

Article

Globally Synchronized Optimization for Rolling Bearing RUL Prediction Under Limited Data with Remora-Optimized WGAN-GP and LSTM

Ying Bai * , Kunpeng Chen and Man Li

School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China;
kunpengchen_hfuu@126.com (K.C.); manli_hfuu@126.com (M.L.)

* Correspondence: baiying@hfuu.edu.cn

Abstract

In recent years, data-driven remaining useful life (RUL) prediction for bearings has gained significant attention in industrial prognostics. However, real-world applications face challenges due to the scarcity and noise of run-to-failure data, as well as the high sensitivity of deep models to hyperparameters. Traditional methods often optimize data enhancement and prediction modules separately, which may limit end-to-end performance improvement and reduce generalization in real-world scenarios. To address this, this paper introduces ROA-DE-LSTM, a novel framework that jointly enhances prediction accuracy and generalization under limited and imperfect data conditions. The framework features an adaptive Remora Optimization Algorithm (ROA) that dynamically adjusts optimization strategies, thereby improving both data augmentation and RUL prediction. Specifically, an ROA-optimized Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) is employed to generate high-quality augmented samples, while an ROA-optimized LSTM network captures long-term temporal dependencies for stable RUL estimation. Unlike traditional approaches, ROA-DE-LSTM synchronously optimizes the data enhancement and prediction stages within a unified framework. Extensive experiments on the XJTU-SY and IEEE PHM2012 datasets demonstrate that ROA-DE-LSTM outperforms representative baselines and exhibits competitive advantages over other optimization-based strategies. Cross-dataset experiments verify the effectiveness of the proposed data enhancement and optimization strategies, showing that the framework maintains stable and robust performance under different operating conditions.

Keywords: data enhancement; deep learning; global optimization; remaining useful life prediction; remora optimization algorithm



Academic Editor: Rui Araújo

Received: 11 June 2026

Revised: 7 July 2026

Accepted: 9 July 2026

Published: 11 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

With industrial equipment becoming increasingly complex and intelligent, the health monitoring and fault prediction of critical components have attracted growing attention [1]. Rolling bearings, as indispensable elements in rotating machinery, exert a direct and decisive influence on overall system performance and reliability [2]. In real industrial scenarios, bearing failures may lead to unexpected downtime, increased safety risks, and substantial maintenance costs. Consequently, accurately predicting the remaining useful life (RUL) of bearings and enabling predictive maintenance are essential for ensuring safe operation and reducing operating expenses [3]. From an engineering perspective, the primary objective is not only to achieve accurate RUL prediction under ideal laboratory conditions,

but also to ensure reliable and stable performance in real-world scenarios, where degradation data are often limited and contaminated by noise. Benefiting from advances in sensor technology and the rapid development of learning algorithms, data-driven and deep learning-based bearing RUL prediction has made notable progress in recent years [4,5]. A typical data-driven workflow collects condition-monitoring signals and applies signal processing techniques to establish a mapping between measured observations and bearing degradation states [6]. Degradation indicators are then integrated with deep learning models to predict the evolution trajectory and, subsequently, estimate the RUL. Catelani et al. [7] derived degradation indicators from data-driven condition monitoring and adopted recurrent neural networks (RNNs) for RUL estimation; Zhou et al. [8] constructed unsupervised health indicators and employed gated recurrent units (GRUs) with augmented memory; and Lu et al. [9] developed a monotonicity-based HI (Health Index) and introduced a physically guided long short-term memory (LSTM) network for RUL prediction. These studies indicate that sequence modeling and degradation indicator forecasting can effectively support RUL estimation, provided that sufficient high-quality run-to-failure data are available.

However, when transitioning from controlled laboratory settings to real industrial environments, the performance of existing RUL prediction models is often constrained by two tightly coupled practical challenges. First, data availability and data quality remain fundamental bottlenecks. Many approaches implicitly assume that sensor data are complete and accurate; yet, comprehensive high-quality data acquisition is costly and frequently impractical, leading to uneven data quality and insufficient sample sizes [10,11]. Under such conditions, deep models may fit spurious patterns and yield unstable RUL estimates. Second, prediction performance is highly sensitive to hyperparameter configurations. Even with a fixed deep learning architecture, hyperparameter settings can substantially influence convergence behavior, predictive accuracy, and generalization [12]. Manual tuning is time-consuming and labor-intensive, and it is difficult to obtain globally optimal configurations—especially for frameworks that include both a data enhancement module and a downstream time-series predictor. More importantly, these two challenges are mutually reinforcing: limited data increase hyperparameter sensitivity, while suboptimal hyperparameter settings may suppress the effectiveness of data enhancement strategies, thereby limiting end-to-end predictive reliability.

To mitigate data scarcity and quality issues, recent studies have explored various data enhancement techniques. Grasso et al. [13] proposed a data-driven automated approach based on optimized empirical modal decomposition (EMD) for enhancing rolling bearing vibration signals. Guo et al. [14] adopted a GAN-based enhancement strategy using an auxiliary classifier GAN to generate bearing samples. Yu et al. [15] introduced a reparametrized residual denoising diffusion probabilistic model (ReF-DDPM) to generate high-quality samples by transforming Gaussian noise and original samples through Markov chains. These studies demonstrate that data augmentation can improve RUL prediction performance under limited-data conditions. However, a key limitation remains: existing frameworks often treat data enhancement and predictor optimization as separate processes, whereas in practice, the effectiveness of enhancement (i.e., whether generated samples are truly beneficial for prediction) and the generalization of the predictor are both strongly governed by their hyperparameters. Without a unified, globally optimized configuration, data enhancement may fail to translate into consistent predictive gains, and improvements can be difficult to reproduce across varying operating conditions.

Based on the above analysis, this paper proposes an integrated framework, named ROA-DE-LSTM, for bearing RUL prediction under limited and imperfect data conditions. The Health Index (HI) is constructed based on a priori knowledge to link condition-

monitoring data with RUL [16]. A Wasserstein GAN with gradient penalty (WGAN-GP) is employed for data enhancement to improve the diversity and quality of training samples, while an LSTM network is used to model long-term dependencies in HI time series for stable RUL prediction. Moreover, a Remora Optimization Algorithm (ROA) [17] is introduced as a global synchronous tuning strategy to jointly optimize the data enhancement process and the predictor hyperparameters, thereby improving end-to-end performance.

Therefore, the critical contributions of this paper can be summarized as follows:

(1) ROA-optimized WGAN-GP for high-quality data enhancement (ROA-WGAN-GP): To address scarce bearing degradation data, the WGAN-GP enhancement module is globally optimized by ROA to generate high-quality simulation samples, increasing training data diversity and improving downstream RUL prediction performance.

(2) ROA-optimized LSTM predictor for HI-based RUL estimation. ROA is used to globally optimize LSTM hyperparameters, improving learning efficiency and prediction accuracy by more effectively modeling long-term dependencies in HI time series.

(3) End-to-end ROA-DE-LSTM with global synchronous optimization across enhancement and prediction. Unlike approaches that tune enhancement and prediction separately, the proposed framework integrates ROA-WGAN-GP into the ROA-LSTM predictor and jointly optimizes hyperparameters for both stages, thereby improving predictive accuracy and generalization and achieving substantial, quantifiable gains in prediction metrics.

The remainder of this paper is organized as follows: Section 2 reviews related work on data enhancement, time-series prediction, and intelligent optimization algorithms. Section 3 presents the proposed ROA-DE-LSTM rolling-bearing RUL prediction network. Section 4 describes test cases, model parameters, and evaluation metrics. Finally, Section 5 reports and discusses the experimental results, demonstrating the effectiveness of the proposed approach.

2. Related Work

2.1. Data Enhancement

In machine learning and deep learning, data enhancement is a widely used technique aimed at expanding datasets by generating new and informative training samples to improve model generalization and prediction accuracy [18,19]. Data augmentation for rolling bearings is particularly important because the data generated during actual operation are often limited and contaminated by noise and outliers. In recent years, Generative Adversarial Networks (GANs) have attracted widespread attention due to their ability to generate high-quality synthetic data [20]. In particular, WGAN-GP [21] effectively addresses the issues of unstable training and mode collapse in the original GAN by introducing a gradient penalty term, thereby improving the diversity and quality of the generated data. Inspired by WGAN-GP and Auxiliary Classifier GAN (ACGAN), Li et al. [22] proposed the ACWGAN-GP method, which incorporates a gradient penalty into ACGAN to generate samples for minority classes, thereby addressing the issue of poor data quality. Chen et al. [23] proposed a method for enhancing rolling bearing data by jointly optimizing the pre-training process and auxiliary classifiers within a WGAN-GP framework. Gu et al. [24] introduced the CSWGAN-GP approach, which incorporates cosine similarity to refine the gradient penalty term for bearing data enhancement. The effectiveness of this method has been validated through experimental results. Despite these advancements, challenges such as insufficient samples, limited variability, and data imbalance still persist, especially in bearing datasets. Although existing methods have improved data diversity, issues such as poor training stability, data imbalance, and insufficient data quality remain unresolved. While WGAN-GP has shown success in generating diverse data, it still faces limitations in hyperparameter optimization and training efficiency. By combining heuristic optimization

algorithms such as ROA with WGAN-GP, these challenges can be effectively addressed, thereby improving data generation quality and significantly enhancing the performance of downstream prediction models.

2.2. Time Series Prediction

The vibration signals of rolling bearings are typical time-series data containing rich state information. By extracting time-domain features from vibration signals, abnormal conditions in bearings can be detected early, which is essential for effective predictive maintenance strategies [25].

Long Short-Term Memory (LSTM) has become a mainstream model for bearing RUL prediction due to its ability to capture long-term dependencies in time-series data. Sun et al. [26] employed a lightweight bidirectional LSTM model with automatic pruning for RUL prediction. Chen et al. [27] combined an attention mechanism with LSTM for this task. Chang et al. [28] proposed a multilayer mesh-optimized LSTM prediction method, with experimental results demonstrating its effectiveness. Pater et al. [29] designed an LSTM auto encoder for RUL prediction of aircraft turbofan engines. Zhu et al. [30] applied a multi-head attention mechanism to filter important temporal information from rolling data and capture temporal dependencies using LSTM. Lin et al. [31] proposed a novel data-driven framework based on multi-scale sequences and LSTM networks for estimating the RUL of complex systems, which proved effective in experiments.

Despite advancements in improving LSTM prediction accuracy through various optimized architectures, complex models often reduce generalization performance, and increasing model complexity leads to longer training time. Existing LSTM models still face challenges such as data imbalance and data scarcity [20,24]. For example, generated data may fail to fully capture the diversity of states, and most optimization algorithms are computationally expensive and time-consuming [32,33].

Thus, optimizing LSTM using lightweight and fast-tuning strategies is critical for improving prediction performance and generalization, especially in scenarios with scarce and imbalanced data. The need for efficient and lightweight model adjustments is increasingly urgent to enhance LSTM's applicability in real-world, data-deficient environments.

2.3. Intelligent Optimization Algorithms

Hyperparameter tuning is essential for optimizing deep learning models. Traditional methods are often inefficient, especially in high-dimensional parameter spaces, where locating the global optimum is challenging. This limitation has led to the adoption of intelligent optimization algorithms for automatically adjusting hyperparameters and enhancing model performance. Zang et al. [32] proposed the use of the Slime Mold Algorithm (SMA) to optimize LSTM hyperparameters for bearing RUL prediction. Gao et al. [33] applied Particle Swarm Optimization (PSO) to optimize a Least Squares Support Vector Machine combined with Convolutional Neural Networks (CNNs) for bearing RUL prediction. The Remora Optimization Algorithm (ROA), which simulates the foraging and attachment behaviors of remora fish, is a novel bio-inspired optimization algorithm [16]. ROA excels in global search capability, fast convergence speed, and simple parameter settings, making it particularly suitable for hyperparameter optimization in deep learning models. Chen et al. [34] utilized ROA to compute characteristic parameter values and compared its computational accuracy and speed with the traditional PSO algorithm. Ragab et al. [35] applied ROA to optimize the hyperparameters of the EfficientNetB3 model, demonstrating that ROA-based optimization outperforms existing methods. Alanazi et al. [36] employed ROA for maximum power point tracking in photovoltaic systems, achieving faster tracking speeds compared to PSO and Manta Ray Foraging Optimization (MRFO).

However, the switching mechanism between the remora-based strategy and the whale optimization strategy in the conventional ROA is based on predefined conditions. While this switching mechanism enhances the diversity of the search to some extent, it lacks an effective method for adapting to specific problems. In particular, when dealing with certain optimization stages and problem complexities, the algorithm may fail to dynamically adjust the search strategy, leading to suboptimal performance. Therefore, it is crucial to develop more flexible and adaptive mechanisms to enable the algorithm to adjust its strategies according to the problem. Furthermore, in deep learning models like LSTM, even small changes in hyperparameters can significantly affect model performance, necessitating more fine-tuned and adaptive adjustments.

As summarized in Table 1, most existing studies focus either on data enhancement or on prediction model optimization, while the coordinated optimization of both stages has rarely been considered. This separation may limit the ability of existing methods to fully exploit the interaction between enhanced data quality and downstream RUL prediction performance. Therefore, a jointly optimized framework that simultaneously considers data enhancement and time-series prediction is important for improving the robustness, generalization ability, and prediction accuracy of bearing RUL models. Recent studies have demonstrated continuous progress in condition monitoring, intelligent fault diagnosis, and data-driven prognostics for rotating machinery, providing important technical foundations for bearing RUL prediction [37,38]. A comparative analysis of related works is presented in Table 1.

Table 1. A Summary of the Related Works.

Class	General Idea	Main Issue
Condition monitoring	Torque/speed equilibrium monitoring based on accelerometric signal processing [37]	Not designed for RUL prediction
Bearing fault diagnosis	Convolutional Capsule Network for bearing diagnosis [38]	Focuses on fault diagnosis rather than RUL estimation
Data Enhancement	Generating real damage data with an Auxiliary Classifier GAN [13] Generate new samples by adding noise to the time step [14] Classification method combining multi-scale residual attention and GANs [20]	Insufficient samples and limited variability Time-consuming sampling process Limited-sample condition
	ACWGAN-GP model for minority-class sample generation [22]	Optimizes data generation for imbalanced diagnosis, without coordination with downstream RUL prediction
	Sample-enhanced diagnostic method based on pre-training and ACWGAN-GP [23]	Optimization is confined to data enhancement, without joint optimization with RUL prediction
	Cosine-similarity-based penalty term for bearing data enhancement [24]	Enhancement-stage optimization only, without synchronized tuning of the prediction stage

Table 1. Cont.

Class	General Idea	Main Issue
Time Series Prediction and Intelligent Optimization Algorithms	LSTM framework based on spatial correlation and temporal attention [25]	Verified mainly on simulation data
	Lightweight Bi-LSTM based on automated model pruning [26]	Experimental environments are relatively simplified and well controlled
	LSTM prediction algorithm based on multilayer grid search optimization [28]	Computationally expensive and resource-intensive
	Data-driven framework for estimating RUL of complex systems [31]	Model framework still requires further optimization
	Optimization of LSTM networks using the Slime Mold Algorithm [32]	Optimizes the prediction model only, without coordinated optimization of the data enhancement stage
	Model optimization approach combining CNN with PSO [33]	Focuses on prediction-side modeling, without jointly considering data enhancement and prediction; Case studies have limitations

Based on this analysis, this study aims to construct an efficient bearing degradation state prediction model by combining ROA global optimization with WGAN-GP-based data enhancement and LSTM-based time-series prediction. Different from existing studies that optimize data enhancement and prediction separately, the proposed framework performs coordinated optimization across both stages. By dynamically adjusting hyperparameters and utilizing flexible search strategies, the model efficiently navigates through complex, high-dimensional parameter spaces. This significantly enhances both the effectiveness and efficiency of joint optimization between LSTM and GAN models. Furthermore, these innovative configurations enable the algorithm to outperform traditional ROA methods, overcoming their inherent limitations. This approach effectively addresses the challenges of data scarcity and imbalance in bearing RUL prediction, ultimately leading to improved overall performance of the predictive model.

3. Methodology

The ROA-DE-LSTM network framework proposed in this paper for bearing RUL prediction is illustrated in Figure 1.

The framework collects bearing operation data from the test bench and extracts health indicators (HIs). These indicators are processed through feature extraction and feature selection to generate key inputs for the prediction model. To address the challenges posed by scarce bearing degradation data, the framework integrates a Wasserstein GAN with Gradient Penalty (WGAN-GP) optimized by the Remora Optimization Algorithm (ROA). This approach generates high-quality simulation data, significantly enhancing data diversity and quantity. Based on this, the framework employs ROA to optimize the hyperparameters of the LSTM model, improving learning efficiency and prediction accuracy. The data enhancement module (DE) is integrated into the ROA-LSTM network to further enhance the model's generalization ability. Finally, the framework achieves performance optimization through combined methods and outputs accurate RUL predictions. The proposed framework integrates data enhancement, hyperparameter optimization, and model prediction, significantly improving the accuracy and robustness of RUL predictions.

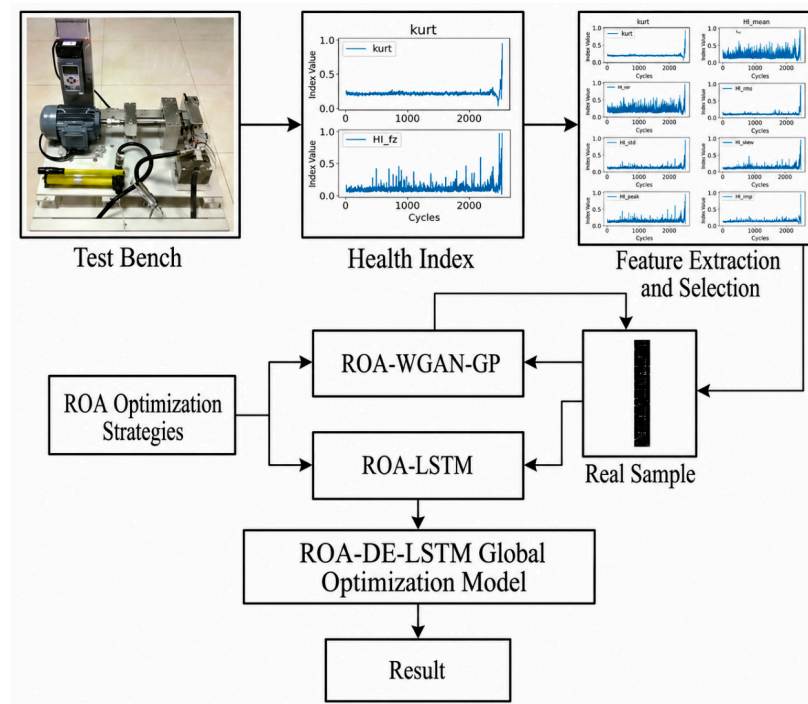


Figure 1. Overview of the proposed rolling bearing RUL prediction framework.

Vibration signals of bearings are typically collected using vibration collectors, which provide essential degradation information. To fully utilize the features in the degradation information, 13 types of time-domain features are extracted based on the existing literature. These features include maximum, mean, variance, standard deviation, root mean square, peak-to-peak value, skewness, kurtosis, waveform factor, peak factor, impulse factor, margin factor, and spectral kurtosis [39]

To estimate the RUL of rolling bearings, a bearing health index (HI) based on spectral kurtosis was constructed [40]. Grey relational analysis is then applied to identify the time-domain features of vibration signals that are highly correlated with the HI, which are used to create a new dataset.

WGAN-GP is used to enhance bearing degradation data, addressing the issue of data scarcity. The enhanced data are then integrated into the prediction model to enrich the dataset. To verify the effectiveness of ROA-based optimization, ROA is used to optimize the hyperparameters of the WGAN-GP network model, and the DE module is constructed by optimizing the learning rates of the model generator and discriminator.

Subsequently, an RUL prediction model is constructed to compare and analyze the enhancement effect of the DE model. Experiments are carried out using LSTM and CNN models on both enhanced and unenhanced datasets to verify the robustness of the DE module. Additionally, ROA is used to globally optimize the LSTM learning rate. Various intelligent optimization algorithms are then selected for comparative analysis to thoroughly validate the effectiveness and superiority of the ROA-DE-LSTM model.

3.1. WGAN-GP Network with ROA Optimization

In this paper, the WGAN-GP network is employed to generate synthetic bearing degradation samples, addressing the issue of data scarcity. The structure of the WGAN-GP network is illustrated in Figure 2.

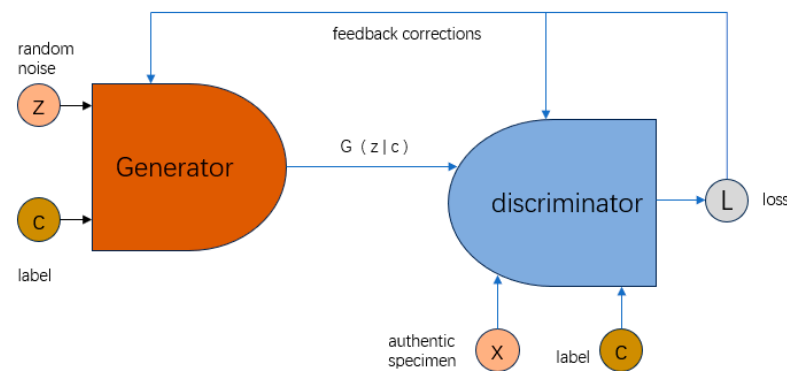


Figure 2. WGAN-GP network structure.

The WGAN-GP network structure consists of a generator and a discriminator, where the generator feedback loss includes both false sample loss and regularization loss, while the discriminator feedback loss consists of true sample loss and gradient penalty loss. By introducing Wasserstein distance loss and gradient penalty, WGAN-GP enhances training stability and alleviates problems such as gradient vanishing and model collapse. To further optimize the data generation process, the Remora Optimization Algorithm (ROA) is integrated into the WGAN-GP framework to dynamically adjust the learning process of both the generator and discriminator, addressing the challenge of optimizing their performance. By combining the global search capabilities of ROA with local optimization of the generator and discriminator learning rates, this approach enables effective switching between whale optimization and sailfish optimization, improving the overall learning efficiency.

This dynamic adjustment allows the GAN to flexibly handle data from different bearing degradation stages, ensuring that the generated data accurately reflects various degradation states. The integration of ROA enhances data diversity and quality by progressively fine-tuning the generated data, addressing data scarcity, and improving the LSTM model's RUL prediction accuracy. This multi-step optimization process improves both the overall data generation and the model's predictive performance.

The adversarial training objectives of the WGAN-GP model optimized by ROA are defined as shown in Equation (1):

$$\begin{cases} \mathcal{L}_D = \mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] - \mathbb{E}_{x \sim p_r}[D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \\ \mathcal{L}_G = -\mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] \end{cases} \quad (1)$$

where $\mathcal{L}_D$ and $\mathcal{L}_G$ denote the discriminator loss and the generator loss, respectively. p_r represents the distribution of real bearing degradation samples, and p_g represents the distribution of generated samples. $\tilde{x} = G(z)$ denotes a generated sample produced by the generator from input noise z . $\hat{x}$ is sampled along the straight line between real and generated samples. The gradient penalty term is introduced to enforce the Lipschitz continuity constraint and improve the training stability of WGAN-GP.

In this framework, ROA is employed as an outer-loop optimizer to tune the generator learning rate and discriminator learning rate of WGAN-GP. It does not change the mathematical form of the WGAN-GP adversarial loss. The generated samples are further evaluated according to their contribution to the downstream LSTM prediction performance.

3.2. ROA-LSTM Optimization Mechanism After Data Enhancement

Bearing vibration signals have strong temporal sequence characteristics, and LSTM can effectively capture long-term dependencies in degradation sequences. In this study, the synthetic samples generated by the ROA-WGAN-GP module are combined with the

original training samples to form an enhanced training set. The LSTM network is then trained on the enhanced dataset to predict the health indicator or RUL value. The ROA is used to tune the learning rate of the LSTM predictor, thereby improving its convergence stability and prediction accuracy under limited training data. The network structure of LSTM is shown in Figure 3.

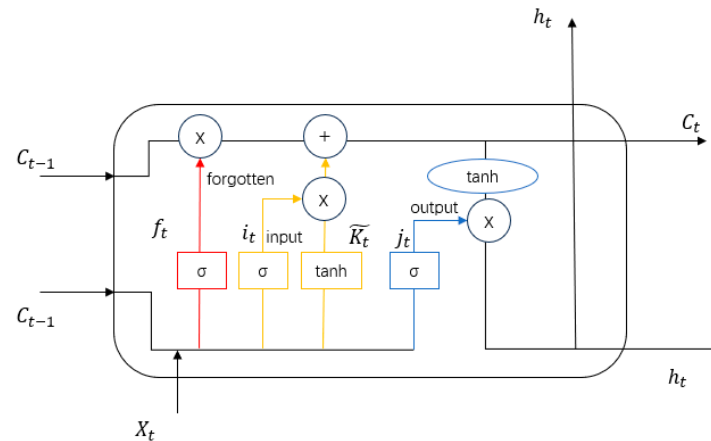


Figure 3. LSTM network structure.

However, the performance of LSTM heavily depends on hyperparameter configurations, and traditional optimization methods often face issues such as local optima and overfitting. To address these problems, this study introduces ROA to optimize the learning rate of the LSTM predictor. In this framework, the ROA-WGAN-GP module focuses on improving the quality and diversity of training samples, while the ROA-LSTM module focuses on improving the accuracy and generalization ability of RUL prediction. The loss function of the LSTM predictor is the mean squared error, as shown in Equation (2):

$$L_{\text{LSTM}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

where n is the number of samples, y_i is the ground-true value of the i -th sample, and $\hat{y}_i$ denotes the corresponding predicted value. This loss is used to train the LSTM predictor. Unlike the WGAN-GP adversarial loss, the LSTM loss directly measures the prediction error of the downstream RUL estimation model.

3.3. Remora Optimization Algorithm

(1) Overview of the Remora Optimization Algorithm (ROA)

The Remora Optimization Algorithm (ROA) is a bio-inspired heuristic optimization algorithm inspired by the feeding behavior of remoras in nature. The algorithm simulates two main phases: the global exploration phase and the local exploitation phase, mimicking the remora's behavior of searching for and attaching to hosts. This algorithm was introduced by Heming Jia et al. in 2021 [17] and is designed to optimize complex problem spaces by balancing exploration and exploitation.

However, a limitation of the traditional ROA is its fixed switching mechanism between the Sailfish Optimization (SFO) and Whale Optimization Algorithm (WOA) strategies, based on preset conditions. While this enhances the diversity of the search, it lacks the flexibility needed to adapt to the specific complexities and optimization stages of certain problems. Therefore, this study introduces an adaptive ROA framework that dynamically adjusts its optimization strategy to better handle the varying stages and complexities of tasks like Remaining Useful Life (RUL) prediction for rolling bearings. This innovation

significantly improves the algorithm's performance in deep learning tasks by enhancing its adaptability during the optimization process.

(2) Mathematical Model and Optimization Process

a. Initialization

In ROA, the initial positions of the population are determined using random initialization, ensuring a diverse starting population that allows the algorithm to effectively explore the solution space. The position of each individual is initialized using Equation (3):

$$x_{i,j} = lb_j + r_{i,j}(ub_j - lb_j), \quad j = 1, 2, \dots, d \quad (3)$$

where $x_{i,j}$ denotes the j -th dimensional component of the i -th remora individual, lb_j and ub_j are the lower and upper bounds of the j -th decision variable, respectively, $r_{i,j} \sim U(0, 1)$ is a uniformly distributed random number, and d is the dimension of the hyperparameter search space. In this study, each remora individual represents a candidate hyperparameter vector to be optimized by ROA.

b. Fitness Function Calculation

The fitness function is designed to evaluate the final prediction performance of the LSTM model after data enhancement. In the ROA optimization process, each remora individual represents a candidate hyperparameter combination, including the generator learning rate, discriminator learning rate, and LSTM learning rate. For each candidate solution, WGAN-GP is first trained using the corresponding generator and discriminator learning rates. The trained generator then produces synthetic degradation samples, which are combined with the original training samples to form an enhanced training set. Subsequently, the LSTM predictor is trained on the enhanced training set using the corresponding LSTM learning rate. Finally, the validation MSE of the trained LSTM predictor is calculated as the fitness value. A smaller fitness value indicates better prediction performance and a more suitable hyperparameter configuration. Therefore, the fitness function used in ROA optimization is defined as the validation MSE of the LSTM predictor, as follows:

$$\xi(X_i) = \frac{1}{n_v} \sum_{i=1}^{n_v} (y_i - \hat{y}_i)^2 \quad (4)$$

where $X_i = [\eta_G, \eta_D, \eta_L]$ denotes the i -th candidate hyperparameter vector. η_G , η_D , and η_L represent the learning rates of the generator, discriminator, and LSTM predictor, respectively. n_v is the number of validation samples, y_j is the true value, and $\hat{y}_j$ is the predicted value.

The fitness function evaluates the final RUL prediction performance of the LSTM model on the validation set. For each candidate hyperparameter combination, WGAN-GP is first trained using the corresponding generator and discriminator learning rates. The trained generator then produces synthetic degradation samples, which are combined with the original training samples to form an enhanced training set. Subsequently, the LSTM predictor is trained using the corresponding LSTM learning rate. A smaller fitness value indicates better prediction performance and a more suitable hyperparameter combination. It should be noted that Equation (4) is not a weighted combination of the WGAN-GP loss and the LSTM loss. Instead, ROA jointly searches the hyperparameters of the data enhancement and prediction modules using the validation MSE of the downstream LSTM predictor as the unified evaluation objective.

c. SFO and WOA Strategies

When remoras attach to fast hosts such as sailfish, the algorithm uses the optimal individual guidance mechanism from the Sailfish Optimization (SFO) strategy, augmented with a random selection rule. The position update equation is as follows:

$$X_i^{t+1} = X_{\text{best}}^t - [r \cdot 2X_{\text{best}}^t + X_{\text{rand}}^t - X_{\text{rand}}^t], r \sim U(0, 1) \quad (5)$$

where t is the current iteration, X_{best}^t is the current optimal remora individual, and X_{rand}^t is the current random remora individual. $r \sim U(0, 1)$ is a random coefficient.

In contrast, when remoras attach to larger hosts like whales, ROA adopts the Whale Optimization Algorithm (WOA) with a spiral search strategy for updating positions. The mathematical models for WOA are:

$$X_i^{t+1} = D_i e^{l} \cos(2\pi l) + X_{\text{best}}^t \quad (6)$$

$$D_i = X_{\text{best}}^t - X_i^t \quad (7)$$

$$l = (a - 1)r + 1, r \sim U(0, 1) \quad (8)$$

$$a = -\left(1 + \frac{t}{T}\right) \quad (9)$$

where D_i denotes the distance between the current individual and the current best individual, l is a spiral random coefficient, $r \sim U(0, 1)$, and a decreases linearly from -1 to -2 as the iteration number t increases from 0 to T .

The dynamic switching between SFO and WOA strategies in traditional ROA is based on preset conditions, but in this study, we propose an adaptive mechanism that dynamically adjusts between the two strategies according to the task's needs and optimization stage.

d. Global Exploration Stage

During the global exploration stage, remoras need to constantly make small movements around their host to accumulate experience and determine whether a host change is necessary. The formula for this experience accumulation is:

$$X_{\text{att}} = X_i^t + (X_i^t - X_{\text{pre}}) \times \text{randn} \quad (10)$$

where X_{att} is a tentative remora movement, X_{pre} is the position of the previous remora, and randn is a random number with a normal distribution.

To determine whether a host change is required, the following conditions are used:

$$f(X_i^t) > f(X_{\text{att}}) \quad (11)$$

$$H(i) = \text{round}(\text{rand}) \quad (12)$$

where $H(i)$ decides whether the remora attaches to a whale ($H(i) = 0$) or a sailfish ($H(i) = 1$). The fitness function values $f(X_i^t)$ and $f(X_{\text{att}})$ are compared to make this decision. The innovation in this step is the dynamic adjustment of the host-switching criteria, allowing the algorithm to adapt based on the optimization stage and complexity.

e. Local Exploitation Stage

In the local exploitation phase, the search range is reduced, and remoras optimize their positions by using the resources around the host. The update formula is:

$$X_i^{t+1} = X_i^t + A \quad (13)$$

where A represents the distance traveled by the remora, and it is computed as:

$$A = B \times (X_i^t - C \times X_{\text{best}}) \quad (14)$$

$$B = 2 \times V \times \text{rand} - V \quad (15)$$

$$V = 2 \left(1 - \frac{t}{T} \right) \quad (16)$$

where B and V represent the volumes of the host and remora, respectively. Remora factor C is used to limit the remora's position. This phase helps refine the solution by exploiting the local search region.

(3) ROA with GAN and LSTM Optimization Coupling Mechanism

a. Data Augmentation with GAN Optimization

To address the data scarcity issue in RUL prediction, WGAN-GP is used to generate high-quality synthetic data for training. ROA optimizes the hyperparameters of both the generator and discriminator to enhance data diversity and quality. The adaptive strategy dynamically adjusts the learning rates of the generator and discriminator, enabling the generation of more diverse and realistic samples that better represent the degradation stages of rolling bearings.

b. LSTM Model Optimization

LSTM networks are used to predict the RUL of bearings based on vibration signal data. The performance of LSTM is highly sensitive to hyperparameters. ROA optimizes the learning rate of the LSTM predictor through both global exploration and local exploitation strategies. The adaptive mechanism ensures that hyperparameters are dynamically adjusted based on the task complexity, improving the model's training stability and convergence.

(4) Algorithm Workflow

The ROA-DE-LSTM model integrates ROA with the data enhancement module to optimize the hyperparameters of WGAN-GP and LSTM for improved RUL prediction. The workflow begins with initializing the remora population in the search space, where each remora represents a three-dimensional candidate solution consisting of the generator learning rate, discriminator learning rate, and LSTM learning rate. For each candidate solution, WGAN-GP is first trained to generate synthetic degradation samples, and the LSTM predictor is then trained on the enhanced training set. The validation MSE of the trained LSTM predictor is used as the fitness value. Based on these evaluations, the remora positions are updated using the SFO and WOA strategies, and the optimization continues until the maximum number of iterations is reached. The best solution represents the optimized hyperparameter combination for the data enhancement and prediction modules. The overall process is summarized in Algorithm 1.

Algorithm 1 Network Model Hyperparameter Optimization Procedure.

1. **Input:** Training set, validation set, search ranges, population size N , maximum iterations T
 2. **Output:** Optimal solution $X_i = [\eta_G, \eta_D, \eta_L]$
 3. Start
 4. Initialize N Remoras with 3D positions:
 5. $X_i = [\eta_G, \eta_D, \eta_L]$
 6. Evaluate fitness for each X_i
 7. Train WGAN-GP with η_G and η_D .
 8. Generate synthetic degradation samples using the trained WGAN-GP
 9. Combine the generated samples with the original training samples and train LSTM using η_L
-

Algorithm 1 *Cont.*

```

10.   Fitness( $X_i$ ) = validation MSE of the LSTM predictor
11.    $X_{best} = \operatorname{argmin} \operatorname{Fitness}(X_i)$ 
12.   Set  $t = 0$ 
13.   While  $t < T$  do
14.       For each Remora  $i = 1$  to  $N$  do
15.           Check boundary constraints for current position and repair if violated
16.           Calculate fitness ( $i$ ) using Equation (4)
17.           Update  $X_{best}$  if  $\operatorname{fitness}(i) < \operatorname{fitness}(X_{best})$ 
18.           Update parameters  $k$ ,  $a$ ,  $V$  using Equations (8), (9) and (16)
19.           If  $H_i == 0$  then
20.               Update the position according to the WOA strategy using Equation (6)
21.           Else if  $H_i == 1$  then
22.               Update the position according to the SFO strategy using Equation (5)
23.           End if
24.           Make host-step prediction using Equation (10)
25.           Determine host replacement using Equations (11) and (12)
26.           If host is not replaced then
27.               Apply host feeding mode using Equation (13)
28.           End if
29.       End for
30.        $t = t + 1$ 
31.   End while
32.   Return  $X_{best}$ 

```

4. Test Examples and Performance Metrics

4.1. Health Indicator Construction

The experimental dataset used in this study is obtained from the bearing life test rig at Xi'an Jiaotong University [40], as shown in Figure 4. The dataset consists of horizontal and vertical vibration signals from bearings under a 40 Hz/10 kN operating condition. Two datasets, Bearing 3-1 and Bearing 3-4, are selected to validate the effectiveness of the proposed ROA-DE-LSTM framework. To simulate the limited-data condition commonly encountered in industrial applications, only Bearing 3-1 is used for model development, while Bearing 3-4 is reserved as an independent test bearing.

The horizontal vibration signal of Bearing 3-1 is used for health indicator construction, feature extraction, model training, validation, and data enhancement. After time-domain feature extraction and sliding-window segmentation, 2500 original samples are constructed from Bearing 3-1. These samples are divided into 2000 training samples and 500 validation samples following an 8:2 split ratio. The validation set is used exclusively for hyperparameter selection and early stopping.

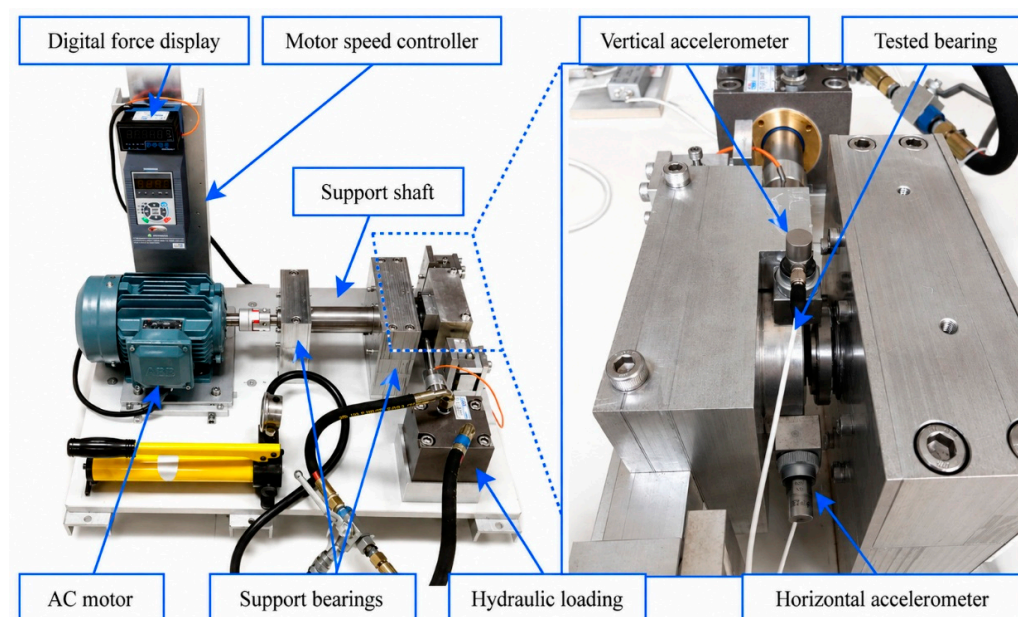


Figure 4. XJTU-SY Bearing Life Test Bench.

Data enhancement is performed using ROA-WGAN-GP based solely on the training samples of Bearing 3-1. In the proposed setting, 2500 synthetic samples are generated accordingly. To analyze the effect of generated data size, two augmentation strategies are considered: full generated data, in which all 2500 synthetic samples are added to the training set, and half generated data, in which 1250 generated samples are randomly selected for training. Therefore, the final training sets consist of 2000 original samples for the baseline model, 3250 samples for the half-generation setting, and 4500 samples for the full-generation setting. The data usage protocol is summarized in Table 2.

Table 2. Usage of the XJTU-SY dataset across different stages of the proposed framework.

Processing Stage	Bearing	Signal Used	Purpose	Leakage Control
Feature extraction	Bearing 3-1	Horizontal vibration signal	Construct HI	Only Bearing 3-1 is used
Training and validation	Bearing 3-1	Horizontal vibration signal	Train the models	Divided into 2000 training samples and 500 validation samples No Bearing 3-4 data are used
Data augmentation	Generated from Bearing 3-1	Synthetic samples	Expand training data	Generated only from the training samples
Testing	Bearing 3-4	Horizontal vibration signal	Independent evaluation	Excluded from training validation, and augmentation

To prevent data leakage, the train/validation/test split is performed before data augmentation. Bearing 3-4 is not used in WGAN-GP training, ROA-LSTM training, validation, hyperparameter tuning, early stopping, or normalization-statistics estimation. All generated samples are produced only from the training samples of Bearing 3-1 and are added only to the training set. The validation and test sets contain only original samples. In addition, the consistency between original and generated samples is evaluated using statistical metrics such as the Population Stability Index (PSI).

In this study, the limited-data setting refers to the case where only the full-life degradation data of a single bearing are available for model development. All comparative models are evaluated under the same protocol, with the same training samples, validation samples, test bearing, model architecture, hyperparameter search space, random seed, early-stopping strategy, and gradient-clipping strategy.

Using the horizontal vibration signal as test data, a health indicator (HI)-based prediction model is constructed to more accurately reflect the degradation trend and remaining useful life (RUL) of bearings. The spectral kurtosis-based HI is constructed using prior knowledge, as described in Equation (17):

$$HI = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right)^2} \quad (17)$$

Given the large volume of detection data, normalization and gray correlation analysis are applied to preprocess the data and extract features. Features exhibiting high correlation coefficients (greater than 0.8) with HI are selected for the new dataset. Table 3 presents the gray correlation analysis results for the extracted time-domain features, including maximum (max), minimum (min), variance (var), root mean square (rms), skewness (skew), and waveform factor (bx).

Table 3. Gray correlation analysis results.

	max	min	var	std	rms	ffz
	0.641	0.539	0.708	0.61	0.61	0.609
	bx	mc	yd	skew	kurt	fz
HI	0.794	0.917	0.652	0.603	0.947	0.919
	HI_var	HI_std	HI_rms	HI-ffz	HI_fz	HI_yd
	0.818	0.856	0.756	0.852	0.887	0.92

The analysis shows that the eight features, including vibration signal features (kurtosis (kurt), peak (fz), and impulse (mc)) and health indicators (variance (HI_var), standard deviation (HI_std), peak (HI_fz), peak-to-peak (HI_ffz), and margin (HI_yd)), are particularly effective in capturing the bearing degradation process, as illustrated in Figure 5. Key features like HI_std and HI_ffz show pronounced peaks approaching failure, making them especially useful for detecting early degradation. Variance (HI_var) and kurtosis (Kurt) display abrupt changes, identifying sudden shifts in system conditions. This further demonstrates the reliability of the health indicators in reflecting bearing health.

4.2. Model Parameter Configuration

The detailed configurations and structural settings of the WGAN-GP-based data enhancement model and the LSTM prediction model are summarized in Table 4.

Table 4. Model parameters for WGAN-GP and LSTM.

Model	Configuration Parameters	Structural Parameters
WGAN-GP	Batch size: 16 Loss function: MSE G/D Optimizer: Adam	Generator G: (Linear, BatchNorm1d) × 3 + Sigmoid Discriminator D: Linear × 3 + L1/L2 + Sigmoid

Table 4. Cont.

Model	Configuration Parameters	Structural Parameters
LSTM	Batch size: 16 Loss function: MSE Optimizer: Adam	LSTM: (130, dropout) $\times$ 2 Dense
CNNRegressor	Batch size: 64 Loss function: MSE Momentum: 0.9 Weight-decay: 0.0001	(Conv1d, BatchNorm1d) $\times$ 3 + Conv1d

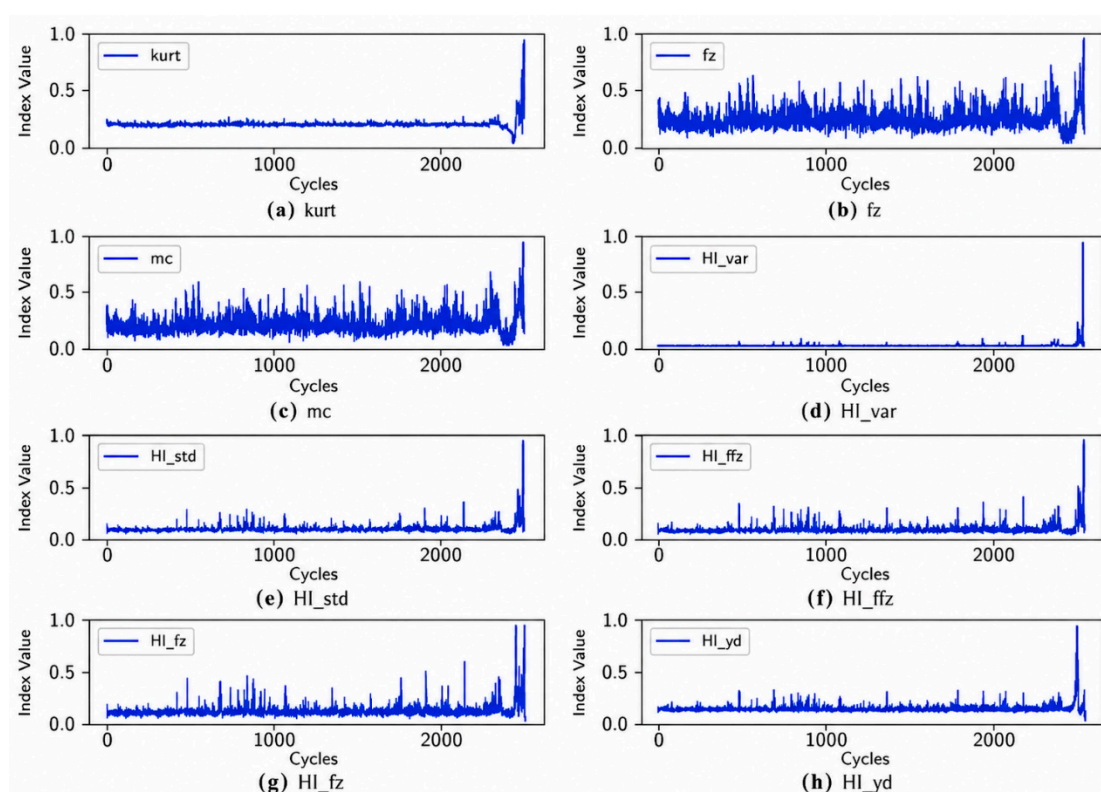


Figure 5. Feature extraction results for Bearing 3-1: (a) kurtosis (kurt); (b) peak value (fz); (c) impulse factor (mc); (d) health-index variance (HI_var); (e) health-index standard deviation (HI_std); (f) health-index peak-to-peak value (HI_ffz); (g) health-index peak value (HI_fz); and (h) health-index margin factor (HI_yd).

4.3. ROA Optimization

To ensure reproducibility and fairness of the comparative optimization experiments, PSO [41], HHO [42], and ROA are configured under the same optimization budget. All three algorithms optimize identical hyperparameters, including the generator learning rate and discriminator learning rate of WGAN-GP, as well as the learning rate of the LSTM predictor, as summarized in Table 5.

In the optimization process, each candidate solution is represented as a three-dimensional hyperparameter vector (LSTM learning rate, generator learning rate, discriminator learning rate). For each candidate solution, WGAN-GP is first trained using the corresponding generator and discriminator learning rates to generate synthetic degradation samples. The generated samples are then combined with the original training samples to train the LSTM predictor using the corresponding LSTM learning rate. The validation MSE of the trained LSTM predictor is used as the fitness value. Therefore, the optimization process jointly searches for hyperparameters of the data enhancement and prediction modules under the same evaluation objective. The search ranges, population size, maximum

number of iterations, internal training epochs, and independent runs are kept identical for all algorithms.

Table 5. Optimization settings of different metaheuristic algorithms.

Algorithm	Optimized Hyperparameters	Search Range	Population Size	Max Iterations	Internal Training Epochs	Random Seeds
PSO	LSTM learning rate; WGAN-GP generator and discriminator learning rates	$[1 \times 10^{-5}, 1 \times 10^{-1}]$ for LSTM; $[1 \times 10^{-5}, 1 \times 10^{-2}]$ for WGAN-GP	30	100	10	10
HHO	LSTM learning rate; WGAN-GP generator and discriminator learning rates	$[1 \times 10^{-5}, 1 \times 10^{-1}]$ for LSTM; $[1 \times 10^{-5}, 1 \times 10^{-2}]$ for WGAN-GP	30	100	10	10
ROA	LSTM learning rate; WGAN-GP generator and discriminator learning rates	$[1 \times 10^{-5}, 1 \times 10^{-1}]$ for LSTM; $[1 \times 10^{-5}, 1 \times 10^{-2}]$ for WGAN-GP	30	100	10	10

The stopping criterion of each optimization algorithm is reaching the maximum number of iterations. Each algorithm performs an identical number of fitness evaluations under the same internal training setting. To reduce the influence of randomness, each optimization algorithm is independently repeated ten times using different random seeds, and the same random seed is used across PSO, HHO, and ROA in the corresponding independent run. The final optimization results are reported based on the average performance of the repeated runs, ensuring that the comparison is primarily determined by the optimization strategy itself rather than differences in search range, computational budget, or parameter configuration.

4.4. Experimental Settings

The detailed training configuration, including the input setting, WGAN-GP architecture, LSTM structure, optimizer, loss function, and regularization-related operations, is summarized in Table 6. Under this unified experimental protocol, the network structure, data construction strategy, training configuration, and optimization budget remain unchanged for all compared methods. As a result, the proposed ROA improves performance through optimization rather than additional computational resources or inconsistent experimental settings.

Table 6. Model and training parameter configuration.

Module	Parameter	Setting
Input data	Input feature type	Selected degradation-related features from the horizontal vibration signal
Input data	Sliding window length	50
Input data	Sliding step	1
WGAN-GP	Batch size	16
WGAN-GP	Optimizer	Adam
WGAN-GP	Generator architecture	Linear + BatchNorm1d blocks followed by activation
WGAN-GP	Discriminator architecture	Linear layers with regularization

Table 6. Cont.

Module	Parameter	Setting
LSTM	Batch size	16
LSTM	Hidden units	130
LSTM	Number of LSTM layers	2
LSTM	Optimizer	Adam
LSTM	Loss function	Mean squared error
Training	Gradient clipping	Used if applied in the final implementation

4.5. Performance Indicators

To evaluate the effectiveness of the proposed ROA-DE-LSTM framework, performance indicators such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-Square (R^2) are selected for comparative analysis [43]. The calculations for these indicators are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \right) \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

5. Result and Discussion

5.1. Comparison of WGAN-GP-Based Optimization

This section presents the results of applying different intelligent optimization algorithms to optimize the WGAN-GP model for data enhancement. The performance is evaluated using mean squared error (MSE) as the loss function. The total generator loss (Loss G) is computed by combining the fake sample loss and the regularization loss. The total discriminator loss (Loss D) is computed by combining the discriminator loss, regularization loss, and gradient penalty for both real and fake samples.

In the experiments, Bearing 3-1 is used as the research object. The baseline WGAN-GP model is compared with the models optimized using Particle Swarm Optimization (PSO), Harris Hawk Optimization (HHO), and Remora Optimization Algorithm (ROA). The training process runs for approximately 3000 iterations. The performance of different optimization algorithms is evaluated by analyzing the loss convergence of both the generator and the discriminator.

(1) Generator Performance

In the generator convergence analysis, the ROA-optimized model shows better stability and faster convergence speed, as illustrated in Figure 6.

The initial maximum loss of the ROA-based model is approximately 0.5, which is significantly lower than that of the baseline model (5.0). The ROA model converges in approximately 200 iterations, compared with 500 iterations for the baseline model. After 2000 iterations, the average loss of the ROA model is 0.02, which is lower than that of the PSO- and HHO-optimized models, whose average losses are 0.05 and 0.03, respectively. These results indicate that the ROA-optimized generator achieves faster convergence and improved training stability.

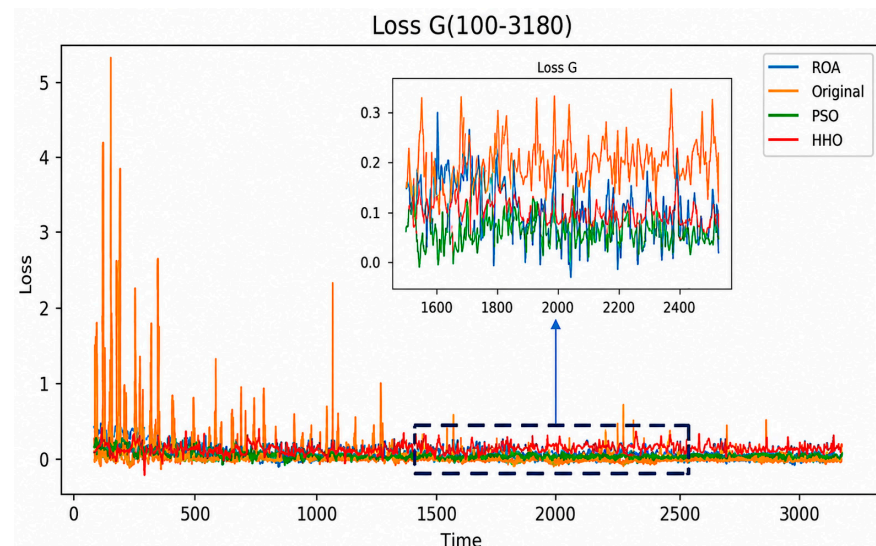


Figure 6. Data-Enhanced Model Generator Training Process.

(2) Discriminator Performance

In the discriminator convergence analysis, the ROA-optimized model also performs better, as shown in Figure 7. The initial loss fluctuation is limited to a maximum of 0.3, which is lower than the fluctuations observed in the PSO and HHO models, with maximum values of 0.6 and 0.7, respectively. The ROA model reaches stability within 300 iterations, while the baseline model requires approximately 700 iterations. The average loss at the final stable stage is 0.01, which is significantly lower than the baseline model's value of 0.08. These findings confirm that the ROA-optimized discriminator has better stability, consistency, and discriminatory ability.

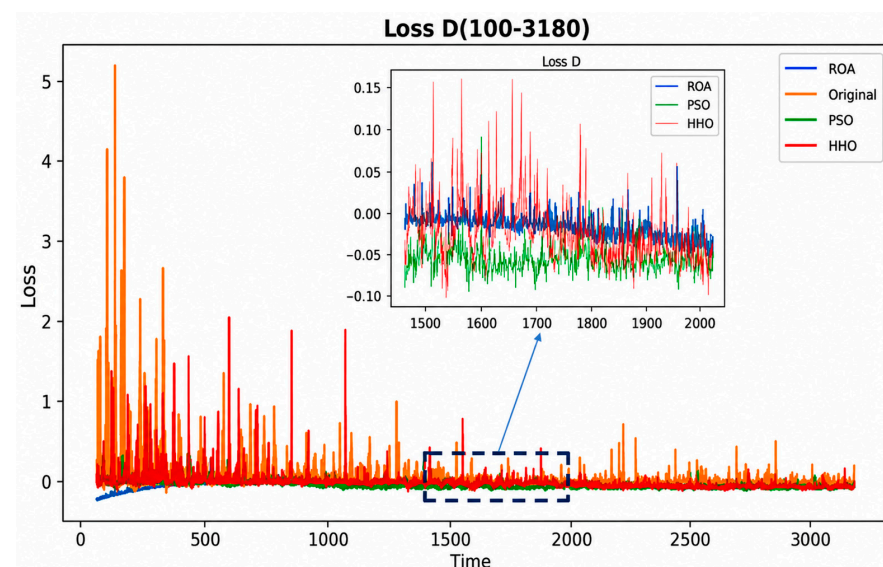


Figure 7. Data-Enhanced Model Discriminator Training Process.

(3) Quantitative Comparison

Table 7 compares the average losses of different WGAN-GP optimization strategies. The relative difference, denoted as *gr*, indicates the deviation of each model from ROA in decimal form. A positive “*gr*” value indicates that the proposed method outperforms the reference object on this metric, whereas a negative value indicates underperformance.

Table 7. Model total loss and gradient penalty term.

	Loss G		D_GP		Loss D	
	Avg	gr	Avg	gr	Avg	gr
Origin	0.168	+0.055	0.160	+0.153	0.157	+0.119
PSO	0.074	−0.039	0.013	+0.006	0.051	+0.013
HHO	0.154	+0.041	0.050	+0.043	0.064	+0.026
ROA	0.113	+0	0.007	+0	0.038	+0

ROA achieves the lowest discriminator loss and gradient penalty, indicating improved training stability and better control of the gradient penalty term. Although PSO achieves a lower generator loss, ROA provides a better overall balance for WGAN-GP optimization and supports more stable data enhancement for RUL prediction.

(4) Distribution Similarity Analysis

Although the generator and discriminator loss curves demonstrate the convergence of the WGAN-GP training process, they do not directly indicate whether the generated degradation samples preserve the statistical characteristics of the original data. Therefore, multiple distribution similarity metrics are further employed to quantitatively evaluate the generated samples, including the Population Stability Index (PSI), Kolmogorov–Smirnov (KS) statistic, Wasserstein distance, and maximum mean discrepancy with the radial basis function kernel (MMD-RBF).

As summarized in Table 8, the generated samples exhibit high similarity to the original degradation data according to the PSI, KS statistic, Wasserstein distance, and MMD-RBF metrics. The PSI values remain within 0.11–0.24, indicating that the distribution differences are within a stable or moderately stable range. The KS statistic ranges from 0.04 to 0.15, indicating that the cumulative distributions of the original and generated samples are close. In addition, the Wasserstein distance remains below 0.10 and the MMD-RBF value remains below 0.16, indicating that both the overall distribution distance and the higher-order feature discrepancy are small.

Table 8. Distribution similarity between the original and generated degradation samples.

Metric	Value	Description
PSI	0.11–0.24	Distribution stability
KS statistic	0.04–0.15	Cumulative distribution similarity
Wasserstein distance	<0.10	Distribution distance
MMD-RBF	<0.16	Feature distribution discrepancy
PCA explained variance	81.09%	Variance explained by the first two principal components

Figure 8 further illustrates the values of the four similarity metrics across different statistical features. Overall, the generated samples exhibit minor deviations from the original data across all evaluated features. In particular, the Wasserstein distance and MMD-RBF values remain consistently low, indicating that both the overall data distribution and the feature-level distribution are well preserved. These results demonstrate that the proposed WGAN-GP effectively learns the intrinsic statistical characteristics of the bearing degradation process while maintaining sufficient diversity for data enhancement.

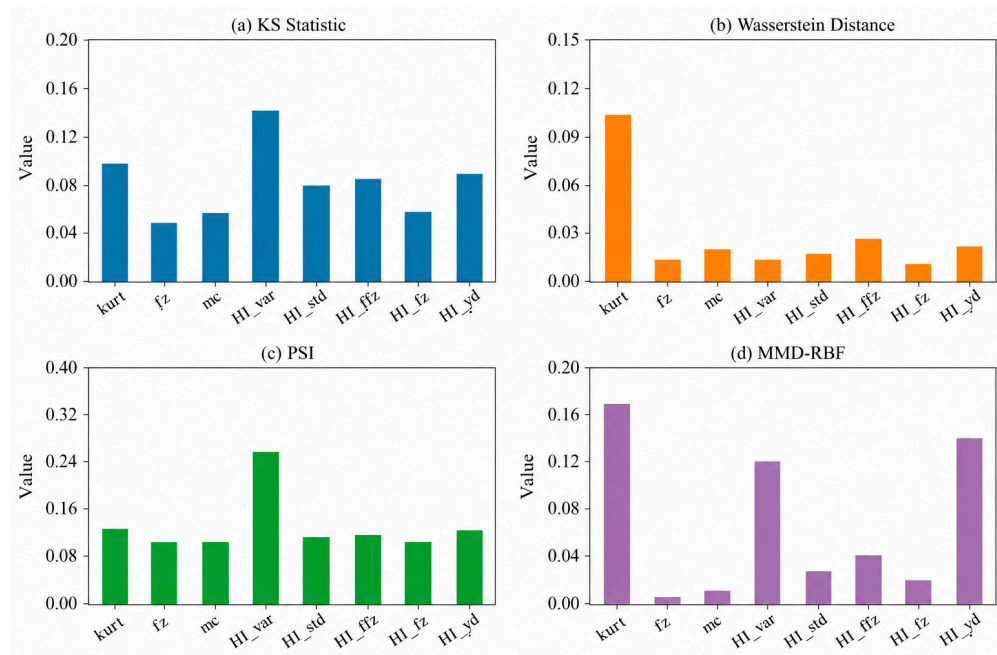


Figure 8. Distribution Similarity Analysis of the Original and Generated Degradation Data.

(5) Feature-space Visualization

To further verify whether the generated samples preserve the intrinsic feature distribution of the original degradation data, principal component analysis (PCA) is performed. The first two principal components explain 81.09% of the total variance, indicating that the two-dimensional PCA space can represent the main structure of the degradation features. As illustrated in Figure 9, the generated samples largely overlap with the original samples in the principal component space while maintaining appropriate diversity. This result further confirms that the proposed WGAN-GP successfully captures the underlying degradation characteristics of the bearing degradation process.

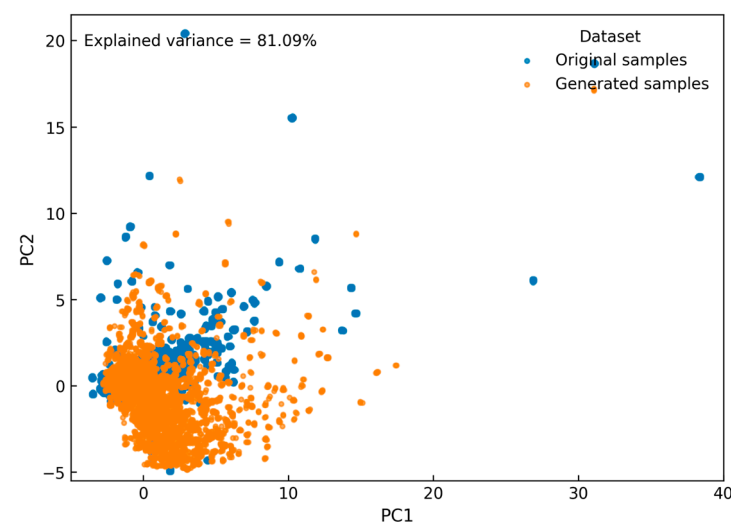


Figure 9. PCA visualization of the original and generated degradation samples.

5.2. ROA-DE-LSTM Prediction Optimization Comparison

To validate the effectiveness of the proposed ROA-DE-LSTM framework, we compare it with multiple prediction methods. CNN and LSTM are used as base models, and PSO, HHO, and ROA are further employed to optimize the data-enhanced LSTM model. Given the long-term periodicity of the health indicator, the sliding window method is applied in

the LSTM network. This method dynamically reads the data with a window size of 50 and a step size of 1 until the entire dataset is processed.

All comparative models are evaluated under the same data usage protocol. Bearing 3-1 is used for model training and validation, while Bearing 3-4 is used as the independent test bearing. The test bearing is not involved in WGAN-GP training, model training, hyperparameter tuning, early stopping, or normalization-statistics estimation. Therefore, the prediction results can be used to evaluate the generalization ability of the proposed method under an independent bearing degradation process.

(1) Network Model Selection and Data Enhancement

After completing data enhancement, independent experiments are conducted on CNN and LSTM models with and without data enhancement. As shown in Figure 10, all models can track the true degradation trend to a certain extent, but their prediction accuracy and stability are different. The LSTM model without enhancement shows larger prediction deviations in several degradation stages. The DE-LSTM model produces a smoother prediction curve and follows the true values more closely in the middle degradation stage. In the late stage, all models show some deviations due to the rapid degradation of the bearing, but the data-enhanced models still provide more stable prediction trajectories.

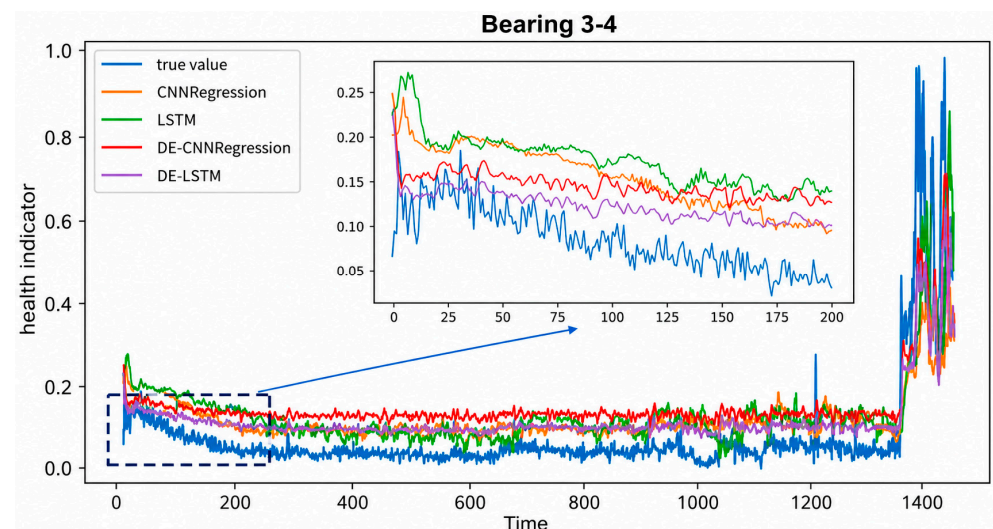


Figure 10. Comparison of Model Health Indicator Predictions.

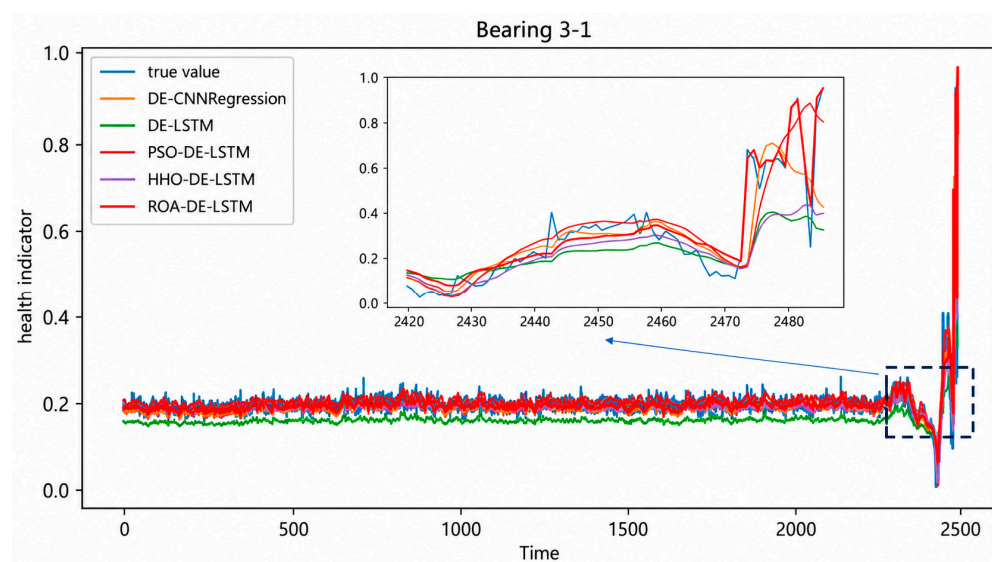
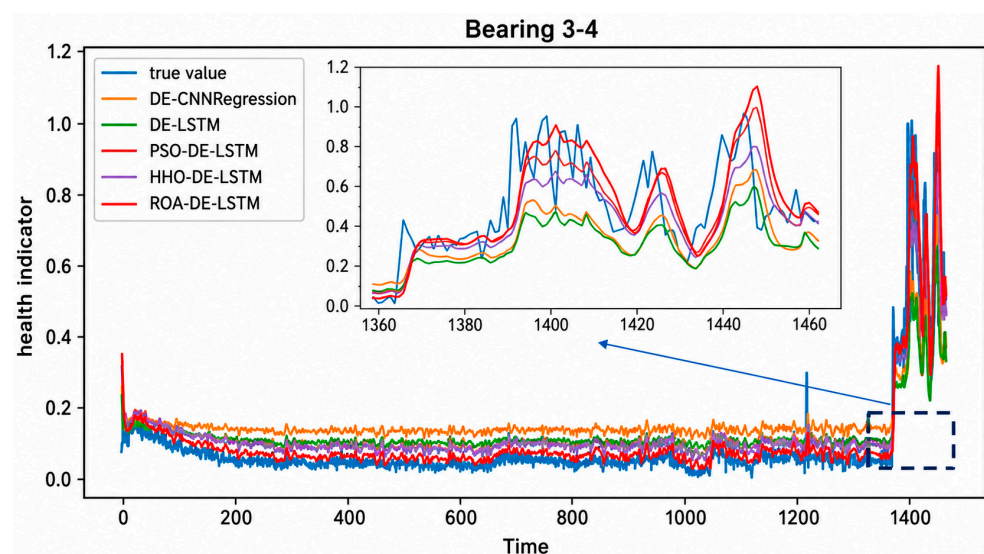
As shown in Table 9, data enhancement improves the prediction performance of both CNNRegression and LSTM models. Compared with CNNRegression, DE-CNNRegression reduces MAE from 0.096 to 0.074, while R^2 increases from 0.300 to 0.460, indicating that the generated degradation samples improve the prediction accuracy and fitting ability of the CNN-based model. Compared with the original LSTM, DE-LSTM slightly increases MAE from 0.062 to 0.063, but reduces RMSE from 0.099 to 0.097 and improves R^2 from 0.477 to 0.498. This indicates that data enhancement improves the overall fitting ability, although its effect on MAE is limited. In addition, DE-LSTM outperforms DE-CNNRegression in all four evaluation metrics, with lower MAE, RMSE, and MAPE and a higher R^2 value. This result shows that, under the same data enhancement strategy, the LSTM network has stronger capability in modeling temporal degradation patterns than the CNNRegression model. Therefore, the DE-LSTM structure is selected as the basic prediction network for the subsequent optimization comparison.

Table 9. Comparison of indicators for predicting degradation status.

	MAE		RMSE		MAPE		R ²	
	Value	gr	Value	gr	Value	gr	Value	gr
CNNRegression	0.096	+0.033	0.114	+0.017	256.62	+76.30	0.300	−0.198
LSTM	0.062	−0.001	0.099	+0.002	184.98	+4.66	0.477	−0.021
DE-CNNRegression	0.074	+0.011	0.100	+0.003	211.87	+31.55	0.460	−0.038
DE-LSTM	0.063	+0	0.097	+0	180.32	+0	0.498	+0

(2) Comparison of Optimization Algorithms in Data-Enhanced Models

After comparing the basic network models, we further evaluate the effects of different optimization algorithms on the DE-LSTM network. The validation results on the training and test sets are shown in Figures 11 and 12. The stability of each optimized DE prediction model is significantly improved, but clear differences can still be observed in the later degradation stage. The ROA-DE-LSTM model produces predictions closest to the true values, while the other optimized models show larger deviations and local fluctuations.

**Figure 11.** Training Set Health Indicator Prediction.**Figure 12.** Test Set Health Indicator Prediction.

To evaluate the statistical stability of different prediction models, all experiments are independently conducted over five runs under identical experimental settings. The results in Table 10 are presented as the mean values with their corresponding standard deviations. As shown in Table 10, the proposed ROA-DE-LSTM achieves the best overall prediction performance among all compared methods, with the lowest mean MAE of 0.035 ± 0.001 , and the highest mean R^2 of 0.846 ± 0.002 . In addition, the relatively small standard deviations across five independent runs indicate that the proposed method provides stable and consistent prediction performance.

Table 10. Results reported as the mean $\pm$ standard deviation over five independent runs.

	MAE		RMSE		MAPE		R^2	
	Value	gr	Value	gr	Value	gr	Value	gr
LSTM	0.062 ± 0.002	+0.027	0.099 ± 0.003	+0.046	184.98 ± 0.31	+68.94	0.477 ± 0.002	−0.369
DE-LSTM	0.063 ± 0.002	+0.028	0.097 ± 0.002	+0.044	180.32 ± 0.28	+64.28	0.498 ± 0.002	−0.348
PSO-DE-LSTM	0.050 ± 0.001	+0.015	0.064 ± 0.001	+0.011	195.73 ± 0.36	+79.69	0.773 ± 0.003	−0.073
HHO-DE-LSTM	0.048 ± 0.002	+0.013	0.063 ± 0.003	+0.010	183.04 ± 0.32	+67.00	0.782 ± 0.002	−0.064
ROA-DE-LSTM	0.035 ± 0.001	+0	0.053 ± 0.002	+0	116.04 ± 0.22	+0	0.846 ± 0.002	+0

(3) Ablation Study

To further verify the contribution of the WGAN-GP data enhancement module, an ablation study is conducted under identical experimental settings. Based on the data enhancement strategy described in Section 4.1, the original training set contains 2000 samples from Bearing 3-1. The full-generation setting adds all 2500 synthetic samples to the original training set, resulting in 4500 training samples. The half-generation setting randomly selects 1250 synthetic samples and adds them to the original training set, resulting in 3250 training samples. These settings are used to evaluate whether the generated degradation samples are beneficial to prediction performance and whether the amount of generated data affects the final prediction results.

Five model configurations are evaluated in the ablation study. The LSTM model denotes the baseline prediction model without data enhancement and without ROA optimization, using only the 2000 original training samples. The DE-LSTM model denotes the prediction model with full generated samples but without ROA optimization, where 2500 generated samples are added to the original training set and the final training set contains 4500 samples. The ROA-LSTM model denotes the prediction model optimized by ROA but without data enhancement, using only the 2000 original training samples. The ROA-DE-LSTM-half model denotes the ROA-optimized prediction model trained with the half-generation setting, in which 1250 generated samples are added and the final training set contains 3250 samples. The ROA-DE-LSTM model denotes the complete proposed model with ROA global optimization and full generated data, where all 2500 generated samples are added and the final training set contains 4500 samples. The quantitative results are summarized in Table 11.

Table 11. Ablation study of different model configurations.

Model Configuration	MAE		RMSE		MAPE (%)		R^2	
	Value	gr	Value	gr	Value	gr	Value	gr
LSTM	0.062	+0.027	0.099	+0.046	184.98	+68.94	0.477	−0.369
DE-LSTM	0.063	+0.028	0.097	+0.044	180.32	+64.28	0.498	−0.348
ROA-LSTM	0.050	+0.015	0.082	+0.029	155.77	+39.73	0.642	−0.204
ROA-DE-LSTM-half	0.046	+0.011	0.079	+0.026	149.77	+33.73	0.667	−0.179
ROA-DE-LSTM	0.035	+0	0.053	+0	116.04	+0	0.846	+0

As shown in Table 11, Figures 13 and 14, WGAN-GP-based data enhancement alone provides limited improvement for the baseline LSTM. The DE-LSTM model slightly reduces RMSE and improves R^2 . This indicates that generated samples alone cannot fully improve the prediction ability of the model without an effective optimization strategy.

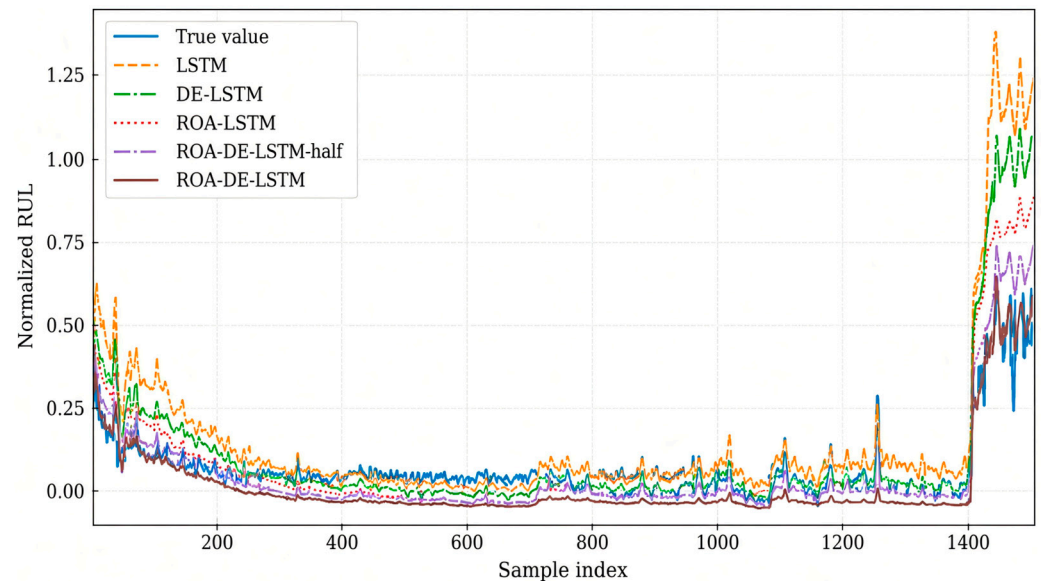


Figure 13. Prediction curves with and without generated samples.

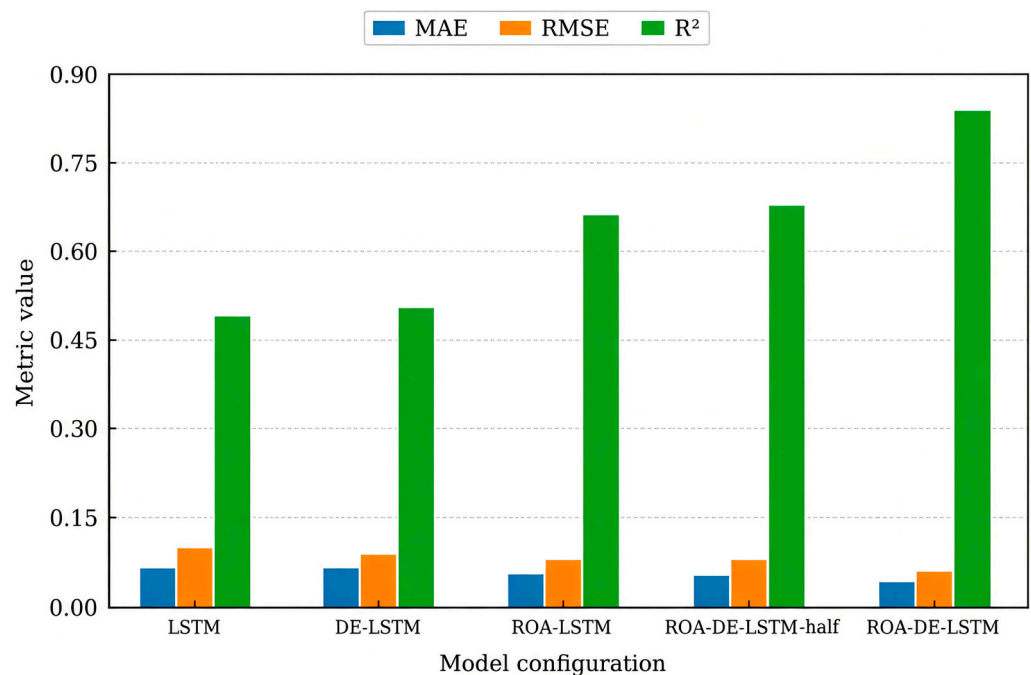


Figure 14. Prediction error distributions with and without generated samples.

In contrast, ROA optimization significantly improves the prediction performance of the LSTM model. Compared with the baseline LSTM, ROA-LSTM reduces RMSE from 0.099 to 0.082, while R^2 increases from 0.477 to 0.642. These results indicate that ROA optimization enhances the parameter-search capability of the prediction model even without data enhancement. When WGAN-GP-based data enhancement is further incorporated into the ROA optimization framework, the prediction performance is further improved. The ROA-DE-LSTM-half model performs better than ROA-LSTM, with RMSE decreasing from

0.082 to 0.079, and R^2 increasing from 0.642 to 0.667. This result indicates that the joint optimization strategy can make more effective use of the generated degradation information and improve the prediction capability of the model. The complete ROA-DE-LSTM model, which adopts ROA-based global optimization and uses the full generated dataset, further reduces RMSE to 0.053 while increasing R^2 to 0.846. Compared with ROA-LSTM, the complete model reduces RMSE by 35.4% and improves R^2 by 31.8%. This suggests that introducing more effective generated samples under the joint optimization framework can further improve the overall prediction performance.

Figure 13 shows the prediction curves with and without generated samples. The complete model produces a degradation trajectory closer to the actual trend, especially in the middle and late degradation stages. Figure 14 shows the prediction error distributions of different configurations. Compared with the baseline and partially enhanced models, the prediction errors of the complete model are more concentrated around zero, indicating better robustness and smaller prediction bias. These results confirm that the proposed WGAN-GP data enhancement module, when jointly optimized with the LSTM prediction model through ROA, is beneficial for improving prediction accuracy and generalization under limited-data conditions.

5.3. Robustness Validation

(1) IEEE PHM2012 Dataset

To further evaluate the robustness and generalization ability of the proposed framework, additional validation experiments are conducted on the IEEE PHM2012 bearing dataset [44]. The IEEE PHM2012 dataset is a widely used public benchmark dataset for bearing remaining useful life prediction. It contains complete run-to-failure degradation data collected from rolling bearings under accelerated degradation tests. During the experiments, vibration signals were continuously acquired under different operating conditions until bearing failure occurred, providing complete degradation trajectories for model validation.

Compared with the XJTU-SY dataset, the IEEE PHM2012 dataset has different experimental equipment, operating conditions, degradation processes, and lifetime distributions. Therefore, it can be used as an independent dataset to examine whether the proposed ROA-DE-LSTM framework maintains effective prediction performance under different data characteristics. To ensure a fair comparison, the same preprocessing procedure, health indicator construction strategy, WGAN-GP data enhancement method, LSTM network structure, and ROA optimization strategy are directly applied to the IEEE PHM2012 dataset without requiring additional model redesign.

(2) Validation on the IEEE PHM2012 Dataset

To verify the prediction performance of the proposed framework on the IEEE PHM2012 dataset, different models are evaluated under identical experimental settings, including LSTM, DE-LSTM-half, DE-LSTM, ROA-LSTM, ROA-DE-LSTM-half, and ROA-DE-LSTM.

The LSTM model denotes the baseline prediction model without data enhancement and without ROA optimization, using only the 2800 original training samples. The DE-LSTM model denotes the prediction model with full generated samples but without ROA optimization, where 2800 generated samples are added to the original training set and the final training set contains 5600 samples. The DE-LSTM-half model denotes the prediction model trained with half generated samples but without ROA optimization, where 1400 generated samples are added to the original training set and the final training set contains 4200 samples. The ROA-LSTM model denotes the prediction model optimized by ROA but without data enhancement, using only the 2800 original training samples from the IEEE PHM 2012 dataset. The ROA-DE-LSTM-half model denotes the ROA-optimized

prediction model trained with the half-generation setting, in which 1400 generated samples are added and the final training set contains 4200 samples. The ROA-DE-LSTM model denotes the complete proposed model with ROA global optimization and full generated data, where all 2800 generated samples are added and the final training set contains 5600 samples. Figure 15 compares the predicted degradation trajectories of different models. The baseline LSTM model can roughly follow the overall degradation trend, but obvious deviations still exist. After introducing WGAN-GP-based data enhancement or ROA optimization, the predicted trajectories become closer to the actual degradation trajectories, indicating that both generated samples and parameter optimization contribute to prediction improvement.

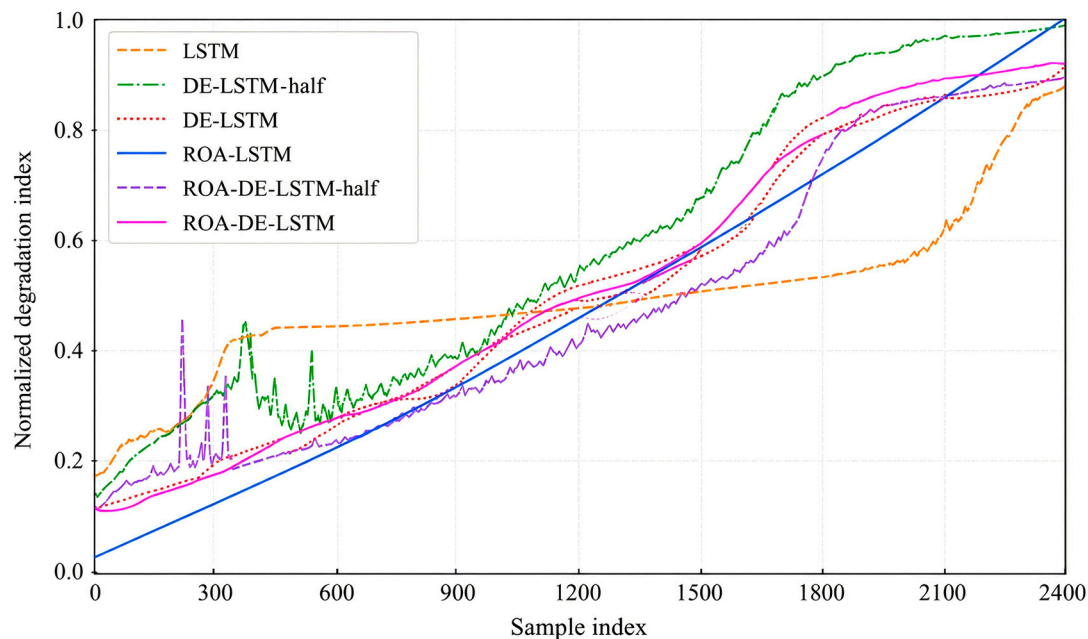


Figure 15. Comparison of degradation trajectories predicted by different model configurations on the IEEE PHM2012 dataset.

The quantitative results are summarized in Table 12. Compared with the baseline LSTM, ROA-LSTM reduces RMSE from 0.183 to 0.068 and increases R^2 from 0.597 to 0.944, showing that ROA optimization significantly improves the prediction ability of the LSTM model. When WGAN-GP-based data enhancement is further incorporated, the complete ROA-DE-LSTM model achieves the lowest RMSE of 0.052 and the highest R^2 of 0.968. Figure 16 shows that the prediction errors of ROA-DE-LSTM are more concentrated around zero, indicating smaller prediction bias and better prediction stability on the tested dataset. These results provide additional evidence that the proposed framework is effective on the IEEE PHM2012 dataset, but its broader applicability to complex real-world industrial scenarios still requires further validation.

Table 12. Performance comparison on the IEEE PHM2012 dataset.

Model Configuration	MAE		RMSE		MAPE (%)		R^2	
	Value	gr	Value	gr	Value	gr	Value	gr
LSTM	0.160	+0.115	0.183	+0.131	43.695	+24.615	0.597	−0.371
DE-LSTM-half	0.112	+0.067	0.128	+0.076	56.008	+36.928	0.802	−0.166
DE-LSTM	0.093	+0.048	0.120	+0.068	27.308	+8.228	0.826	−0.142
ROA-LSTM	0.051	+0.006	0.068	+0.016	18.043	−1.037	0.944	−0.024
ROA-DE-LSTM-half	0.047	+0.002	0.053	+0.001	17.500	−1.580	0.966	−0.002
ROA-DE-LSTM	0.045	+0	0.052	+0	19.080	+0	0.968	+0

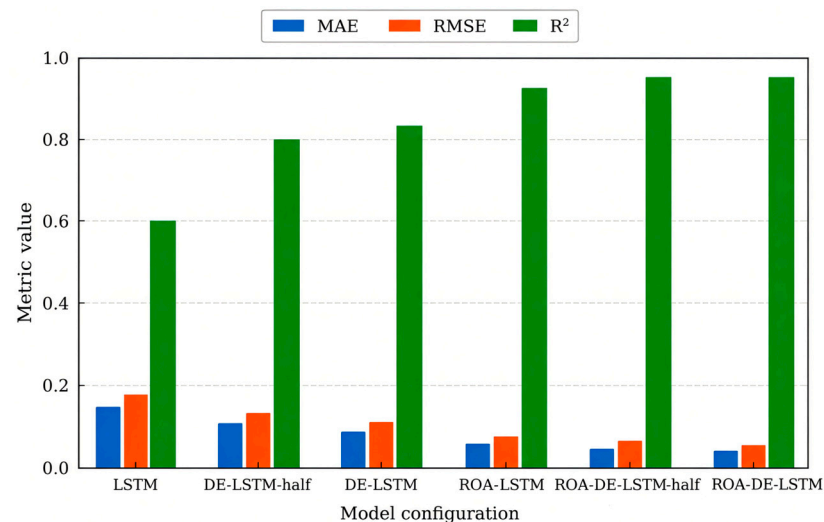


Figure 16. Prediction error distribution on the IEEE PHM2012 dataset.

(3) Comparison with Existing Methods

To further evaluate the competitiveness of the proposed framework, the prediction performance of ROA-DE-LSTM is compared with representative methods reported in the literature on the XJTU-SY and IEEE PHM2012 datasets. The comparison results are summarized in Table 13.

Table 13. Comparison with representative methods reported in the literature.

Dataset	Method	MAE	RMSE	MAPE (%)	R ²
XJTU-SY	ROA-DE-LSTM (Ours)	0.035	0.053	116.04	0.846
	Stacked BiGRU [45]	0.1086	0.0747	—	—
	CNN-GRU-MHA + Transfer Learning [46]	—	0.0693 (avg)	—	—
IEEE PHM2012	ROA-DE-LSTM (Ours)	0.045	0.052	19.080	0.968
	1D-DCAE + Bi-LSTM + TPA [47]	0.035	0.055	—	—
	SSA-BiLSTM + Attention [45]	0.0390	0.0531	—	0.9661
	VQ-VAE [48]	0.063	0.086	—	—
	GMFE-GAN [49]	0.0906	0.1053	47.94	—
	CVGAN+ GRU [50]	0.105	0.221	—	—

For the XJTU-SY dataset, the proposed ROA-DE-LSTM is compared with three representative methods that do not employ data enhancement. As shown in Table 13, ROA-DE-LSTM achieves an MAE of 0.035 and an RMSE of 0.053. The proposed method achieves lower prediction errors. Although CNN-BiLSTM reports a slightly lower RMSE, ROA-DE-LSTM achieves a lower MAE, indicating that the proposed framework maintains competitive prediction accuracy on the XJTU-SY dataset.

For the IEEE PHM2012 dataset, the proposed method is compared with several representative methods. Among them, VQ-VAE, GMFE-GAN, and CVGAN-GRU also introduce data enhancement strategies. As shown in Table 13, ROA-DE-LSTM achieves an RMSE of 0.052 and an R² of 0.968. The proposed data enhancement and optimization strategy can generate more effective degradation representations for RUL prediction.

5.4. Time Consumption Comparison of Optimization Algorithms

Computational time is a key metric for evaluating the efficiency of optimization algorithms. In this study, we compare the time consumption of three algorithms—ROA, HHO,

and PSO—across different test functions. Shorter execution time indicates faster convergence toward the optimal solution, reflecting the algorithm’s efficiency in optimization and search.

Table 14 summarizes the time consumption for each algorithm under various test conditions. The results show that ROA outperforms HHO and PSO in terms of computational efficiency across all test functions. For example, in the Sphere function test, ROA completes the task in just 0.025 s, while HHO and PSO take 0.866 s and 2.529 s, respectively. This demonstrates that ROA achieves faster convergence, reducing computational time.

Table 14. Time consumption of optimization algorithms.

Test Function	ROA	HHO	PSO
Sphere Function	0.025 s	0.866 s	2.529 s
Baseline Function 2	0.032 s	1.047 s	2.724 s
Minimum Absolute Value Function	0.020 s	0.977 s	2.237 s
Rastrigin Function	0.029 s	1.149 s	2.812 s
Ackley Function	0.033 s	1.421 s	3.641 s

Similarly, in the Rastrigin and Ackley function tests, ROA’s execution times are 0.029 s and 0.033 s, respectively, while HHO and PSO take between 1.149 s and 3.641 s. ROA’s execution time is 97.68% faster than HHO and 99.09% faster than PSO. This highlights ROA’s superior efficiency.

The time advantage of ROA is attributed to its unique optimization mechanism, which allows it to explore the solution space globally and converge quickly after identifying promising regions. The results validate the effectiveness of the ROA-DE-LSTM framework in time-sensitive optimization tasks, especially in complex scenarios like Remaining Useful Life (RUL) prediction for rolling bearings.

6. Conclusions

In this study, an ROA-DE-LSTM framework is proposed for remaining useful life (RUL) prediction of rolling bearings under limited-data conditions. The framework combines a Remora Optimization Algorithm (ROA)-based optimization strategy, a Wasserstein GAN with Gradient Penalty (WGAN-GP) for data augmentation, and an LSTM network for temporal modeling, aiming to improve prediction performance in data-scarce industrial scenarios.

The core idea of the proposed method lies in jointly optimizing both data generation and prediction stages under a unified ROA-driven framework. In particular, the adaptive ROA mechanism improves the balance between exploration and exploitation by dynamically switching between Whale Optimization Algorithm (WOA) and Sailfish Optimization (SFO), which enhances the robustness of hyperparameter search in complex optimization spaces.

The main contributions can be summarized as follows:

- (1) ROA-optimized WGAN-GP is introduced to improve the quality and diversity of generated degradation data, alleviating the impact of limited run-to-failure samples.
- (2) ROA-optimized LSTM is designed to better capture long-term dependencies in health indicator sequences, improving prediction stability under limited and noisy data.
- (3) An integrated ROA-DE-LSTM framework is developed to jointly optimize data augmentation and prediction modules, improving consistency between generated data and downstream forecasting tasks.

Experimental results on the XJTU-SY and IEEE PHM2012 datasets show that the proposed method achieves consistently better performance compared with PSO- and HHO-

based optimization strategies in terms of prediction accuracy and convergence behavior. In addition, statistical evaluations and cross-dataset experiments indicate that the proposed framework maintains stable performance under different operating conditions.

It should be noted that the proposed framework is evaluated under a limited number of degradation trajectories and controlled experimental conditions. While the results demonstrate its effectiveness, further validation on more diverse industrial datasets, including gearboxes and other types of equipment, and real-world operating environments is still necessary. Future work will focus on improving generalization capability and reducing computational cost for practical deployment.

Author Contributions: Y.B.: Methodology, Validation, Funding acquisition, Conceptualization, K.C.: Visualization, Writing—original draft. M.L.: Formal analysis, Data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the 2025 Guangdong Province “Bright Action” Project: (grant number: CC/XM-202401JS0202), and the Undergraduate Innovation Research Projects: (grant numbers: 202511059029, S202511059126, S202511059142X).

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

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang, K.; Xu, Y.; Liao, Z.; Song, L.; Chen, P. A novel Fast Entrogram and its applications in rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2021**, *154*, 107582. [\[CrossRef\]](#)
2. Kumar, A.; Vashishtha, G.; Gandhi, C.P.; Tang, H.; Xiang, J. Tacho-less sparse CNN to detect defects in rotor-bearing systems at varying speed. *Eng. Appl. Artif. Intell.* **2021**, *104*, 104401. [\[CrossRef\]](#)
3. Vrignat, P.; Kratz, F.; Avila, M. Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108140. [\[CrossRef\]](#)
4. Ding, W.; Li, J.; Mao, W.; Meng, Z.; Shen, Z. Rolling bearing remaining useful life prediction based on dilated causal convolutional DenseNet and an exponential model. *Reliab. Eng. Syst. Saf.* **2023**, *232*, 109072. [\[CrossRef\]](#)
5. Xu, Z.; Bashir, M.; Liu, Q.; Miao, Z.; Wang, X.; Wang, J.; Ekere, N. A novel health indicator for intelligent prediction of rolling bearing remaining useful life based on unsupervised learning model. *Comput. Ind.* **2023**, *176*, 108999. [\[CrossRef\]](#)
6. Maio, F.D.; Tsui, K.L.; Zio, E. Combining relevance vector machines and exponential regression for bearing residual life estimation. *Mech. Syst. Signal Process.* **2012**, *31*, 405–427. [\[CrossRef\]](#)
7. Catelani, M.; Ciani, L.; Fantacci, R.; Patrizi, G.; Picano, B. Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3524611. [\[CrossRef\]](#)
8. Zhou, J.; Qin, Y.; Chen, D.; Liu, F.; Qian, Q. Remaining useful life prediction of bearings by a new reinforced memory GRU network. *Adv. Eng. Inform.* **2022**, *53*, 101682. [\[CrossRef\]](#)
9. Lu, W.; Wang, Y.; Zhang, M.; Gu, J. Physics guided neural network: Remaining useful life prediction of rolling bearings using long short-term memory network through dynamic weighting of degradation process. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107350. [\[CrossRef\]](#)
10. Feng, Z.; Zhang, Z. A small sample rolling bearing fault diagnosis based on PSD-VME and DS evidence theory enhanced mRVM. *Comput. Electr. Eng.* **2024**, *118*, 109458. [\[CrossRef\]](#)
11. Yao, X.; Lu, X.; Jiang, Q.; Shen, Y.; Xu, F.; Zhu, Q. SSPENet: Semi-supervised prototype enhancement network for rolling bearing fault diagnosis under limited labeled samples. *Adv. Eng. Inform.* **2024**, *61*, 102560. [\[CrossRef\]](#)
12. Deng, J.; Liu, G.; Wang, L.; Liu, G.; Wu, X. Intelligent optimization design of squeeze casting process parameters based on neural network and improved sparrow search algorithm. *J. Ind. Inf. Integr.* **2024**, *39*, 100600. [\[CrossRef\]](#)
13. Grasso, M.; Chatterton, S.; Pennacchi, P.; Colosimo, B.M. A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis. *Mech. Syst. Signal Process.* **2016**, *81*, 126–147. [\[CrossRef\]](#)
14. Guo, Q.; Li, Y.; Song, Y.; Wang, D.; Chen, W. Intelligent fault diagnosis method based on full 1-D convolutional generative adversarial network. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2044–2053. [\[CrossRef\]](#)
15. Yu, T.; Li, C.; Huang, J.; Xiao, X.; Zhang, X.; Li, Y.; Fu, B. ReF-DDPM: A novel DDPM-based data augmentation method for imbalanced rolling bearing fault diagnosis. *Reliab. Eng. Syst. Saf.* **2024**, *251*, 110343. [\[CrossRef\]](#)

16. Saidi, L.; Ali, J.B.; Bechhoefer, E.; Benbouzid, M. Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR. *Appl. Acoust.* **2017**, *120*, 1–8. [\[CrossRef\]](#)
17. Jia, H.; Peng, X.; Lang, C. Remora optimization algorithm. *Expert Syst. Appl.* **2021**, *185*, 115665. [\[CrossRef\]](#)
18. Mathew, J.; Pang, C.K.; Luo, M.; Leong, W.H. Classification of imbalanced data by oversampling in kernel space of support vector machines. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 4065–4076. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Mao, W.; He, L.; Yan, Y.; Wang, J. Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine. *Mech. Syst. Signal Process.* **2017**, *83*, 450–473. [\[CrossRef\]](#)
20. Zhao, J.; Hu, L.; Huang, L.; Wang, C.; Liang, D. MSRA-G: Combination of multi-scale residual attention network and generative adversarial networks for hyperspectral image classification. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106017. [\[CrossRef\]](#)
21. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*; Curran Associates Inc.: Long Beach, CA, USA, 2017; pp. 5769–5779.
22. Li, Z.; Zheng, T.; Wang, Y.; Cao, Z.; Guo, Z.; Fu, H. A Novel Method for Imbalanced Fault Diagnosis of Rotating Machinery Based on Generative Adversarial Networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3500417. [\[CrossRef\]](#)
23. Chen, J.; Yan, Z.; Lin, C.; Yao, B.; Ge, H. Aero-engine high speed bearing fault diagnosis for data imbalance: A sample enhanced diagnostic method based on pre-training WGAN-GP. *Measurement* **2023**, *213*, 112709. [\[CrossRef\]](#)
24. Gu, X.; Yu, Y.; Guo, L.; Gao, H.; Luo, M. CSWGAN-GP: A new method for bearing fault diagnosis under imbalanced condition. *Measurement* **2023**, *217*, 113014. [\[CrossRef\]](#)
25. Tian, H.; Yang, L.; Ju, B. Spatial correlation and temporal attention-based LSTM for remaining useful life prediction of turbofan engine. *Measurement* **2023**, *214*, 112816. [\[CrossRef\]](#)
26. Sun, J.; Zhang, X.; Wang, J. Lightweight bidirectional long short-term memory based on automated model pruning with application to bearing remaining useful life prediction. *Eng. Appl. Artif. Intell.* **2023**, *118*, 105662. [\[CrossRef\]](#)
27. Chen, Z.; Wu, M.; Zhao, R.; Guretno, F.; Yan, R.; Li, X. Machine Remaining Useful Life Prediction via an Attention-Based Deep Learning Approach. *IEEE Trans. Ind. Electron.* **2021**, *68*, 2521–2531. [\[CrossRef\]](#)
28. Chang, Z.; Yuan, W.; Huang, K. Remaining useful life prediction for rolling bearings using multi-layer grid search and LSTM. *Comput. Electr. Eng.* **2022**, *101*, 108083. [\[CrossRef\]](#)
29. Pater, I.D.; Mitici, M. Developing Health Indicators and RUL Prognostics for Systems with Few Failure Instances and Varying Operating Conditions Using a LSTM Autoencoder. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105582. [\[CrossRef\]](#)
30. Zhu, T.; Chen, Z.; Zhou, D.; Xia, T.; Pan, E. Adaptive staged remaining useful life prediction of roller in a hot strip mill based on multi-scale LSTM with multi-head attention. *Reliab. Eng. Syst. Saf.* **2024**, *248*, 110161. [\[CrossRef\]](#)
31. Lin, R.; Yu, Y.; Wang, H.; Che, C.; Ni, X. Remaining useful life prediction in prognostics using multi-scale sequence and Long Short-Term Memory network*. *J. Comput. Sci.* **2022**, *57*, 101508. [\[CrossRef\]](#)
32. Zang, C.; Liu, R.; Guo, W.; Yan, H. Prediction of Remaining Useful Life of Bearing Based on Long Short-Term Memory Network Optimized by Slime Mould Algorithm. In *Proceedings of the 2021 China Automation Congress (CAC)*, Beijing, China, 22–24 October 2021; pp. 6616–6621. [\[CrossRef\]](#)
33. Gao, Y.; Wang, X.; Yan, L. Bearing Remaining Useful Life Prediction by combining CNN with PSO_LSSVM. In *Proceedings of the 2021 33rd Chinese Control and Decision Conference (CCDC)*, Kunming, China, 22–24 May 2021; pp. 7124–7129.
34. Chen, Y.; Wang, X.; Zhang, Z.; Zhou, D.; Jiang, Y.; Weng, J.; Walther, F.; Gong, J. An approach to estimate the low cycle fatigue probabilistic curves of PBF-LB/M 316L steel from small size datasets using the remora optimization algorithm. *Int. J. Fatigue* **2024**, *185*, 108375. [\[CrossRef\]](#)
35. Ragab, M.; Abukhodair, F.; AL-Ghamdi, A.A.-M.; Shabanah, S.; Alfasi, B.; Alyamani, M. Synergizing Remora Optimization Algorithm and Transfer Learning for Visual Places Recognition in Intelligent Transportation Systems and Consumer Electronics. *IEEE Trans. Consum. Electron.* **2024**, *70*, 3731–3739. [\[CrossRef\]](#)
36. Alanazi, A.; Alanazi, M.; Arabi, S.; Sarker, S. A New Maximum Power Point Tracking Framework for Photovoltaic Energy Systems Based on Remora Optimization Algorithm in Partial Shading Conditions. *Appl. Sci.* **2022**, *12*, 3828. [\[CrossRef\]](#)
37. Niola, V.; Cosenza, C.; Fornaro, E.; Malfi, P.; Melluso, F.; Nicolella, A.; Savino, S.; Spirito, M. Torque/Speed Equilibrium Point Monitoring of an Aircraft Hybrid Electric Propulsion System Through Accelerometric Signal Processing. *Appl. Sci.* **2025**, *15*, 2135. [\[CrossRef\]](#)
38. Jiang, G.; Li, D.; Feng, K.; Li, Y.; Zheng, J.; Ni, Q.; Li, H. Rolling Bearing Fault Diagnosis Based On Convolutional Capsule Network. *J. Dyn. Monit. Diagn.* **2023**, *2*, 275–289.
39. Liu, K.; Chehade, A.; Song, C. Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis. *IEEE Trans. Autom. Sci. Eng.* **2015**, *14*, 1504–1514. [\[CrossRef\]](#)
40. Wang, B.; Lei, Y.; Li, N.; Li, N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans. Reliab.* **2018**, *69*, 401–412. [\[CrossRef\]](#)

41. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the 1995 International Conference on Neural Networks (ICNN'95), Perth, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948.
42. Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [[CrossRef](#)]
43. Shen, Y.; Tang, B.; Li, B.; Tan, Q.; Wu, Y. Remaining useful life prediction of rolling bearing based on multi-head attention embedded Bi-LSTM network. *Measurement* **2022**, *202*, 111803. [[CrossRef](#)]
44. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebel-Morello, B.; Zerhouni, N.; Varnier, C. PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests. In Proceedings of the IEEE International Conference on Prognostics and Health Management, Denver, CO, USA, 18–21 June 2012; pp. 1–8.
45. Zou, Y.; Sun, W.; Xu, T.; Wang, B. Research on the Bearing Remaining Useful Life Prediction Method Based on Optimized BiLSTM. *Sensors* **2025**, *25*, 4351. [[CrossRef](#)] [[PubMed](#)]
46. Yu, J.; Shao, J.; Peng, X.; Liu, T.; Yao, Q. Remaining Useful Life of the Rolling Bearings Prediction Method Based on Transfer Learning Integrated with CNN-GRU-MHA. *Appl. Sci.* **2024**, *14*, 9039. [[CrossRef](#)]
47. Wang, C.; Jiang, W.; Shi, L.; Zhang, L. Rolling Bearing Remaining Useful Life Prediction Using Deep Learning Based on High-Quality Representation. *Sci. Rep.* **2025**, *15*, 8228. [[CrossRef](#)] [[PubMed](#)]
48. Wang, J.; Zhang, Q.; Zhu, G.; Sun, G. Utilizing VQ-VAE for End-to-End Health Indicator Generation in Predicting Rolling Bearing RUL. *Meas. Sci. Technol.* **2024**, *35*, 066106. [[CrossRef](#)]
49. Suh, S.; Lukowicz, P.; Lee, Y.O. Generalized Multiscale Feature Extraction for Remaining Useful Life Prediction of Bearings with Generative Adversarial Networks. *Knowl.-Based Syst.* **2022**, *237*, 107866. [[CrossRef](#)]
50. Wang, J.; Zhang, Q.; Zhu, G.; Sun, G. Utilizing Autoregressive Networks for Full Lifecycle Data Generation of Rolling Bearings for RUL Prediction. *arXiv* **2024**, arXiv:2401.01119.

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.