 2, and the loss curves of each model are shown in Figure 4.

Table 2. Ablation experiment results.

Model	Accuracy/%	Change Percentage/%
ISGA-DF-ResNet-FDAM	99.81	0
ISGA-DF-ResNet	99.63	0.18
ISGA-ResNet-FDAM	99.51	0.30
SGA-DF-ResNet-FDAM	99.32	0.49
DF-ResNet-FDAM	95.49	4.33

From the analysis of the results in Table 2, ISGA-DF-ResNet-FDAM achieves optimal performance with an accuracy of 99.81%. After removing the FDAM module, accuracy decreases by 0.18%, demonstrating the effectiveness of the frequency-domain attention mechanism in processing complex spectral features. Using standard ResNet instead of the dual-flexible structure results in a 0.30% performance decrease, proving the advantages of adaptive convolution layers and dynamic skip connection design. Adopting standard SGA instead of ISGA leads to a 0.49% accuracy decrease, validating the effectiveness of improved optimization strategies. DF-ResNet-FDAM achieves only 95.49% accuracy, a

4.33% decrease compared to the complete framework, indicating that the four improvement strategies in ISGA form a synergistically coupled whole. The multi-stage adaptive transition is responsible for switching between the exploration and exploitation stages. In the exploration stage, opposition-based learning and cooperative foraging work together to expand the search coverage and reduce the risk of premature convergence. In the exploitation stage, the elite memory management strategy retains high-quality solutions and ensures convergence accuracy. Therefore, intelligent structure optimization is a key factor in improving diagnostic performance.

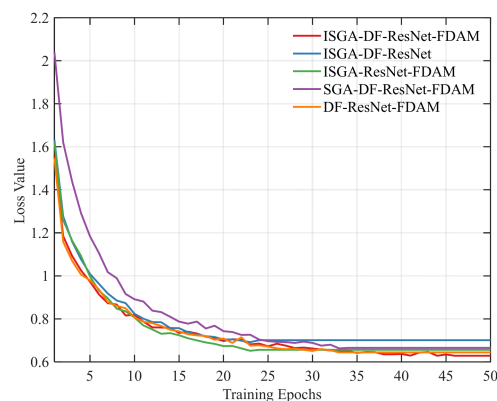


Figure 4. Loss function curves of each model.

From Figure 4, ISGA-DF-ResNet-FDAM ultimately achieves the lowest loss value with relatively smooth loss curves, indicating that the structurally optimized network has better training stability. The synergistic effects of various modules significantly enhance model convergence quality and final performance, demonstrating the feasibility of the proposed method.

Second, to verify the effectiveness of the proposed ISGA-DF-ResNet-FDAM model, DBO-DFResNet-FDAM and GWO-DFResNet-FDAM are selected as comparison methods to validate the optimization performance of the ISGA algorithm. Additionally, state-of-the-art networks, including ResNet, Transformer, and WDCNN, are used for comparative experiments. The parameter settings for each model are as follows: ResNet, DBO-DFResNet-FDAM, and GWO-DFResNet-FDAM models have hyperparameter settings consistent with the ISGA-DF-ResNet-FDAM model; Transformer has four attention heads and four Transformer layers; and WDCNN scales are set to three, five, and seven, with learning rate set to 0.001. The confusion matrix results of each model are shown in Figure 5, and experimental results are shown in Figure 6 and Table 3.

Table 3. Results of each model on the Jiangnan University Dataset.

Diagnostic Model	Accuracy/%	Evaluation Metrics		
		Macro-Averaged Precision	Macro-Averaged Recall	Macro-Averaged F1-Score
ISGA-DF-ResNet-FDAM	99.81	0.9982	0.9981	0.9981
DBO-DFResNet-FDAM	98.77	0.9880	0.9877	0.9876
GWO-DFResNet-FDAM	96.11	0.9630	0.9611	0.9603
ResNet	95.19	0.9550	0.9519	0.9510
WDCNN	93.40	0.9354	0.9340	0.9341
Transformer	91.05	0.9132	0.9105	0.9112

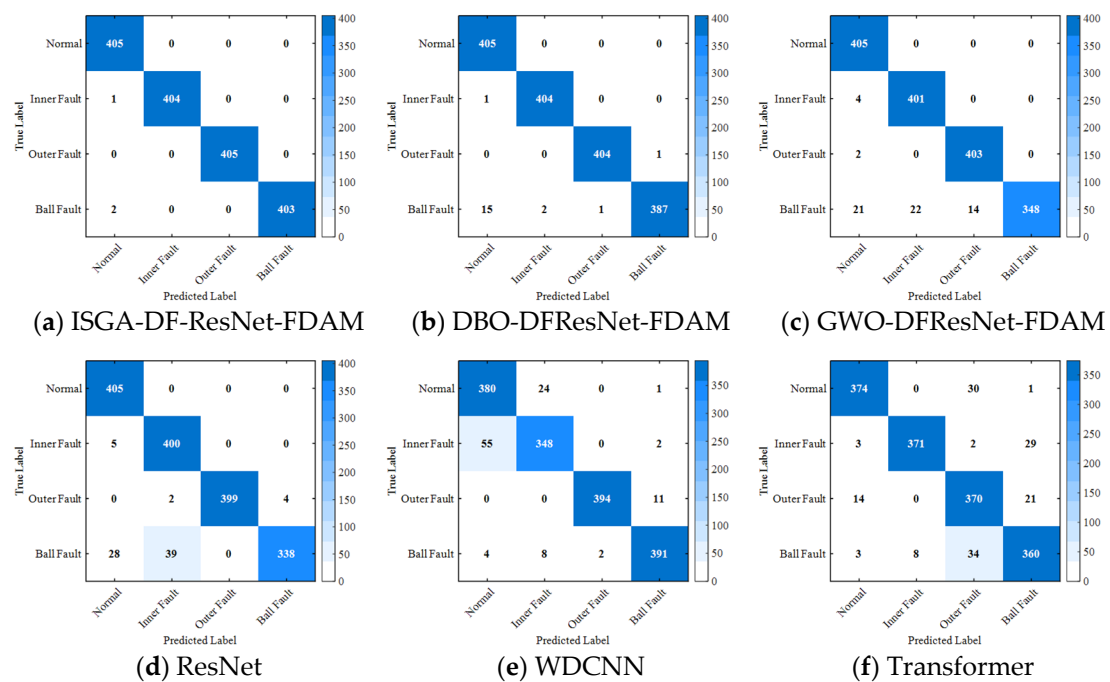


Figure 5. Confusion matrix results of each model.

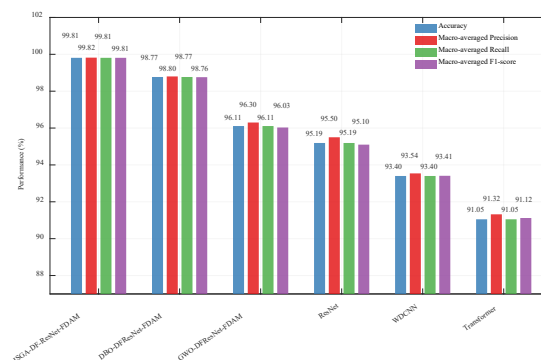


Figure 6. Prediction results of each model.

From the results in Figure 5, the ISGA-DF-ResNet-FDAM model demonstrates excellent recognition capability across all fault categories with very few misclassified samples. The DBO-DF-ResNet-FDAM and GWO-DF-ResNet-FDAM models perform poorly in rolling element fault classification, while other networks exhibit obvious confusion phenomena. This indicates that the proposed method significantly improves bearing fault diagnosis accuracy through intelligent structure optimization and multi-module collaborative design.

From the comparative results in Figure 6 and Table 3, ISGA-DF-ResNet-FDAM achieves optimal performance across all evaluation metrics, with accuracy reaching 99.81% and macro-averaged precision, recall, and F1-score all exceeding 99.8%. The comprehensive diagnostic accuracy improves by 1.04%, 3.70%, 4.62%, 6.41%, and 8.76% compared to DBO-DFResNet-FDAM, GWO-DFResNet-FDAM, ResNet, WDCNN, and Transformer, respectively, validating the superiority of the ISGA-DF-ResNet-FDAM model in bearing fault diagnosis. However, compared with fixed-structure baseline methods, the proposed method incurs greater computational overhead. This is because ISGA needs to evaluate multiple candidate architectures, and thus the search phase increases the overall optimization time.

5.1.4. Model Robustness Analysis

Currently, there are three methods for verifying robustness: modifying data structure, modifying model structure, and adding external modules. Here, the data structure modification approach is used to verify the robustness of the proposed model by reducing the sliding window length from 2048 to 1024. The diagnostic results of each model after modification are shown in Table 4.

Table 4. Experimental results of each model.

Model	Diagnostic Accuracy/%		Change Percentage/%
	Before Modification	After Modification	
ISGA-DF-ResNet-FDAM	99.81	99.38	0.43
DBO-DFResNet-FDAM	98.77	98.15	0.63
GWO-DFResNet-FDAM	96.11	95.06	1.09
ResNet	95.19	91.91	3.45
WDCNN	93.40	90.12	3.51
Transformer	91.05	86.42	5.09

From the results in Table 4, when the model's data structure is modified, the proposed ISGA-DF-ResNet-FDAM model's diagnostic accuracy decreases by only 0.43%, showing the smallest performance degradation. The performance of DBO-DFResNet-FDAM and GWO-DFResNet-FDAM decreases by 0.63% and 1.09%, respectively, still maintaining relatively stable performance. However, ResNet, WDCNN, and Transformer show significant performance degradation of 3.45%, 3.51%, and 5.09%, respectively. This indicates that traditional fixed-structure networks are sensitive to input data changes with poor adaptability, while ISGA-DF-ResNet-FDAM can effectively cope with data structure changes and maintain stable diagnostic performance through adaptive network structure design and intelligent optimization strategies.

5.2. Case 2: CWRU Bearing Dataset

To verify the generalization capability of the proposed ISGA-DF-ResNet-FDAM model, the Case Western Reserve University (CWRU) bearing fault dataset is further employed for cross-dataset validation experiments; a detailed data description can be found in reference [18]. This dataset consists of a 2-horsepower motor, torque encoder, dynamometer, and control electronics, using a 16-channel data acquisition system to collect vibration signals from vibration bearings. The experimental bearing model is SKF6205, with signal sampling frequency of 48 kHz.

The experiments analyze drive-end bearing signals with a sampling frequency of 12 kHz, extracting 300 samples for each fault type, with 70% of samples used as the training dataset and 30% as the test set. The same preprocessing strategy as the Jiangnan University dataset is applied to ensure experimental consistency and comparability. The dataset division is shown in Table 5.

The experimental results on the CWRU dataset are shown in Table 6 and Figures 7 and 8.

Radar chart 7 shows the comprehensive performance of each model across four evaluation metrics, with ISGA-DF-ResNet-FDAM forming the largest coverage area across all dimensions, reflecting its balanced and excellent diagnostic performance. Cross-dataset experimental results validate that ISGA-DF-ResNet-FDAM possesses good generalization capability and robustness.

Table 5. Data distribution.

Running State	Fault Diameter/mm	Sample Number	Label
Normal	/	300	0
Inner Fault	0.1778	300	1
Ball Fault	0.1778	300	2
Outer Fault	0.1778	300	3
Inner Fault	0.3556	300	4
Ball Fault	0.3556	300	5
Outer Fault	0.3556	300	6
Inner Fault	0.5334	300	7
Ball Fault	0.5334	300	8
Outer Fault	0.5334	300	9

Table 6. Diagnostic results of each model.

Diagnostic Model	Accuracy/%	Evaluation Metrics		
		Macro-Averaged Precision	Macro-Averaged Recall	Macro-Averaged F1-Score
ISGA-DF-ResNet-FDAM	99.78	0.9978	0.9978	0.9978
DBO-DFResNet-FDAM	99.44	0.9945	0.9944	0.9944
GWO-DFResNet-FDAM	99.33	0.9935	0.9933	0.9933
ResNet	97.89	0.9801	0.9789	0.9790
WDCNN	92.00	0.9280	0.9200	0.9136
Transformer	94.33	0.9564	0.9433	0.9406

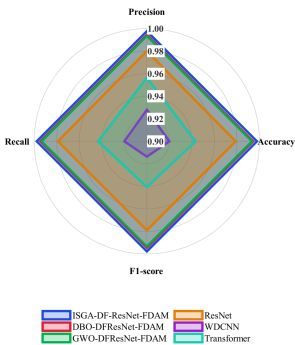


Figure 7. Diagnostic radar chart of each model.

From Table 6 and Figure 8, ISGA-DF-ResNet-FDAM achieves excellent diagnostic performance on the CWRU dataset as well, with all metrics reaching 99.78%, demonstrating superior cross-dataset generalization capability. The proposed method improves accuracy by 0.34%, 0.45%, 1.89%, 5.45%, and 7.78% compared to DBO-DFResNet-FDAM, GWO-DFResNet-FDAM, ResNet, Transformer, and WDCNN, respectively.

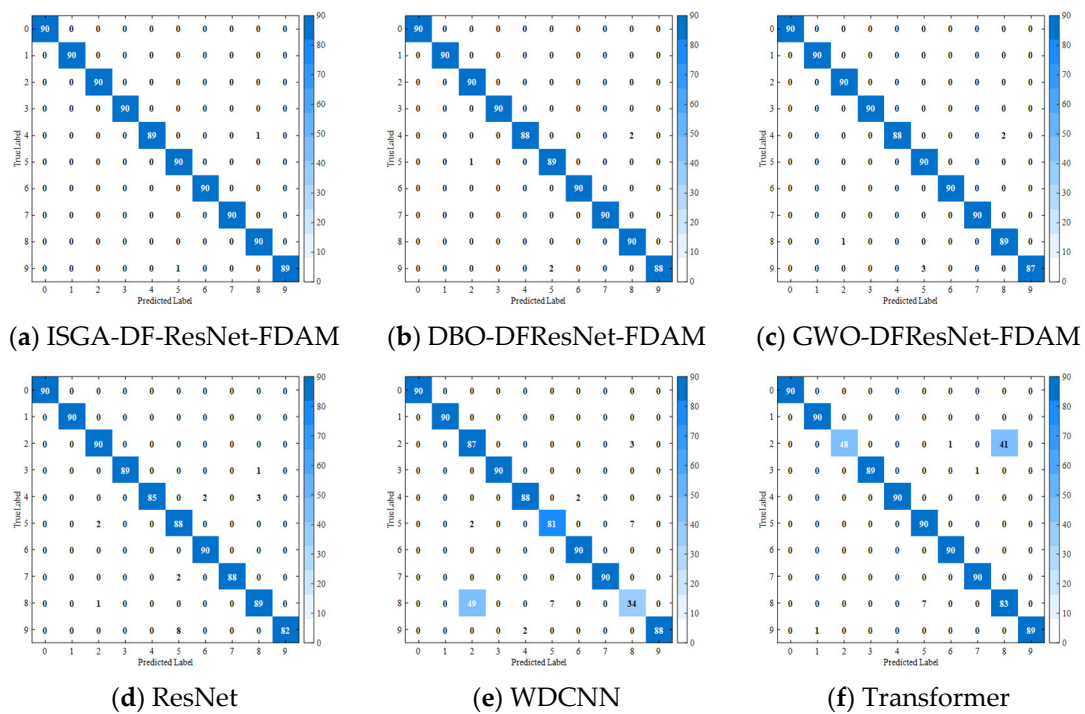


Figure 8. Confusion matrix results of each model on dataset 2.

6. Conclusions

This paper proposes a multi-phase adaptive snow goose algorithm-driven frequency-enhanced residual network bearing fault diagnosis method, addressing the limitations of traditional fixed-structure networks in complex fault pattern recognition. By constructing a dual-flexible residual network, adaptive adjustment of convolution layers and skip connections is achieved, enabling the network to dynamically optimize topological structures according to different fault characteristics. The frequency-domain attention module deeply exploits the frequency-domain discriminative features of vibration signals, effectively compensating for the deficiencies of traditional time-domain analysis. The improved snow goose algorithm integrates multi-phase adaptive mechanisms, opposition-based learning strategies, elite memory management, and collaborative foraging mechanisms, achieving an effective balance between global search and local exploitation in high-dimensional parameter spaces. On the Jiangnan University and CWRU datasets, the proposed method achieves diagnostic accuracies of 99.81% and 99.78%, respectively, representing improvements of 4.62%~8.76% compared to mainstream methods such as ResNet, WDCNN, and Transformer. Robustness testing shows that when data structure changes occur, the method's performance degradation is only 0.43%, while cross-dataset validation confirms the method's good generalization capability, providing reliable assurance for practical engineering applications. This research establishes an integrated framework combining intelligent optimization algorithms with deep network architecture design, providing theoretical support for the field of rotating machinery fault diagnosis. Future research will build on this work to further expand systematic comparisons with more state-of-the-art methods and to conduct more comprehensive robustness evaluations under multiple operating conditions.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, Z.L. (Zhongcheng Lei) software, validation, B.R.; validation, formal analysis, J.W. and Z.L. (Zhou Li); resources, data curation, C.L.; visualization, supervision, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Open Research Fund of National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety (grant no. GJGCZX-JJ-202415).

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

Conflicts of Interest: Author Bo Ren, Junqing Wang, Chengxiang Liu, Zhuoyan Zhong were employed by the company China Yangtze Power 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.

References

1. Wang, Z.; Han, G.J.; Liu, L.; Wang, F.; Zhu, Y. A Globally Interpretable Convolutional Neural Network Combining Bearing Semantics for Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3507713. [\[CrossRef\]](#)
2. Zhou, J.M.; Yang, X.T.; Liu, L.L.; Wang, Y.; Wang, J.; Hou, G. Fuzzy Broad Learning System Combined with Feature-Engineering-Based Fault Diagnosis for Bearings. *Machines* **2022**, *10*, 1229. [\[CrossRef\]](#)
3. Zhang, J.S.; Zhang, K.; An, Y.Y.; Luo, H.; Yin, S. An Integrated Multitasking Intelligent Bearing Fault Diagnosis Scheme Based on Representation Learning Under Imbalanced Sample Condition. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 6231–6242. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, H.; Liu, Z.L.; Peng, D.D.; Cheng, Z. Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. *ISA Trans.* **2022**, *128*, 470–484. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Chen, H.M.; Meng, W.; Li, Y.J.; Xiong, Q. An anti-noise fault diagnosis approach for rolling bearings based on multiscale CNN-LSTM and a deep residual learning model. *Meas. Sci. Technol.* **2023**, *34*, 045013. [\[CrossRef\]](#)
6. Kiakojouri, A.; Wang, L. A Generalized Convolutional Neural Network Model Trained on Simulated Data for Fault Diagnosis in a Wide Range of Bearing Designs. *Sensors* **2025**, *25*, 2378. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Wang, H.T.; Fang, G.Z.; Cui, X.L.; An, X. Parallel Axial Attention and ResNet-Based Bearing Fault Diagnosis Method. *Electronics* **2026**, *15*, 899. [\[CrossRef\]](#)
8. Hou, S.X.; Lian, A.; Chu, Y.D. Bearing fault diagnosis method using the joint feature extraction of Transformer and ResNet. *Meas. Sci. Technol.* **2023**, *34*, 075108. [\[CrossRef\]](#)
9. Fu, L.; Wang, S.N.; Chen, R.E.; Ma, Z.; Ma, J.; Yao, B.; Xu, F. Intelligent Fault Diagnosis of Rolling Bearings Based on an Improved Empirical Wavelet Transform and ResNet Under Variable Conditions. *IEEE Sens. J.* **2023**, *23*, 29097–29108. [\[CrossRef\]](#)
10. Li, L.T.; Jiang, H.K.; Wang, R.X.; Yang, Q. A reinforcement neural architecture search convolutional neural network for rolling bearing fault diagnosis. *Meas. Sci. Technol.* **2023**, *34*, 115122. [\[CrossRef\]](#)
11. Zhang, F.Y.; Yao, P.; Geng, X.Y.; Mu, L.; Paitekul, P.; Viyanit, E.; Jiang, M.; Jia, L. Intelligent rolling bearing compound fault diagnosis based on frequency-domain Gramian angular field and convolutional neural networks with imbalanced data. *J. Vib. Control* **2024**, *30*, 5522–5535. [\[CrossRef\]](#)
12. Tian, A.Q.; Liu, F.F.; Lv, H.X. Snow Geese Algorithm: A novel migration-inspired meta-heuristic algorithm for constrained engineering optimization problems. *Appl. Math. Model.* **2024**, *126*, 327–347. [\[CrossRef\]](#)
13. Cao, Z.Y.; Zhou, J.X.; Xiao, K.; Zhou, Y.; Fei, X. Physics-driven digital twin model and CNN-GRU network for remaining useful life prediction in rolling bearings. *Meas. Sci. Technol.* **2025**, *36*, 096105. [\[CrossRef\]](#)
14. Zhu, W.B.; Ni, H.B.; Li, Z.; Cao, J.; Ni, B.; Chang, J. A Multimodal Progressive Fusion Bearing Fault Diagnosis Algorithm Based on Residual Network. *IEEE Sens. J.* **2025**, *25*, 23857–23868. [\[CrossRef\]](#)
15. Chen, C.; Li, X.F.; Xi, J.T.; Jiantao, S.; Dongdong, Y. Intelligent fault diagnosis method for rolling bearings based on flexible residual neural network. *Chin. J. Eng.* **2025**, *47*, 480–488. [\[CrossRef\]](#)
16. Bian, H.H.; Li, C.; Liu, Y.H.; Tong, Y.; Bing, S.; Chen, J.; Ren, Q.; Zhang, Z. Improved snow geese algorithm for engineering applications and clustering optimization. *Sci. Rep.* **2025**, *15*, 4506. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Yang, Z.Y.; Mao, R.Z.; Ye, L.C.; Liu, Y.; Hu, X.; Li, Y. VSC-ACGAN: Bearing fault diagnosis model applied to imbalanced samples. *Meas. Sci. Technol.* **2025**, *36*, 036212. [\[CrossRef\]](#)
18. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [\[CrossRef\]](#)

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.