applied sciences

MDPI

Article

Research on Remaining Useful Life Prediction of Bearing with Digital Twin and TSMixer

Quanbo Lu [1,2,*], Dong Zhu [3] and Mei Li [4]

[1] School of Integrated Circuit Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2] College of Communication Engineering, Chongqing Polytechnic University of Electronic Technology, Chongqing 401331, China
[3] Sevnce Robotics Co., Ltd., Chongqing 401120, China
[4] School of Information Engineering, China University of Geosciences, Beijing 100083, China
* Correspondence: luquanbo111@163.com

Abstract

To advance the prediction accuracy of remaining useful life (RUL) for rolling element bearings, this study proposes a novel hybrid prognostic framework that synergistically integrates digital twin (DT) technology with the Time-series Mixer (TSMixer) architecture. A high-fidelity virtual bearing digital prototype is constructed via multi-physics fusion modeling of DT, enabling real-time state perception and dynamic evolution simulation of the physical bearing system, which further generates abundant, high-reliability augmented operational data to compensate for the scarcity of measured full-life degradation samples. By fusing multi-domain features extracted from both historical monitoring datasets and DT-augmented simulation sequences, a comprehensive multi-source feature space covering the full degradation trajectory of bearings is established. Subsequently, the TSMixer model, which excels in parallel capturing of local temporal patterns and global long-range dependencies, is employed to learn the implicit time-varying degradation characteristics from the constructed time-series datasets, so as to achieve accurate RUL estimation of the tested bearings. Experimental validation on the publicly available benchmark dataset demonstrates that the proposed framework yields superior prognostic performance, with a prediction accuracy improvement of approximately 9.23% compared with several state-of-the-art baseline models. Furthermore, the proposed hybrid paradigm exhibits remarkable robustness and generalization capability under variable working conditions, which provides a promising technical solution for intelligent predictive maintenance and active fault prevention of high-end mechanical equipment.

Keywords: digital twin; remaining useful life prediction; bearing; TSMixer

Check for updates

Academic Editor: Rui Araújo

Received: 18 May 2026
Revised: 27 June 2026
Accepted: 28 June 2026
Published: 6 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) license.

1. Introduction

With the continuous advancement of advanced manufacturing and industrial intelligentization, modern mechanical equipment represented by high-end CNC machine tools has been evolving towards ultra-precision, high-degree automation and high-speed operation. Nevertheless, long-term cyclic operation under complex working conditions inevitably induces irreversible performance degradation of core components such as rolling bearings. Once the degradation process crosses the critical failure threshold without timely maintenance intervention, it may trigger severe operational accidents, resulting in enormous economic losses, catastrophic mechanical failures and even personal injuries. On

the other hand, blindly implemented over-maintenance and unnecessary component replacement strategies will also cause substantial waste of production resources and increase the overall lifecycle operation cost of equipment [1]. In this context, accurate and reliable remaining useful life (RUL) prediction for critical mechanical components serves as a core enabler for condition-based maintenance, which allows maintenance activities to be scheduled at the optimal time window, so as to maximally guarantee operational safety while improving the comprehensive economic benefits of the entire production system.

At present, RUL prediction methods for mechanical equipment are broadly categorized into three well-recognized classes: failure model-based methods, probabilistic statistical methods, and data-driven methods [2]. While model-based techniques typically yield relatively high prediction accuracy under ideal conditions, their practical deployment is fundamentally constrained by the requirement for highly precise physical governing equations and stringent predefined operational assumptions. For complex bearing systems characterized by multi-source nonlinearity and dynamic load fluctuations, constructing an accurate full-scale physical model that fully captures the underlying degradation process remains a long-standing challenge, which substantially narrows the general applicability of this category of methods. As an alternative branch, probabilistic statistical approaches primarily leverage long-term historical operational datasets to quantify the inherent probabilistic correlations between time-varying parameter degradation trajectories and implicit failure evolution mechanisms. Through quantitatively matching the real-time multi-parameter state space of in-service equipment with the pre-calibrated probabilistic degradation model, the remaining useful life can be statistically inferred with quantified confidence intervals [3]. In sharp contrast, the prediction performance of data-driven methods is highly contingent on the availability of sufficiently large, fully labeled, and high-quality degradation datasets, which often restricts their application scenarios in industrial scenarios with scarce fault data. For instance, the teager–kaiser energy operator (TKEO) has been extensively employed in vibration-based fault diagnosis owing to its ability to estimate the instantaneous energy of nonlinear and nonstationary signals [4]. By emphasizing transient components and modulated characteristics, TKEO can effectively enhance weak fault-related features that are often masked by noise [5].

With the rapid development of Digital Twin (DT) technology [6], new opportunities have emerged for monitoring and predicting the operational state of bearings. Digital Twin technology integrates physical models, sensor data, and operational history to perform multidisciplinary, multi-physics, multi-scale, and probabilistic simulations, thereby enabling a high-fidelity virtual representation of physical entities in cyberspace. Initially, DT technology was applied to monitor and predict the flight status of aircraft in order to ensure safe operation throughout their life cycle. With the continuous advancement of intelligent manufacturing, DT technology has been widely adopted in product lifecycle management [7], visualization and monitoring of equipment operating conditions [8], spindle dynamic characteristic monitoring [9], tool condition monitoring [10], product design and fault diagnosis [11], as well as digital monitoring of thermal characteristics [12].

Building upon these advantages, DT technology provides a novel virtual–physical fusion framework for RUL prediction through high-fidelity modeling and real-time state synchronization. To further enhance prediction accuracy and efficiency, this paper proposes a hybrid RUL prediction method that integrates DT technology with the TSMixer (DT-TSMixer) model.

The main contributions of this paper are outlined as follows.

(1) Hybrid Integration of DT and TSMixer for RUL Prediction: Unlike prior DT-based approaches that often rely on conventional neural networks or LSTM architectures, the proposed framework explicitly integrates the high-fidelity DT model with the TSMixer

architecture. This combination allows simultaneous utilization of physical system knowledge from the DT and data-driven temporal feature extraction via TSMixer, improving predictive accuracy.

(2) Dynamic Correction of DT Model Outputs Using Data-Driven Observations: The hybrid RUL prediction method introduces a closed-loop correction mechanism where the discrepancy between TSMixer predictions and DT theoretical outputs is used to iteratively refine the DT model. This ensures that the virtual model adapts to operational variations in real-time, which is not addressed in existing approaches.

(3) Enhanced Multi-Source Feature Fusion: The proposed framework extracts both time-domain and frequency-domain features from historical and DT-generated data, constructing a comprehensive multi-source dataset. This allows the TSMixer model to capture complex degradation patterns more effectively than standard DT-based RUL prediction methods that often use single-source data.

(4) Improved Prediction Robustness and Reliability: Experimental results demonstrate that the DT-TSMixer framework achieves approximately 9.23% improvement in prediction accuracy over baseline models, and it maintains stable performance across different operating conditions. This validates the framework's practical advantage in intelligent maintenance and fault prevention for industrial bearings.

2. Related Works

2.1. TSMixer-Based RUL Prediction

Lee et al. proposed a TSMixer- and transfer learning-based framework that improves prediction reliability for small-scale solar power systems using limited short-term time-series data [13]. Yang et al. developed an enhanced TSMixer model to improve particulate matter prediction accuracy and support adaptive environmental control under dynamic conditions [14]. Pei et al. introduced a multi-scale TSMixer architecture capable of capturing temporal dependencies at different resolutions for more accurate stock index futures forecasting [15]. Yamazono et al. proposed a permutation-dependent feature mixing strategy within the TSMixer framework to enhance multivariate time-series forecasting performance [16]. Kukadiya et al. improved a modified TSMixer model and validated its robustness on field synchrophasor data for prediction under cyber-attack scenarios [17]. Li et al. proposed the TBMTSMixer model to accurately forecast tunnel boring machine penetration rates in hard rock environments by exploiting temporal feature interactions [18]. Cao et al. developed a residual-corrected TSMixer combined with a Rime-enhanced decomposition strategy to improve photovoltaic power prediction accuracy and stability [19]. Han et al. proposed a hard disk failure prediction framework integrating multilayer convolution, attention mechanisms, and TSMixer to enhance fault prediction capability and reliability [20].

Although the TSMixer explicitly models inter-variable correlations through an attention mechanism and is particularly well-suited for high-dimensional multivariate time-series prediction, the effective integration between models and data remains insufficient. This lack of fusion leads to significant challenges in quantifying bearing defect sizes, thereby constraining improvements in identification accuracy and reliability. DT technology offers a new paradigm for bearing condition monitoring by integrating real-time data with physics-based models, thereby enhancing the accuracy and reliability of bearing health management.

2.2. DT-Based RUL Prediction

Dai et al. proposed a multi-stage tool wear prediction approach based on a multi-dimensional DT framework, enabling real-time tool condition monitoring and accurate

RUL prediction through the integration of virtual–physical data and wear evolution modeling [21]. Lu et al. developed a DT-driven and data-driven framework that integrates virtual–physical interaction with data analytics to achieve accurate RUL prediction of gearboxes [22]. Cao et al. developed a tool-health DT framework that integrates innovation-adaptive sensor fusion and uncertainty-aware prognostics to achieve reliable tool condition monitoring and accurate RUL prediction for CNC predictive maintenance [23]. Feng et al. proposed a DT-driven causal diagnosis framework for high-speed spindle systems, enabling life-cycle health assessment and RUL prediction through the integration of virtual–physical interaction and causal degradation analysis [24]. Kobayashi & Alam presented an explainable and interpretable AI approach for DTs, specifically applied to RUL prediction, ensuring trustworthiness in the decision-making process [25]. Zhang et al. developed a self-adaptive DT model for predicting the RUL of fuel cells, aiming for more accurate lifetime forecasting in energy systems [26]. Ramesh et al. developed a DT-enabled prediction framework integrating ANFIS and a genetic algorithm to achieve real-time machining quality prediction and optimization, providing support for predictive maintenance and RUL assessment in manufacturing systems [27]. Wei et al. proposed an applicability evaluation method for DT technology in bearings, providing a systematic foundation for implementing DT-based condition monitoring, predictive maintenance, and RUL prediction [28].

Although recent studies have integrated DT technology with RUL prediction, most existing models rely on data generated from simulated operating conditions, which cannot fully replicate real-world scenarios. This study provides a novel solution for bearing condition monitoring and RUL prediction. Compared with traditional approaches, the proposed method not only enables real-time bearing condition monitoring but also supplies reliable data support for RUL prediction models. To address the limitation of the TSMixer model, which exhibits high variability under insufficient training data, multi-dimensional real-time state data generated from physical entities are incorporated to effectively augment the training dataset. These real-time data enhance the model's ability to extract degradation features, thereby providing a more reliable basis for equipment health assessment and predictive maintenance.

3. The Proposed Method

3.1. DT Model

The internal structure and dynamic characteristics of bearings are highly complex, exhibiting diverse dynamic responses due to variations in operating conditions and design configurations. This complexity makes it extremely challenging to establish a comprehensive and highly accurate dynamic model capable of capturing all bearing behaviors. Currently, most bearing dynamic models are developed based on simplified assumptions, including two-degree-of-freedom, four-degree-of-freedom, and five-degree-of-freedom models. Among these, the two-degree-of-freedom model has become the most widely adopted due to its balance between simplicity and effectiveness, as illustrated in Figure 1. This model primarily focuses on the translational motion of the inner ring in a two-dimensional plane, while the dynamic behavior of rolling elements is equivalently represented by a nonlinear elastic restoring force. Consequently, a simplified mass–spring–damper dynamic system is constructed. In this study, a fault dynamic model of bearings is developed based on this simplified framework, aiming to investigate the bearing dynamic behavior and provide theoretical support for fault analysis and performance improvement.

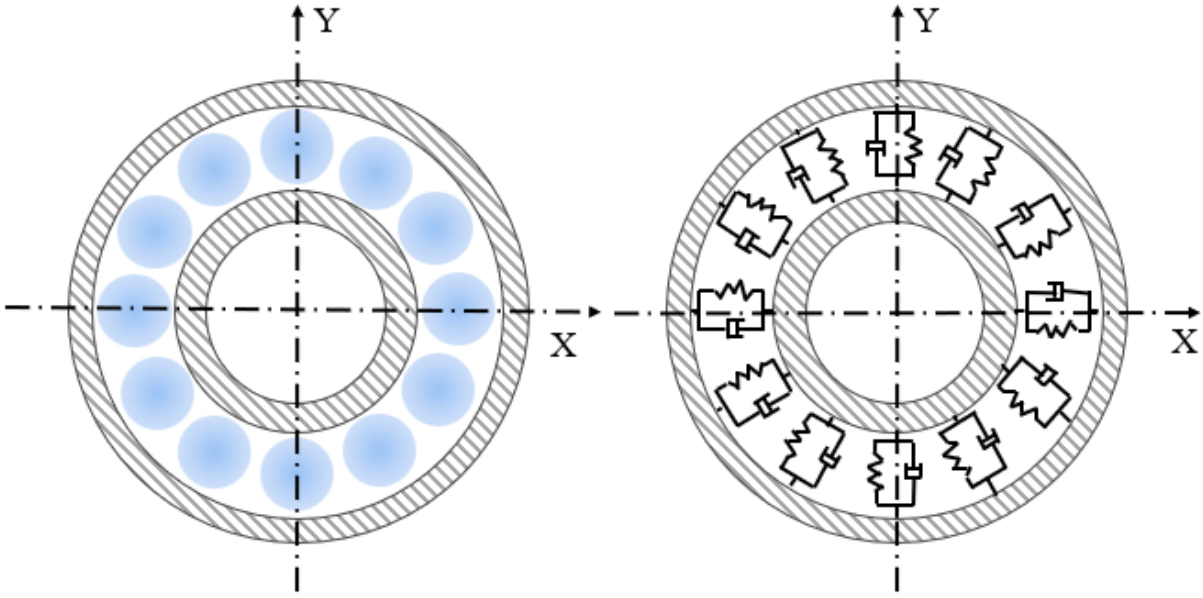


Figure 1. Bearing structure and simplified model drawing.

Based on the simplified two-degree-of-freedom dynamic model of bearings, the dynamic equations are derived by analyzing the force equilibrium of the bearing under different operating conditions. The motion characteristics of the bearing in both axial and radial directions are accurately modeled [26], as expressed in Equation (1).

$$\begin{cases} M\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx + F_x = W_x \\ M\frac{d^2y}{dt^2} + c\frac{dy}{dt} + ky + F_y = W_y \end{cases} \quad (1)$$

where m denotes the equivalent mass of the bearing system, c represents the equivalent damping coefficient, and k is the stiffness coefficient. The variables x and y correspond to the vibration displacement responses of the bearing in the orthogonal $X-Y$ coordinate system, while F_x and F_y represent the internal force components in the respective directions. Additionally, W_x and W_y denote the externally applied radial load components along the X and Y axes.

When a rolling element passes through a localized defect on the outer race, the contact elastic deformation within the contact zone is altered, leading to a redistribution of the Hertzian contact stress. The core of fault modeling lies in capturing the dynamic evolution of the contact clearance to identify fault characteristics. The constructed mathematical model enables parametric visualization of fault features during simulation. This model effectively reflects the mechanical characteristics of the fault region and provides a theoretical foundation for fault diagnosis. Figure 2 illustrates a schematic diagram of a localized outer race defect, where β_0 denotes the central angular position of the defect region.

In this study, the assumed model of an outer raceway defect is visually illustrated in Figure 3, where L denotes the defect length along the rolling direction, W represents the defect width, and H indicates the defect depth. To simplify the modeling process, the defect depth H is assumed to remain constant, and the defect is considered square-shaped with $L = W$. This assumption enables a focused analysis of the influence of defect size on bearing performance while avoiding the additional complexity introduced by depth variation, thereby ensuring model tractability and analytical clarity.

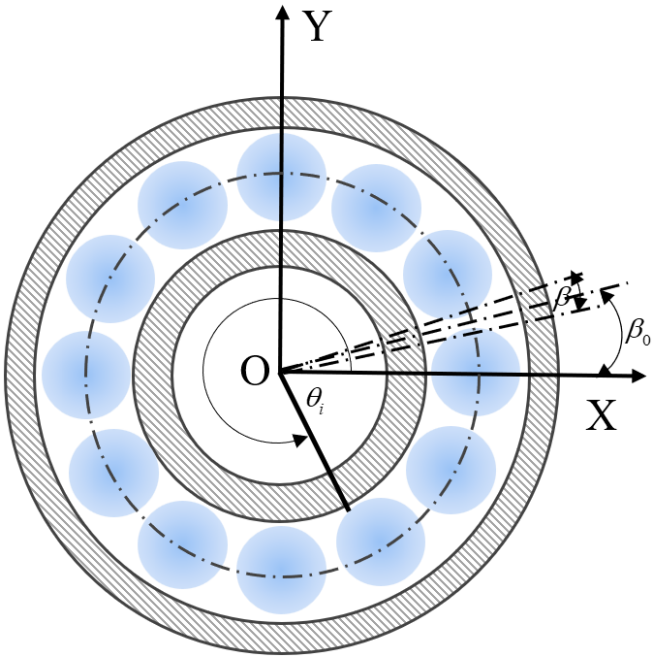


Figure 2. Schematic diagram of localized partial failure of bearing.

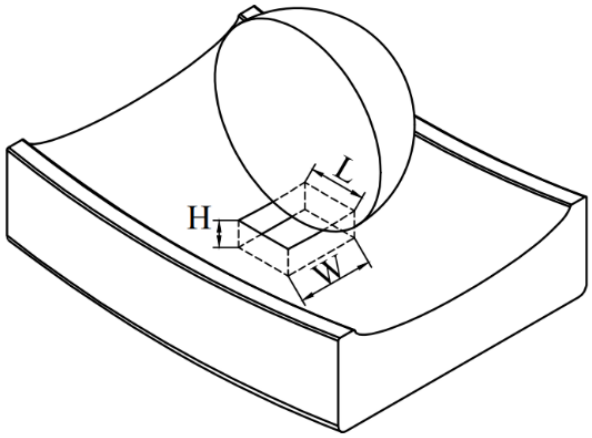


Figure 3. Bearing failure shape schematic.

When a rolling element enters or exits the defect region, the contact deformation δ_i changes accordingly. For bearings with localized defects, the contact deformation δ_i of the i-th rolling element is jointly affected by the radial clearance C_γ and the defect depth [1].

$$\delta_i = x\cos\theta_i + y\sin\theta_i - C_\gamma - H' \tag{2}$$

where x and y denote vibration displacements, θ_i represents the angular position of the i-th rolling element at time t, and H' is the displacement excitation function.

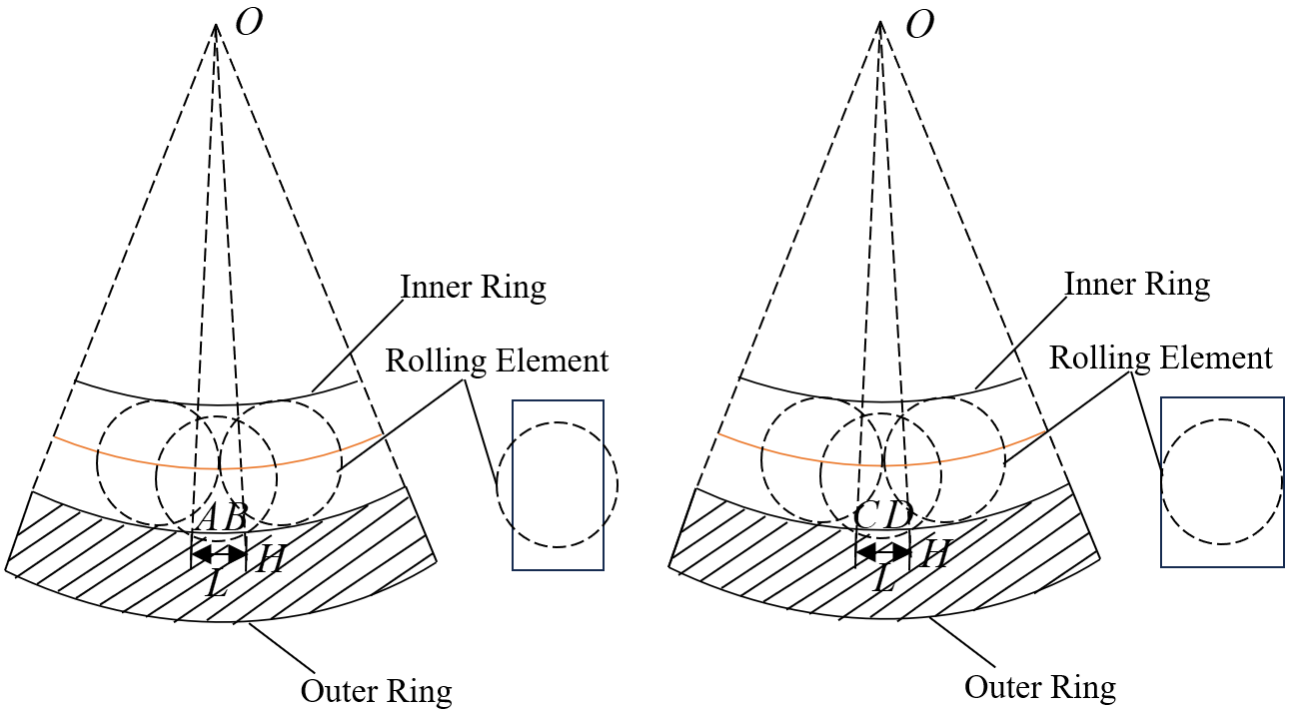
3.1.1. Time-Varying Displacement Induced by Localized Defects

As the rolling element passes through the defect region, an additional time-varying displacement excitation is induced, typically exhibiting a half-sine waveform that increases and then decreases. To accurately describe this behavior, a piecewise function is adopted.

The maximum displacement amplitude H generated when a rolling element passes through the defect region is given by Equation (3) [1].

$$H'_{\max} = \frac{d_b}{2} - \sqrt{\left(\frac{d_b}{2}\right)^2 - \left(\frac{L}{2}\right)^2} \tag{3}$$

The rolling process can be divided into three stages: entry, traversal, and exit, with two possible contact conditions—non-contact with the defect bottom and contact with the defect bottom. Here, H represents the defect depth, while L denotes the circumferential length of the defect region. The displacement excitation induced by bearing defects is typically modeled using a piecewise function, through which the evolution of localized defects in bearings can be categorized into two cases, as illustrated in Figure 4.



(**a**) Early stages of defects (**b**) Defective late stage

Figure 4. Schematic representation of the shape of defects of different sizes in the raceway.

In the first case, when the rolling element diameter exceeds the circumferential size of the defect, the maximum theoretical displacement $H'_{\max}$ is smaller than the defect depth H, indicating that the rolling element only interacts with the defect edges, which is typical of early-stage damage, as is visually illustrated in Figure 4a. Considering the non-ideal rigidity of the bearing, elastic deformation occurs when the rolling element comes into contact with the defect region under a given load. The resulting contact deformation is essentially the coupled effect of the elastic deformations of both the rolling element and the inner and outer race defect regions. As illustrated in Figure 5, the maximum displacement of the rolling element at this stage is defined as the difference between the rolling element deformation C_d and the outer race deformation C_o, which can be calculated as follows [2]:

$$H'_{\max} = C_d - C_o \tag{4}$$

$$\begin{cases} C_o = r_o - \sqrt{r_o^2 - \left(\frac{L}{2}\right)^2} \\ C_d = r_d - \sqrt{r_d^2 - \left(\frac{L}{2}\right)^2} \end{cases} \tag{5}$$

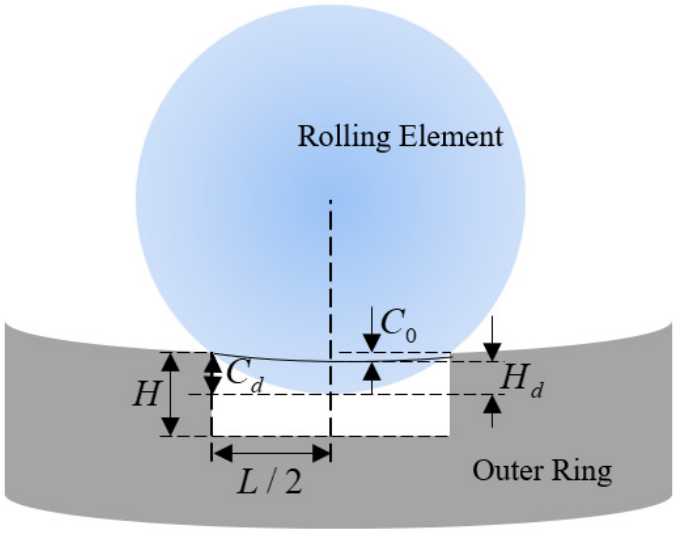


Figure 5. Schematic diagram of the time-varying displacement of the center of the sphere of a rolling body passing through a local fault region of the outer ring.

In the second case, due to the specific geometric characteristics of the defect region, the rolling element comes into contact with the bottom of the defect when entering the damaged area, as illustrated in Figure 4b. At the moment when the rolling element initially contacts the defect bottom, the contact deformation reaches its maximum peak. As the rolling element exits the defect region, the contact deformation exhibits a nonlinear gradual variation. Over a complete cycle of passing through the defect region, the contact deformation amplitude $H'_{\max}$ is related to both the defect depth and its circumferential size. The theoretical radial displacement $H'_{\max}$ is equal to the defect depth H. In this case, the rolling element falls into the defect and contacts its bottom, which typically occurs in the later stage of fault development. Therefore, the expression of $H_{\max}$ is given in Equation (6) [2].

$$H_{\max} = \begin{cases} H'_{\max}, H'_{\max} < H \\ H, H'_{\max} \geq H \end{cases} \tag{6}$$

Accordingly, for the first case, the displacement excitation function is defined as Equation (7) [1].

$$H' = \begin{cases} H_{\max} \sin\{\pi[\mathrm{mod}(\theta_i, 2\pi) - \beta_0]/\beta\}, 0 \leq \mathrm{mod}(\theta_i, 2\pi) - \beta_0 \leq \beta \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad , other \end{cases} \tag{7}$$

While for the second case, it is described by the piecewise function in Equation (8) [1].

$$H' = \begin{cases} H_{\max} \sin\{\pi[\mathrm{mod}(\theta_i, 2\pi) - \beta_0]/\beta\}, 0 \leq \mathrm{mod}(\theta_i, 2\pi) - \beta_0 \leq \beta_q \\ H_{\max} \qquad\qquad\qquad\qquad\qquad , \beta_q \leq \mathrm{mod}(\theta_i, 2\pi) - \beta_0 \leq \beta - \beta_q \\ H_{\max} \sin\{\pi[\mathrm{mod}(\theta_i, 2\pi) - \beta_0]/\beta\}, \beta - \beta_q \leq \mathrm{mod}(\theta_i, 2\pi) - \beta_0 \leq \beta \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad , other \end{cases} \tag{8}$$

where β_0 denotes the initial angular position of the localized defect, β represents the angular extent of the defect region, β_q is the angular position at which the rolling element contacts the defect bottom, and θ_i denotes the angular position of the i-th rolling element at time t.

3.1.2. Influence of Rolling Element Slip

At present, elasto hydrodynamic lubrication theory has become well established. Relevant studies indicate that, under complex operating conditions, a micrometer-scale lubricating film is formed between the rolling elements and the inner and outer raceways of bearings. Since the surface roughness of the raceway is comparable to the thickness of the oil film, the lubrication state is primarily governed by the film thickness, highlighting the critical role of surface roughness in bearing performance.

Under specific load conditions, rolling elements perform periodic circular motion driven by the cage; however, their motion at the contact interface does not always remain pure rolling. In particular, when the load direction angle varies dynamically, slip behavior is likely to occur. Notably, the presence of localized defects can significantly intensify this slip phenomenon, thereby increasing the vibration amplitude of the bearing. The slip behavior between the rolling elements and the raceways not only affects the stability of their motion trajectories but also alters the rotational angular velocity of the cage through contact force transmission. To model this effect, previous studies have introduced uniformly distributed random variables to characterize the slip ratio [24]. Accordingly, the angular position of the i-th rolling element in the load zone at time t under slip conditions can be expressed as Equation (9) [2]:

$$\theta_i = \frac{2\pi(i-1)}{N} + w_c t + \theta_0 + (0.5 - rand) \times \phi_s \tag{9}$$

where ϕ_s denotes the phase deviation induced by slip, typically ranging from 0.01 to 0.02 rad.

3.1.3. Dynamic Model of Localized Progressive Faults

Based on the nonlinear Hertzian contact force–deformation relationship described above, the total nonlinear contact forces of the bearing in the x and y directions can be expressed as Equations (10) and (11) [1], respectively.

$$F_x = K\sum_{i=1}^{z} \lambda(\delta_i)\left[x\cos\theta_i + y\sin\theta_i - C_\gamma - H'\right]^{\frac{3}{2}} \sin\theta_i \tag{10}$$

$$F_y = K\sum_{i=1}^{z} \lambda(\delta_i)\left[x\cos\theta_i + y\sin\theta_i - C_\gamma - H'\right]^{\frac{3}{2}} \cos\theta_i \tag{11}$$

By applying Newton's second law to the established two-degree-of-freedom dynamic model, the governing dynamic equations of the system are derived, as shown in Equation (12) [2].

$$\begin{cases} M\dfrac{d^2x}{dt^2} + C\dfrac{dx}{dt} + K\sum\limits_{i=1}^{z} \lambda(\delta_i)[x\cos\theta_i + y\sin\theta_i - C_\gamma - H']^{\frac{3}{2}} \sin\theta_i = W_x \\ M\dfrac{d^2y}{dt^2} + C\dfrac{dy}{dt} + K\sum\limits_{i=1}^{z} \lambda(\delta_i)[x\cos\theta_i + y\sin\theta_i - C_\gamma - H']^{\frac{3}{2}} \cos\theta_i = W_y \end{cases} \tag{12}$$

where K represents the contact stiffness of the inner and outer raceways, while λ is a switching function that characterizes the contact state between the rolling elements and the

raceways. Equation (13) effectively captures the nonlinear dynamic behavior of bearings under localized fault conditions [1].

$$\lambda(\delta_i) = \begin{cases} 1, \delta_i > 0 \\ 0, \delta_i \leq 0 \end{cases} \tag{13}$$

3.1.4. RUL Mathematical Model of Bearing

The fatigue life of a bearing can be calculated according to [23], as expressed in Equation (14):

$$L10 = (C/P)^p(13) \tag{14}$$

where $L10$ denotes the rated life at which 90% of bearings can operate under a specified load without failure, C represents the basic dynamic load capacity, P is the equivalent dynamic load, and p is the life exponent.

Based on the fatigue life, the RUL of the bearing is defined in Equation (15) [20]:

$$RUL = (L10 - L)/L10 * 100\% \tag{15}$$

where RUL represents the proportion of remaining life, L_{10} is the rated fatigue life, and L denotes the actual service life. Using Equation (15), the RUL of the bearing can be effectively estimated. This estimation provides a basis for maintenance decision-making, enabling timely bearing replacement and reducing potential losses caused by unexpected failures.

3.2. TSMixer

The Time-series Mixer (TSMixer) model is constructed by stacking multiple multilayer perceptrons (MLPs) and efficiently extracts information through joint mixing operations along both the temporal and feature dimensions. Simple univariate linear models can outperform high-capacity sequence deep learning models based on recurrent neural networks or attention mechanisms on several widely used benchmark datasets. Furthermore, effectively leveraging cross-variable and auxiliary information can significantly improve time-series forecasting performance. Therefore, in this study, the TSMixer model is employed to extract temporal features from power grid data, enabling the capture of complex dependencies among time steps and the learning of temporal dynamics within the input sequence.

For each input feature dataset, the data are first fed into the TSMixer model for temporal feature extraction. The process begins with the temporal mixing stage. Given an input tensor $X \in \mathbb{R}^{B \times T \times F}$, where B denotes the batch size, T represents the length of the time series, and F is the feature dimension, the temporal sequence vector $x_f \in \mathbb{R}^{B \times T}$ of each feature channel is processed using an MLP operating along the temporal dimension [13]:

$$\widetilde{x}_f = MLP_f^T(x_f) \tag{16}$$

where f denotes a single feature channel.

After temporal processing, the outputs are concatenated back into tensor form and combined with the original tensor through a residual connection, yielding [13]:

$$H^{(1)} = X + \widetilde{X}^{(T)} \tag{17}$$

where $\widetilde{X}^{(T)} = [\widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3, \ldots, \widetilde{x}_F] \in \mathbb{R}^{B \times T \times F}$.

Following temporal mixing, feature-channel mixing is performed. For the time series, the feature vector at the t-th time step, denoted as $h_t \in \mathbb{R}^{B\times F}$, is processed using another MLP [14]:

$$\widetilde{h}_t = MLP_t^F(h_t) \tag{18}$$

Similarly, after concatenation and residual connection, the output can be expressed as [14]:

$$H^{(2)} = H^{(1)} + \widetilde{X}^{(F)} \tag{19}$$

where $\widetilde{X}^{(F)} = [\widetilde{h}_1, \widetilde{h}_2, \widetilde{h}_3, \ldots, \widetilde{h}_T] \in \mathbb{R}^{B\times T\times F}$.

The MLP adopts a standard two-layer architecture, formulated as [14]:

$$MLP(x) = W_2 \cdot (W_1 \cdot x + b_1) + b_2 \tag{20}$$

where W_1 and W_2 denote the weight matrices from the input layer to the hidden layer and from the hidden layer to the output layer, respectively; b_1 and b_2 are the bias vectors; and $\sigma(\cdot)$ represents the nonlinear activation function.

3.3. DT-TSMixer Method

3.3.1. Center Particle Swarm Optimization

A particle swarm consists of N particles moving in a d-dimensional search space. At the t-th iteration, the position and velocity of the i-th particle are denoted as $X_i^{(t)} = (x_{i1}, x_{i2}, \cdots, x_{id})$ and $V_i^{(t)} = (v_{i1}, v_{i2}, \cdots, v_{id})$, respectively. The personal best position achieved by the particle is represented as $P_i^{(t)} = (p_{i1}, p_{i2}, \cdots, p_{id})$, while the global best position of the entire swarm is denoted as $G_i^{(t)} = (g_{i1}, g_{i2}, \cdots, g_{id})$. The position and velocity of the i-th particle in the next iteration are updated according to Equations (21) and (22) [20].

$$V_{id}^{(t+1)} = w \times V_{id}^t + c_1 \times rand(\,) \times \left(p_{id} - X_{id}^t\right) + c_2 \times rand(\,) \times \left(p_{id} - X_{id}^t\right) \tag{21}$$

$$X_{id}^{(t+1)} = X_{id}^t + X_{id}^{(t+1)} \tag{22}$$

where c_1 and c_2 denote the cognitive and social learning factors, respectively. $Rand(\,)$ are random variables uniformly distributed within [0, 1]. The inertia weight ω decreases linearly from 0.9 to 0.4 during the optimization process to balance global exploration and local exploitation. The particle velocity is constrained within a predefined range $[V_{\min}, V_{\max}]$. If any component of the velocity exceeds the specified threshold $V_{\min}$ or $V_{\max}$, it is clipped to the corresponding boundary.

The center particle swarm optimization (CPSO) algorithm is inspired by collective swarm behavior [20]. As indicated by the velocity update Equation (22), the particle velocity is influenced by three components: the previous velocity, individual cognition, and social cognition. The social component drives particles toward the global optimum, while the individual cognition and inertia terms preserve particle diversity and guide the search process.

In CPSO, $N-1$ particles update their positions using the standard particle swarm optimization mechanism, as expressed in Equation (23) [20].

$$X_{cd}^{(t+1)} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_{id}^{(t+1)} \tag{23}$$

Unlike conventional particles, the central particle not only updates its position but also participates in additional processes, such as fitness evaluation and competition for the global optimum, thereby enhancing the convergence performance of the algorithm.

Based on this mechanism, a CPSO-based hybrid optimization algorithm is developed. The Algorithm 1 pseudocode is summarized as follows:

Algorithm 1: CPSO-based Optimization

Input: Particle positions $X_i^{(t)} = (x_{i1}, x_{i2}, \cdots, x_{id})$, Particle velocities $V_i^{(t)} = (v_{i1}, v_{i2}, \cdots, v_{id})$

Output: Optimized particle positions $X_{cd}^{(t+1)} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_{id}^{(t+1)}$

1. Initialize particle swarm
2. Initialize parameters
3. For $t = 1$ to maximum iteration i
4. $(Adaptivevalue)_i^{(t)} = Adaptivevalueassessment\left(X_c^t\right)$
5. Update personal best and global best
6. For $i = 1$ to $N - 1$
7. $(Adaptivevalue)_i^{(t)} = Adaptivevalueassessment\left(X_i^t\right)$
8. Update velocity $V_i^{(t+1)}$
9. Apply velocity constraints $V_i^{(t+1)}$
10. Update position
11. End for
12. Update central particle position
13. If convergence criterion is satisfied, terminate
14. End for

To quantitatively evaluate the prediction performance of the DT-TSMixer model, a score function is introduced as an additional performance metric. Unlike conventional error-based indicators such as MAE and RMSE, the score function provides a comprehensive assessment of prediction quality by considering the deviation between the predicted and actual RUL values. This metric is particularly useful for measuring the overall prognostic capability of the model and facilitating a fair comparison among different prediction methods. The mathematical formulation of the score function is defined as follows:

$$Accuracy = \left(1 - \frac{1}{N}\sum_{i=1}^{N} \frac{\left|\hat{RUL}_i - RUL_i\right|}{RUL_i} \times 100\%\right) \tag{24}$$

where $\hat{RUL}_i$ denotes predicted RUL. RUL_i denotes actual RUL. N denotes the number of test samples.

3.3.2. DT-TSMixer

Figure 6 illustrates the intelligent prediction framework of the DT-TSMixer method. The TSMixer-based data-driven model includes data acquisition, preprocessing, sample partitioning, and normalization. Meanwhile, the DT model is constructed with multiple functional layers, including data acquisition, signal processing, real-time control, mechanism analysis, logic execution, and application layers, enabling multi-level coupling. The DT model provides theoretical predictions of system states, while the TSMixer model generates data-driven predictions. The discrepancy between these two outputs is utilized to correct the DT model, thereby improving prediction accuracy for the operating state of bearing.

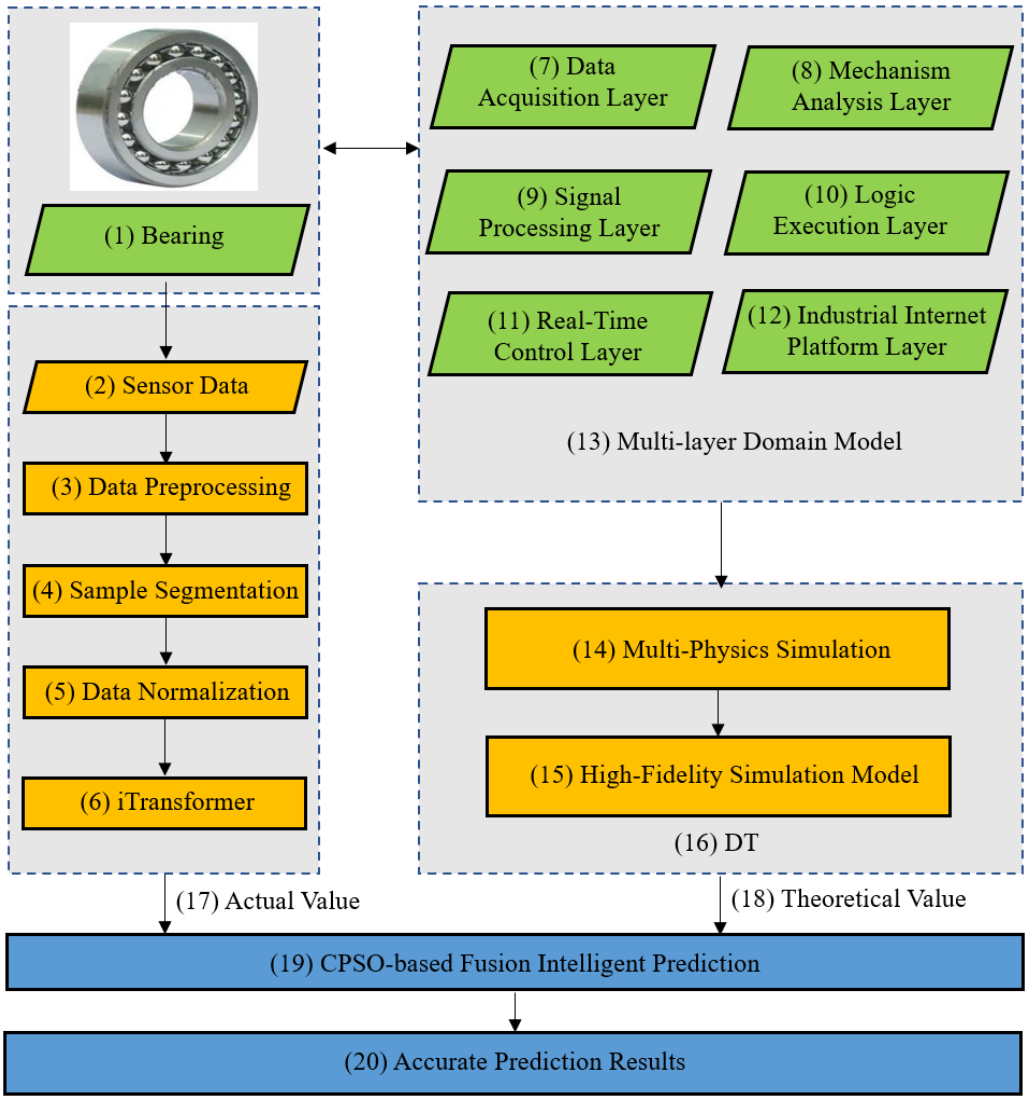


Figure 6. Architecture of the proposed DT-TSMixer prognostic framework.

The core idea of DT-TSMixer is to integrate data-driven modeling with mechanism-driven DT modeling for enhanced predictive performance. The workflow, shown in Figure 7, is summarized as follows:

(1) Construct the TSMixer-based data-driven model and train it using historical data to predict the RUL of REBs.
(2) Transform the state-space model and estimate internal states based on degradation patterns derived from the DT model.
(3) Initialize the CPSO algorithm and use the data-driven observations to correct the DT model outputs, yielding refined RUL predictions.
(4) Compare the predicted results with predefined thresholds: if exceeded, maintenance actions are triggered; otherwise, return to Step (2) for iterative refinement.

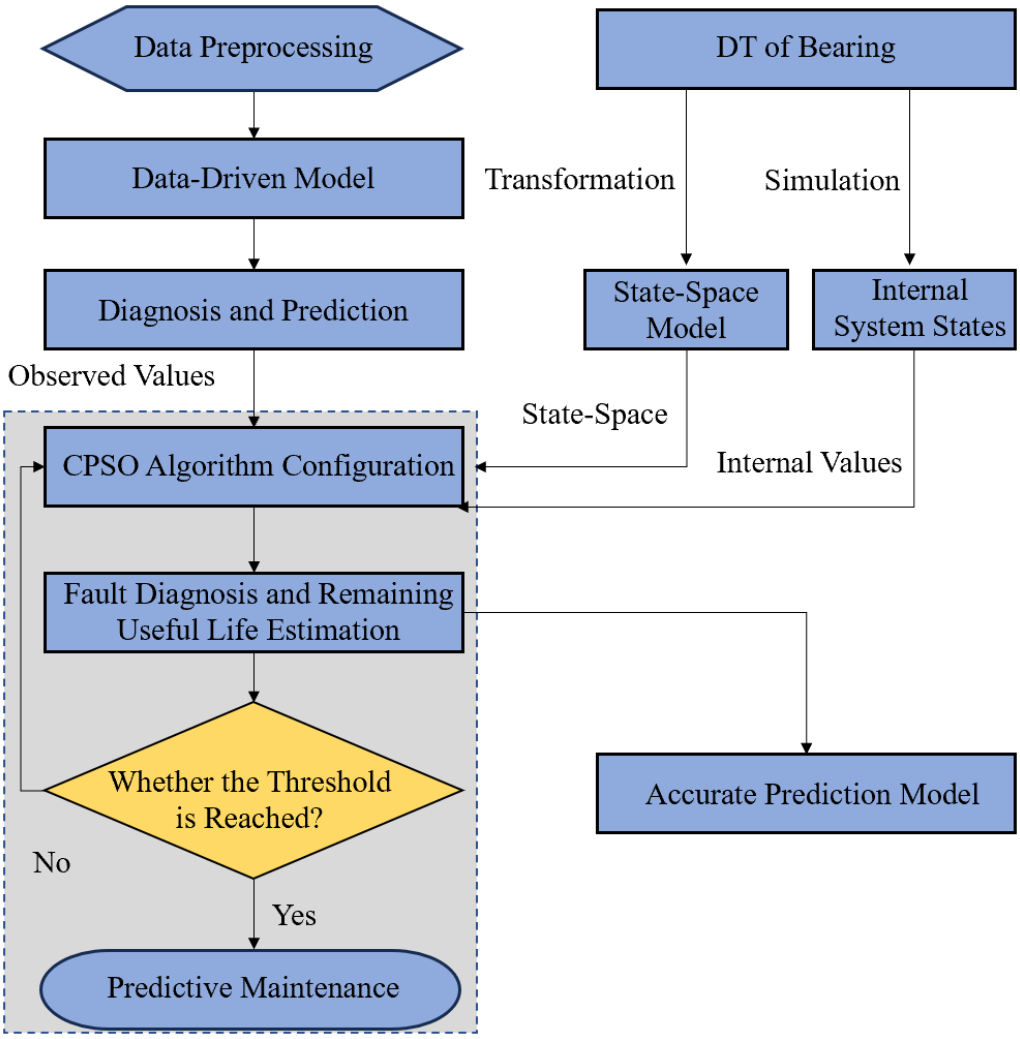


Figure 7. Workflow of the DT-TSMixer-based RUL prediction process.

In Figure 8, the theoretical RUL of the bearing is obtained from Equation (15) and is used as the system state equation to initialize the CPSO algorithm. The predicted value h_t generated by the TSMixer model is derived from Equation (20) and serves as the observation of the CPSO system, with the number of particles set to 150. Algorithm 2 presents the pseudocode for implementing the DT-TSMixer-based RUL prediction method for bearing.

Algorithm 2: Pseudocode for DT-TSMixer-Based RUL Prediction

Input: Theoretical predictions from the DT model and observed predictions from the TSMixer model
Output: Estimated RUL of REB
(1) Initialize CPSO parameters and particle population
(2) $RUL = (L10 - L)/L10 * 100\%$
(3) $h_t = o_t \cdot \tanh(C_t)$
(4) Construct prediction samples
(5) Calculate particle weights based on prediction accuracy h_{ti}
(6) Normalize particle weights $RUL(\mathrm{i})$
(7) Perform resampling based on normalized weights
(8) Output the final RUL prediction of the REB h_f

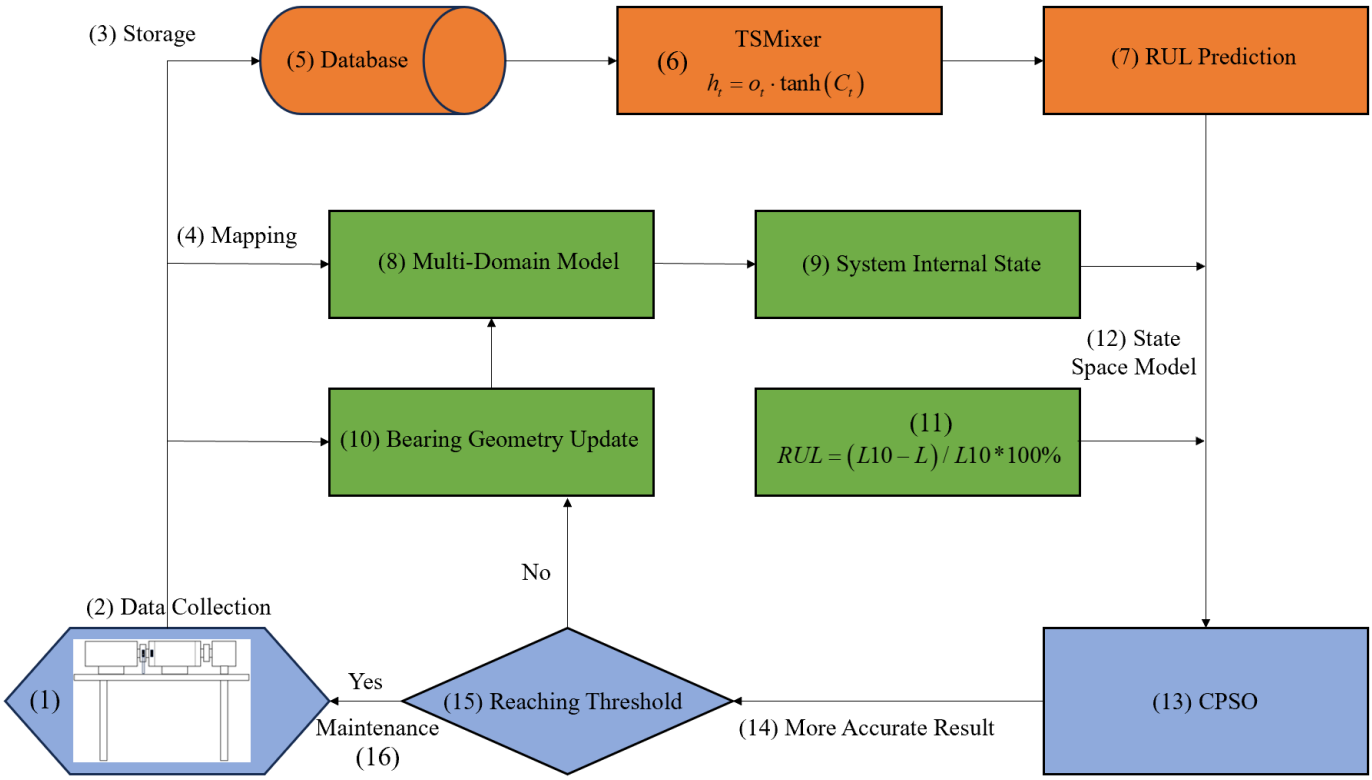


Figure 8. Flowchart of Algorithm 2.

4. Case Study

4.1. Experimental Platform Setup

The experimental setup established in this study consists of two main components: the system unit and the test platform. The former includes a dynamic signal acquisition and analysis system, acceleration sensors, and a photoelectric speed sensor, while the latter comprises a bearing housing, a vibration platform, a magnetic powder brake, and a power cabinet. The basic operating principle of the vibration signal acquisition system is as follows: the power cabinet supplies energy to the bearing housing, and the acceleration signals are collected through the dynamic vibration signal acquisition system and transmitted to the computer-based analysis system for processing. A schematic diagram of the experimental platform is shown in Figure 9.

The bearing parameters are listed in Table 1. The vibration signals are collected using an acceleration sensor mounted on the bearing housing. For all operating conditions, the signals are sampled at a frequency of 25.6 kHz, with each sampling lasting 1.28 s, and the interval between two consecutive samplings set to 1 min.

Table 1. Parameters of bearing.

Pitch Circle Diameter	Rolling Body Diameter	Number of Rolling Bodies	Contact Angle
38.5	20.6	9	0°

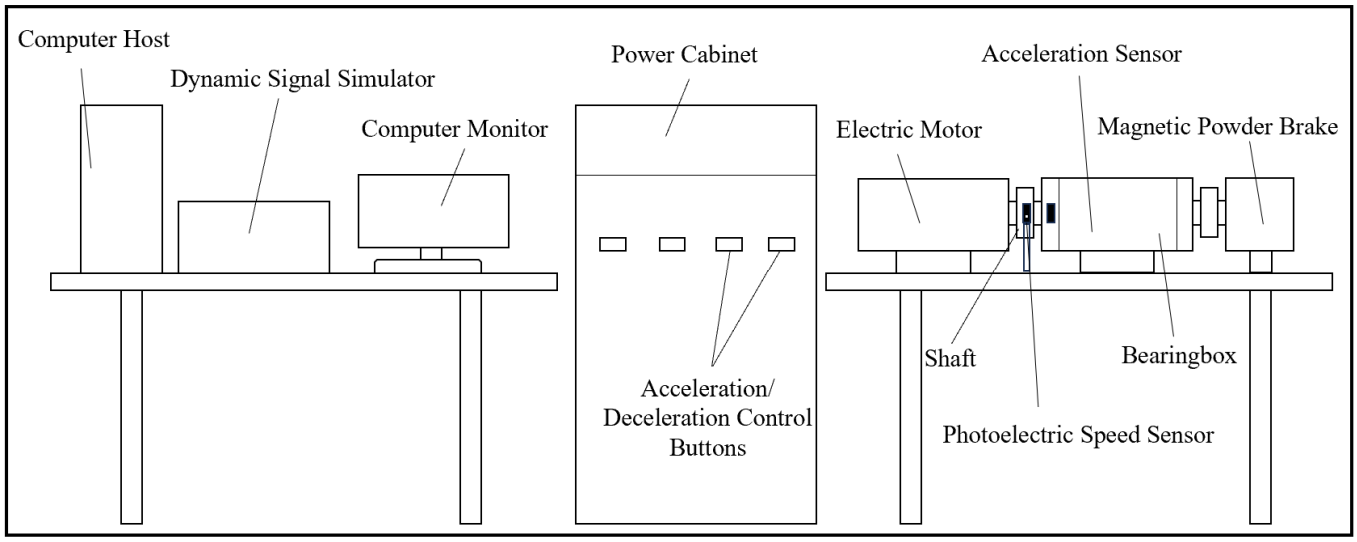


Figure 9. The experimental platform.

4.2. Construction of the Bearing DT Model

Guided by application requirements, the model is simplified and reduced in order while ensuring an accurate representation of the physical entity characteristics. The effects of sealing rings and dust covers on the dynamic analysis are neglected, and pure rolling is assumed between the inner and outer rings and the rolling elements. All bearing components are considered as linear elastic materials. GCr15 steel is adopted, with a density of 7830 kg/m^3, an elastic modulus of 206 GPa, and a Poisson's ratio of 0.3. A two-dimensional automatic surface-to-surface contact model is employed, where the rolling elements are defined as the contact surfaces, and the inner ring, outer ring, and cage are defined as target surfaces. A total of nine contact pairs are established between the balls and the inner/outer rings as well as the cage. A refined mesh is applied to ensure both model accuracy and computational efficiency. Based on the actual operating conditions of the bearing test rig, a fixed constraint is applied to the outer ring, while a rotational constraint is imposed on the inner surface of the inner ring. During the model validation stage, specific driving conditions and loads are applied to simulate real operating conditions. Figure 10 illustrates the initially established DT model of the bearing.

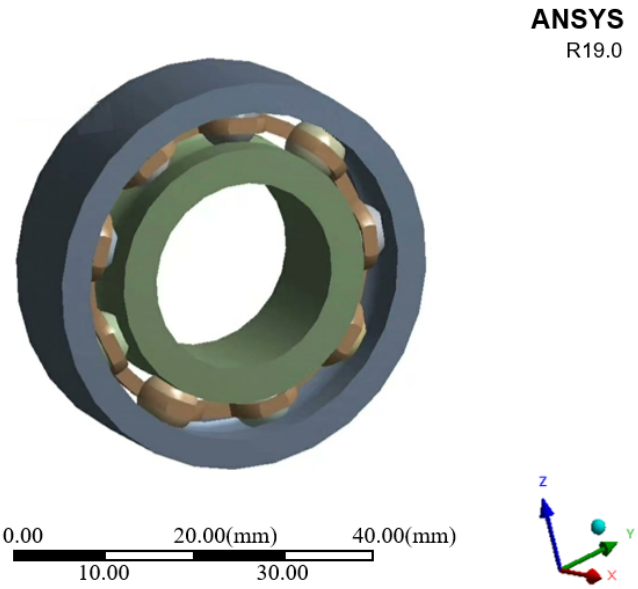


Figure 10. DT model of bearing.

After model construction, the inner ring rotational speed is set to 2000 r/min, a radial load of 1000 N is applied, and the outer ring rotational speed is set to 0. The numerical simulation results are compared with theoretical calculations to determine whether the model satisfies practical requirements. If significant discrepancies exist, the model must be further refined. Model validation and calibration constitute an iterative process, and subsequent analyses are conducted only after the model achieves the required accuracy. The force analysis of the bearing is shown in Figure 11.

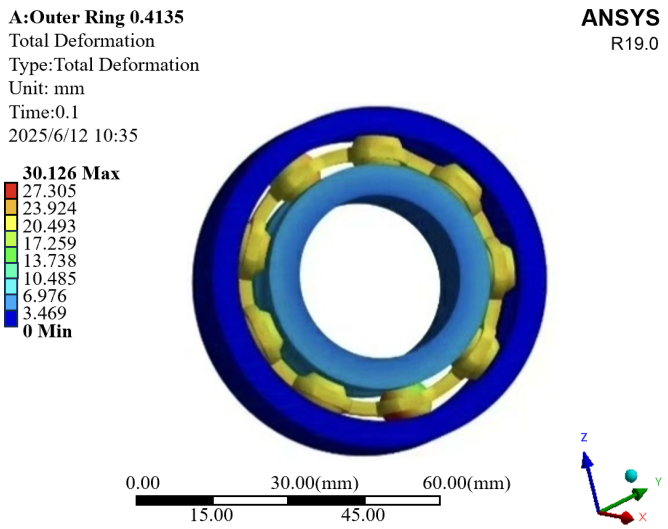


Figure 11. Stress diagram of bearing.

When the rolling elements are assumed to undergo pure rolling, no relative sliding occurs at the contact points. In this case, the velocity of the rolling element is related to the linear velocities of both the inner and outer rings, and can be calculated accordingly. Based on the bearing parameters and the imposed conditions—namely, a constant rotational speed of the inner ring and a fully constrained outer ring—the angle α is introduced to account for the rolling and radial components. Its exact value depends on the bearing design and geometry. For deep groove ball bearing, the rolling motion between the inner and outer rings is typically assumed to be pure rolling; thus, α can be approximated as zero. Substituting into Equation (11) yields a rolling element velocity of v = 2.4701 m/s [15].

$$V = \frac{\pi D_m \left[(N_a + N_b) + \left[(N_b - N_a) \frac{d_m}{D_m} \cos \alpha \right] \right]}{2} \tag{25}$$

To mitigate randomness, two rolling elements are randomly selected, and their velocity curves are illustrated in Figure 12. During operation, the bearing experiences vibration due to factors such as radial clearance and cage pocket clearance, leading to fluctuations in rotational speed. This phenomenon indicates that the developed DT model can effectively capture the real operating characteristics of the bearing. By comparing the data in Figure 11, the errors between the simulated rotational speeds of the two rolling elements and the theoretical values are 1.24% and 0.96%, respectively, demonstrating the validity and feasibility of the proposed bearing DT model.

Figure 13 presents a comparison between the simulated and real vibration signals of the target bearing, while Figure 14 compares the simulated bearing load with the experimental real load. The quantitative results indicate that the accuracy of the bearing vibration signal simulated by the DT model exceeds 98.3%, and the accuracy of the simulated load distribution reaches up to 97.6%. This verifies that the proposed DT model for bearings can accurately reflect the real operating characteristics of physical bearings.

Table 2 presents a comparison between the measured bearing vibration signals and the corresponding DT-generated signals under an operating speed of 2000 r/min. The feature analysis results indicate that the discrepancy between the real and simulated vibration signals is negligible, demonstrating that the proposed DT model can accurately capture the dynamic characteristics of the physical bearing system and maintain a high level of modeling fidelity.

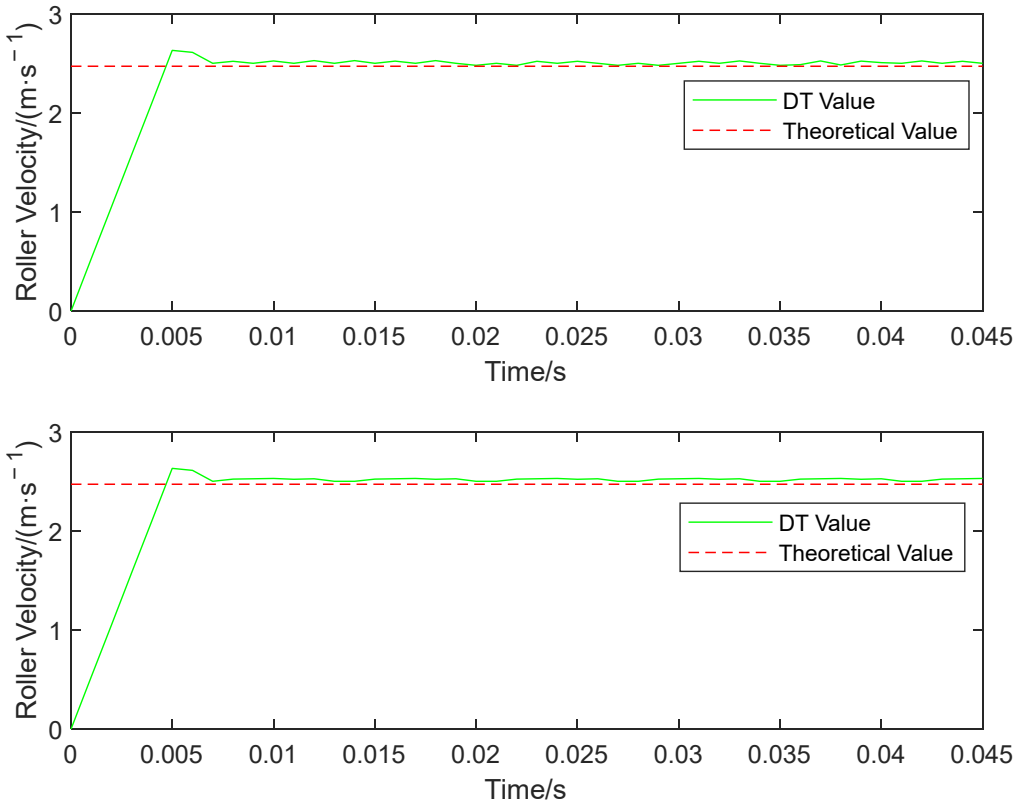


Figure 12. Roller velocity of bearing.

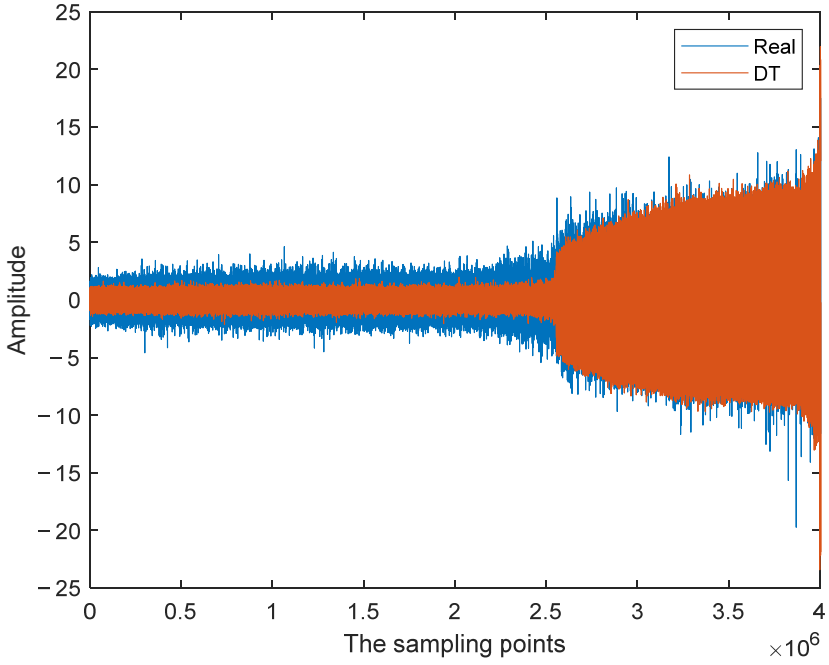


Figure 13. The vibration signal of bearing.

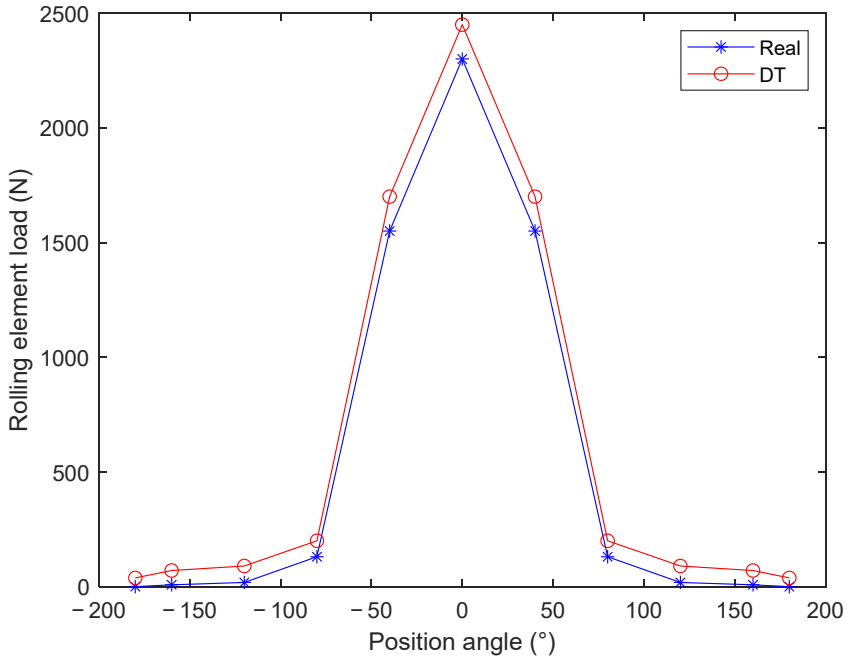


Figure 14. The load distribution of bearing.

Table 2. The comparison of the DT data and actual data.

Data	Peak Value	RMS Value	Skewness	Frequency Domain Feature
Real Data	8.1 g	12.1 g	60^2	69.74 Hz
DT Data	8.4 g	12.5 g	61^2	71.21 Hz
Error	3.70%	3.31%	3.36%	2.11%

4.3. The Realization of TSMixer Model

Normalization, as a fundamental method of data standardization, aims to eliminate the dimensional discrepancies among multi-dimensional features. In this study, the min–max normalization method is adopted to standardize the state data. This approach performs a linear mapping based on the minimum and maximum values within each feature dimension, transforming the original dataset into a normalized range of [0, 1]. The mathematical formulation is given in Equation (26) [16].

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{26}$$

where the minimum and maximum values of each feature are denoted by x_{min} and x_{max}, respectively. The normalized feature values are then combined to form a new dataset x_1'. The processed data significantly improve computational efficiency during model training and reduce the risk of overfitting.

To determine the most suitable network configuration, a parameter sensitivity analysis was performed by systematically varying the number of TSMixer layers and hidden nodes. Table 3 summarizes the corresponding maximum residual errors (MREs) for gear RUL prediction. Among all tested configurations, the model with three TSMixer layers and eighteen hidden nodes achieved the lowest MRE, indicating the best predictive capability and overall model performance.

The processed dataset is divided into a training set and a testing set with a ratio of 7:3. All experiments are conducted on a high-performance virtualized computing platform equipped with NVIDIA virtual GPU technology, including 4 virtual CPU cores, 16 GB

memory, and an NVIDIA VGPU50-12 GPU. The software environment is based on Python 3.8 and the PyTorch 2.0.0 framework. The model is trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 256. All models are trained for 50 epochs, and performance is evaluated on the validation set after each epoch. The model with the lowest validation loss is saved for final testing. To ensure reproducibility, random seeds for PyTorch, NumPy, and Python are fixed, and all hyperparameters are strictly controlled and recorded.

Table 3. The MRE corresponding to different TSMixer layers and hidden node numbers.

Model Structure	TSMixer Three-Layer MRE (μm)	TSMixer Four-Layer MRE (μm)	TSMixer Five-Layer MRE (μm)
ten hidden nodes	10.3	9.3	16
fourteen hidden nodes	8.5	11.2	13
eighteen hidden nodes	5	10.4	21.4
twenty-two hidden nodes	9.1	10.3	17.6

In addition, the model training parameters are configured as follows: the Adam optimizer is adopted with an initial learning rate of 0.001, and the mean squared error (MSE) is used as the loss function to measure the deviation between predicted and true values [16]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (27)$$

where y_i and $\hat{y}_i$ denote the true and predicted values, respectively. The number of training epochs is set to 50, and data are fed in mini-batches. The optimal model is selected based on the minimum average MSE loss on the training set. As illustrated in Figure 15, the MSE decreases rapidly during the early training stage (Epoch 1–10), indicating efficient feature learning. From Epoch 10 to 50, the MSE gradually stabilizes at a low level, suggesting that the model has captured the intrinsic data patterns and reached convergence, thereby demonstrating the effectiveness and stability of the selected hyperparameters.

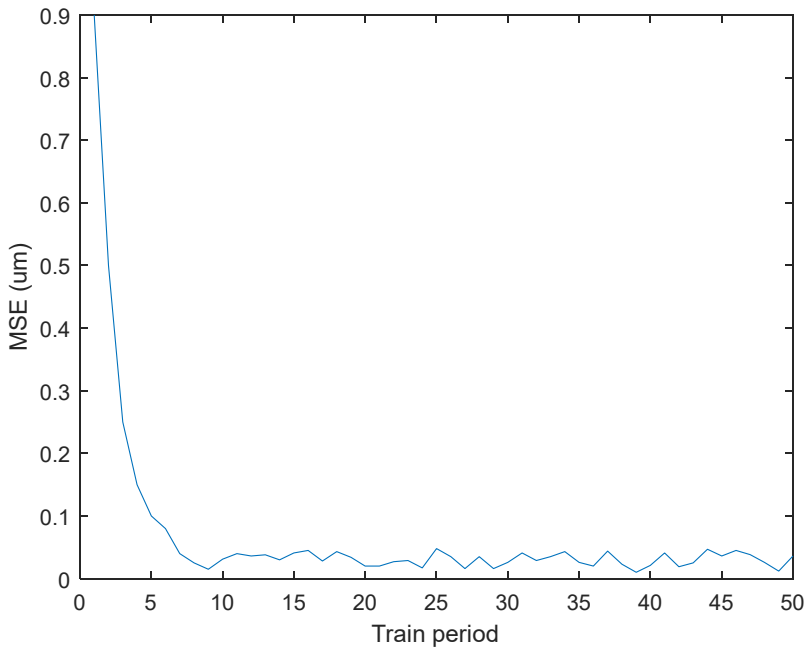


Figure 15. Trend of MSE variation with training epochs.

4.4. The Realization of DT-TSMixer Model

Taking the bearing experiment as an example, the overall prediction trajectory of the DT-TSMixer model exhibits the closest agreement with the true RUL curve. The dataset is divided into a training set and a testing set with a ratio of 7:3. The training dataset is established by integrating historical bearing operation data with high-fidelity DT-generated data under consistent operating conditions. Specifically, time-domain and frequency-domain features are extracted from both data sources and fused to construct a unified multi-source feature representation for TSMixer training. In this hybrid dataset, historical data provide realistic degradation characteristics of the physical bearing system, while DT-generated data enrich the diversity of degradation patterns and compensate for the limited availability of fault samples by providing additional state information. The comparative prediction results of different models are illustrated in Figure 16.

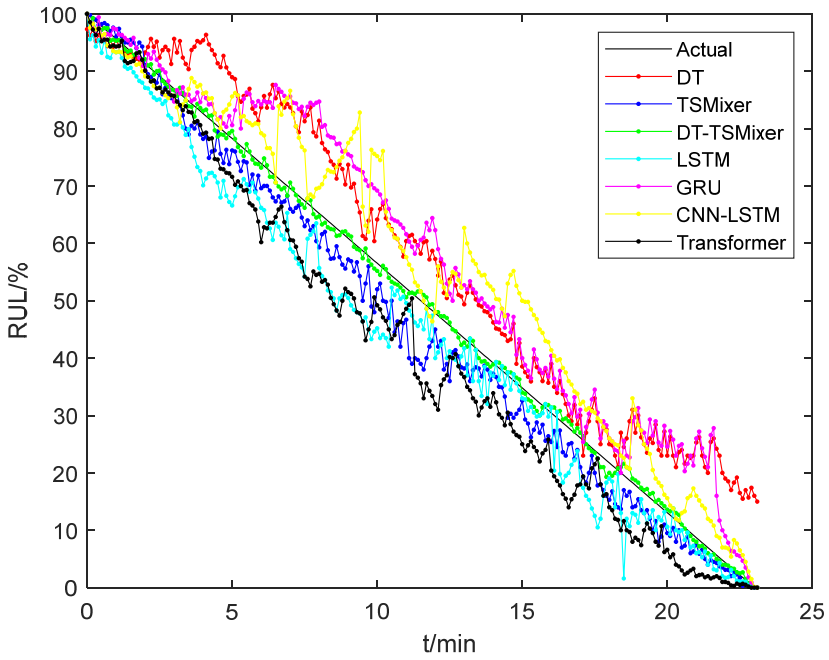


Figure 16. Prediction performance on bearing test data (2000 r/min, 1000 N).

Table 4 shows the accuracy comparison of different models. The results demonstrate that the proposed DT-TSMixer model achieves high prediction accuracy, outperforming baseline models by approximately 9.23%, thereby enabling more precise monitoring of bearing operating conditions. As shown in Figure 16, the DT-TSMixer model captures the degradation trend more rapidly in the early stage. Although slight fluctuations appear in the later stage of degradation, the model maintains superior overall prediction accuracy due to its early-stage advantage. Compared with DT-CNN, CNN, and MLCNN, the experimental results further demonstrate the effectiveness and robustness of the proposed DT-TSMixer model across different evaluation metrics and operating conditions.

To clarify the contribution of the CPSO algorithm, this paper conducts an ablation study. Specifically, the prediction performance of the proposed DT-TSMixer framework was evaluated both with and without the CPSO-based correction mechanism. The comparative results demonstrate that CPSO effectively reduces the discrepancy between the Digital Twin predictions and the data-driven predictions, leading to improvements in the score function, MAE, and RMSE metrics. These results confirm that CPSO contributes significantly to the overall prediction accuracy and robustness of the proposed framework. Figure 17 presents a comparison of several hybrid optimization algorithms, including the Kalman Filter,

Ensemble Learning, and CPSO. The results demonstrate that the proposed CPSO-based approach achieves superior performance, highlighting its enhanced optimization capability. As shown in Table 5, the CPSO-based method achieves the best performance among all compared approaches, yielding the lowest MAE and RMSE values, which demonstrates its superior predictive capability in the RUL estimation task.

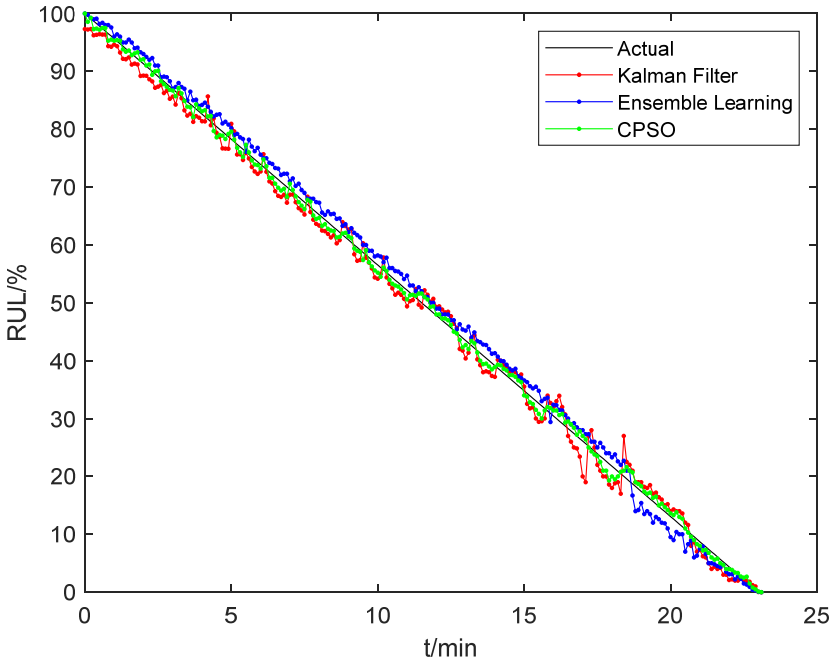


Figure 17. The ablation experiment of CPSO.

Table 4. The accuracy (score function) comparison of different models.

Model	Accuracy
LSTM	80.69%
GRU	82.19%
CNN-LSTM	84.39%
Transformer	87.83%
DT	90.34%
TSMixer	91.39%
DT-TSMixer	99.57%

Table 5. The comparison of the ablation experiment results.

Model	MAE	RMSE
Kalman Filter	0.2069 ± 0.0857	0.2167 ± 0.0589
Ensemble Learning	0.1026 ± 0.0982	0.1695 ± 0.0862
CPSO	0.0528 ± 0.0039	0.0739 ± 0.0098

In addition, this study compares the number of parameters and computational cost of different algorithms, as summarized in Table 6. As shown in Table 6, the CPSO model demonstrates significant advantages in terms of both parameter size and computational complexity. Specifically, the Kalman Filter and Ensemble Learning models contain a relatively large number of parameters, which results in higher memory consumption during both training and inference, thereby increasing hardware resource requirements and computational burden. In contrast, the CPSO model has only 1.23×10^6 parameters, which is the smallest among all compared methods, significantly reducing memory usage.

Moreover, the computational cost of CPSO is considerably lower than that of the other models, demonstrating the effectiveness of its lightweight design and its suitability for efficient implementation.

Table 6. The parameter counts and operations comparison of different models.

Model	Number of Parameters/10^6	Computation Amount/(10^6 FLOPs)
Kalman Filter	12.69	168
Ensemble Learning	10.68	103
CPSO	1.23	15

To further assess the statistical significance of the performance improvement, the DT-TSMixer model and all baseline models are evaluated through 10 independent runs with different random seeds. A paired t-test is employed to analyze the differences in performance metrics. The t-statistic is calculated as [18]:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{28}$$

where $\bar{d}$ represents the mean difference in performance metrics between the DT-TSMixer and baseline models, s_d denotes the standard deviation of the differences, and n = 10 is the number of repeated experiments. The corresponding p-value is obtained based on the t-distribution [18]:

$$p = 2 \times P(T \geq |t|) \tag{29}$$

where T follows a t-distribution with $n - 1$ degrees of freedom.

Table 7 presents the mean and standard deviation over the 10 repeated experiments, while Table 8 summarizes the statistical significance results. The findings indicate that the DT-TSMixer model achieves the best performance in the RUL prediction task, with MAE (0.0867 ± 0.0072) and RMSE (0.0927 ± 0.0036), both lower than those of all comparison models. Furthermore, the p-values for all comparisons are less than 0.01, confirming that the performance improvements of the proposed model are statistically significant.

Table 7. The performance comparison of different models.

Model	MAE	RMSE
LSTM	0.2689 ± 0.0276	0.2829 ± 0.0217
GRU	0.2397 ± 0.0692	0.2681 ± 0.0638
CNN-LSTM	0.2138 ± 0.0528	0.2571 ± 0.0258
Transformer	0.2689 ± 0.0576	0.3068 ± 0.0286
DT	0.2568 ± 0.0658	0.2937 ± 0.0369
TSMixer	0.1736 ± 0.0576	0.2136 ± 0.0571
DT-TSMixer	0.0867 ± 0.0072	0.0927 ± 0.0036

Although the current study is validated using a benchmark bearing life-cycle dataset under controlled operating conditions, the DT-TSMixer framework is designed as a general hybrid prediction architecture that can be extended to different operating conditions and other rotating machinery systems. To verify the feasibility and effectiveness of the proposed DT-TSMixer framework, this paper conducts a complete life-cycle bearing dataset under two representative operating conditions (1500 r/min and 1000 N radial load; 2000 r/min and 1500 N radial load;), as shown in Figures 18 and 19. Table 9 compares the accuracy of different existing methods, and the experimental results indicate that the DT-TSMixer

method significantly outperforms both the TL-ConvNext Method and the DT Drive Method, with an accuracy improvement of around 4.12%.

Table 8. The results of statistical significance tests.

Comparison Model	Comparison MAE	Comparison RMSE
LSTM	t = 27.3587 $p = 3.37 \times 10^{-15}$	t = 42.9625 $p = 6.89 \times 10^{-16}$
GRU	t = 16.2687 $p = 2.65 \times 10^{-17}$	t = 15.2857 $p = 3.82 \times 10^{-18}$
CNN-LSTM	t = 18.3589 $p = 3.26 \times 10^{-18}$	t = 20.5821 $p = 4.96 \times 10^{-19}$
Transformer	t = 28.6872 $p = 1.53 \times 10^{-16}$	t = 27.6873 $p = 7.35 \times 10^{-17}$
DT	t = 37.3692 $p = 2.28 \times 10^{-19}$	t = 42.2581 $p = 7.36 \times 10^{-20}$
TSMixer	t = 6.5438 $p = 1.42 \times 10^{-6}$	t = 6.2143 $p = 3.67 \times 10^{-6}$
DT-TSMixer	t = 8.5249 $p = 7.16 \times 10^{-8}$	t = 10.2368 $p = 4.07 \times 10^{-9}$

Table 9. The accuracy comparison of various methods.

Method	Accuracy
TL-ConvNext Method [29]	94.6%
DT Drive Method [30]	95.6%
DT-TSMixer	98.72%

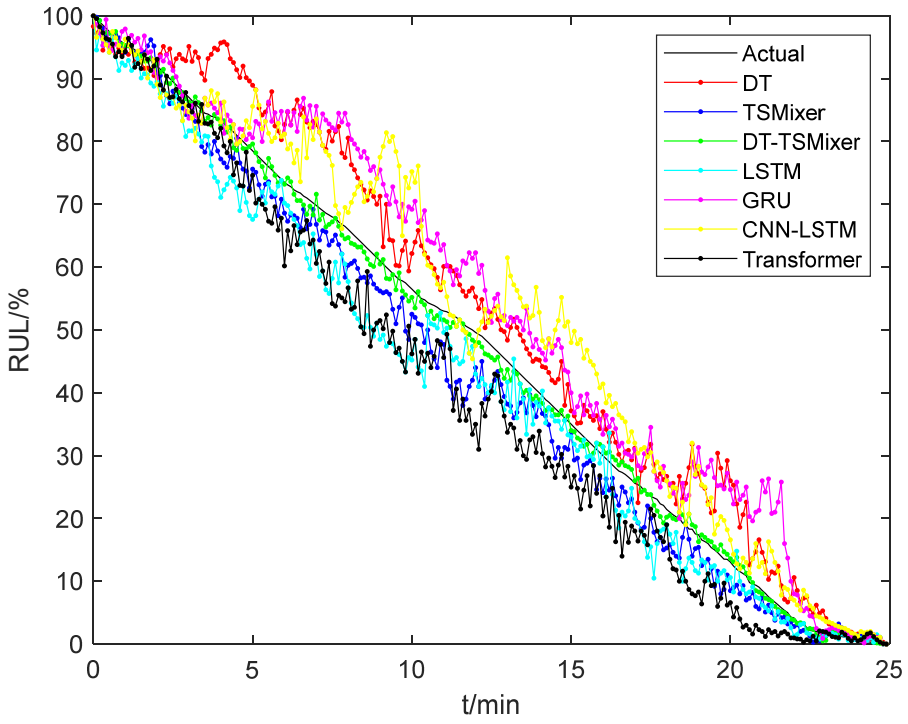


Figure 18. Prediction performance on bearing test data (1500 r/min, 1000 N).

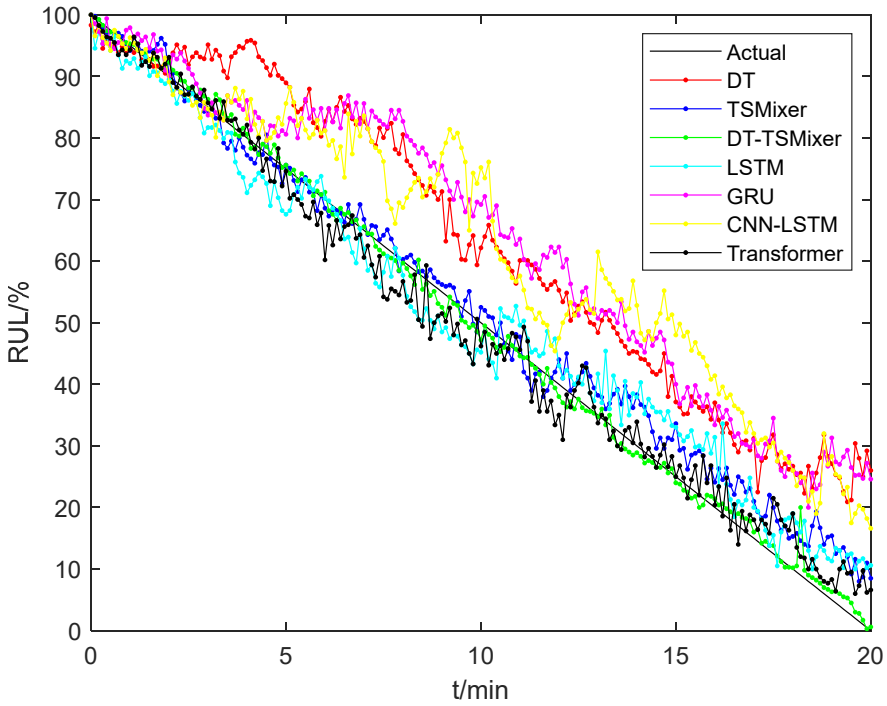


Figure 19. Prediction performance on bearing test data (2000 r/min, 1500 N).

5. Conclusions

This work addresses the critical data scarcity problem that limits the RUL prediction performance of bearings, and proposes a hybrid intelligent prediction approach that organically combines DT technology with advanced data-driven modeling. To be specific, a DT model constructed with physical constraint guidance is introduced to supplement the high-precision real-time degradation features that are missing in limited sensor data. The DT-corrected real-time data are merged with historical datasets and processed through a feature alignment module, which outputs a sufficiently large and highly representative dataset to support the full training of the TSMixer model. Extensive ablation and comparison experiments verify that the proposed DT-TSMixer model achieves exceptional RUL prediction accuracy, surpassing the mainstream baseline models by approximately 9.23%, and realizes more fine-grained perception of the bearing operation degradation process. The research results provide an effective solution for high-precision bearing RUL prediction, making it possible to detect and eliminate potential faults in advance before the occurrence of complete failure, and further improve the overall operation efficiency and reliability of the mechanical equipment.

Compared with attention-based architectures, TSMixer adopts a lightweight MLP-based structure, which significantly reduces the computational burden and enables efficient online inference; the integration of the DT model and CPSO-based correction mechanism inevitably increases the computational burden during model updating and state estimation. In addition, the DT model is developed based on several simplifying assumptions, such as a reduced-order bearing dynamic model and idealized contact conditions, which may limit its ability to fully represent highly complex real-world operating environments. Practical deployment also faces challenges related to the acquisition of high-quality sensor data, real-time synchronization between physical and virtual systems, and the computational requirements for online predictive analysis. Future research will therefore focus on developing adaptive DT models with online parameter updating, integrating multi-source heterogeneous sensor data, reducing computational complexity through lightweight model

design, and extending the proposed framework to more complex industrial scenarios with variable operating conditions and multiple fault modes.

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

Funding: This work was supported by the Project Supported by the Chongqing Natural Science Foundation General Project (Grant No. CSTB2025NSCQ-GPX0115), the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202503114).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting this study can be available upon reasonable request to the authors.

Conflicts of Interest: Author Mr. Dong Zhu was employed by the company Sevnce Robotics Co., Ltd., Chongqing. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RUL	Remaining Useful Life
DT	Digital Twin
TSMixer	Time-series Mixer
TKEO	Teager–Kaiser Energy Operator
MLPs	Multiple Multilayer Perceptrons
CPSO	Center Particle Swarm Optimization
MRE	Maximum Rresidual Error
MSE	Mean Squared Error

References

1. Deng, L.; Li, W.; Yan, X. An intelligent hybrid deep learning model for rolling bearing remaining useful life prediction. *Nondestruct. Test. Eval.* **2025**, *40*, 2670–2697.
2. Fan, J.; Liang, W.; Zhang, W.Q. SARNet: A spike-aware consecutive validation framework for accurate remaining useful life prediction. In Proceedings of the 2026 IEEE International Conference on Acoustics, Speech and Signal Processing, Beijing, China, 3–8 May 2026; pp. 1201–1205.
3. Zhou, J.; Yang, J.; Xiang, S.; Qin, Y. Remaining useful life prediction methodologies with health indicator dependence for rotating machinery: A comprehensive review. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3528519. [CrossRef]
4. Li, X.; Zhou, P.; Aruin, A.S. Teager–Kaiser energy operation of surface EMG improves muscle activity onset detection. *Ann. Biomed. Eng.* **2007**, *35*, 1532–1538. [CrossRef] [PubMed]
5. Tran, V.T.; AlThobiani, F.; Ball, A. An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert Syst. Appl.* **2014**, *41*, 4113–4122. [CrossRef]
6. Tang, Y.; Wang, K.; Niyato, D.; Li, J.; Dobre, O.A.; Duong, T.Q. Secure data sharing and prediction with digital twin and blockchain in healthcare. *IEEE Commun. Mag.* **2025**, *63*, 170–178. [CrossRef]
7. Oksuz Gurdal, B.; Testik, O.M. A framework for product life cycle management based digital twin implementation in the aerospace industry. *Appl. Stoch. Models Bus. Ind.* **2025**, *41*, e70001. [CrossRef]
8. Mehdipour, S.; Habibzadeh, A.; Esmi, N.; Heravi, P.; Mirsaraee, M.; Shahbahrami, A.; Javanmardi, Z. Integrating digital twins, data management and artificial intelligence for predictive maintenance in aviation: A comprehensive review. *Digit. Twins Appl.* **2026**, *3*, e70029. [CrossRef]
9. Zhang, G.; Zhang, K.; Shen, W.; Xu, G.; Lin, Y.-J. Intelligent fault diagnosis of multi-way directional valves in hydraulic systems using digital twin and deep learning approaches. *Mech. Syst. Signal Process.* **2025**, *230*, 112579. [CrossRef]

10. Zhou, T.; Zhang, M.; Hu, T.; Meng, L.; Yi, M.; Zhang, J.; Xu, C. Research on the online monitoring method of cutting tool wear based on the mechanism-data fusion concept of digital twin. *J. Manuf. Process.* **2025**, *148*, 386–407. [CrossRef]
11. Wang, Q.; Zhang, X.; Lu, J.; Xiao, G.; Ren, Y.; Li, W. Digital twin-driven physically constrained generative adversarial network for industrial boiler fault diagnosis. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3530215. [CrossRef]
12. Ma, S.; Leng, J.; Zheng, P.; Chen, Z.; Li, B.; Li, W.; Liu, Q.; Chen, X. A digital twin-assisted deep transfer learning method towards intelligent thermal error modeling of electric spindles. *J. Intell. Manuf.* **2025**, *36*, 1659–1688.
13. Lee, Y.; Jeong, J. TSMixer- and transfer learning-based highly reliable prediction with short-term time series data in small-scale solar power generation systems. *Energies* **2025**, *18*, 765.
14. Yang, C.; Li, H.; Ma, Y.; Huang, Y.; Chu, X. Enhanced TSMixer model for the prediction and control of particulate matter. *Sustainability* **2025**, *17*, 2933. [CrossRef]
15. Pei, Z.; Yan, J.; Yan, J.; Yang, B.; Liu, X. Multi-scale TsMixer: A novel time-series architecture for predicting a-share stock index futures. *Mathematics* **2025**, *13*, 1415.
16. Yamazono, R.; Hachiya, H. Permutation dependent feature mixing for multivariate time series forecasting. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Porto, Portugal, 15–19 September 2024; pp. 301–316.
17. Kukadiya, P.; Prasad, A.E.; Pandey, A.K.; Jain, T.; Hubballi, N.; Shukla, S. Enhancement of modified TSMixer model and validation on field data for synchrophasor measurement prediction under cyber attacks. *IEEE Trans. Ind. Appl.* **2026**, *62*, 1–13. [CrossRef]
18. Li, L.; Zheng, B. Penetration rate forecasting for hard rock TBMs using the TBMTSMixer model. *Geotech. Geol. Eng.* **2025**, *43*, 256. [CrossRef]
19. Cao, L.; Shao, X.; Wang, C.; Gao, J. Residual-corrected TSMixer with a RIME-enhanced decomposition strategy for photovoltaic power prediction. *Electr. Power Syst. Res.* **2026**, *253*, 112554.
20. Han, X.; Guo, Z.; Zhang, J.; Li, Y. Hard disk failure prediction model based on multilayer convolution, attention mechanism and TSMixer. In Proceedings of the 17th International Conference on Machine Learning and Computing, Guangzhou, China, 14–17 February 2025; pp. 671–681.
21. Dai, W.; Liu, S.; Lv, S.; Hu, J.; Wang, X. Multi-stage tool wear prediction method based on multi-dimensional digital twin. *J. Intell. Manuf.* **2026**, *37*, 597–614.
22. Lu, Q.; Li, M.; Huang, X. Digital twin and data-driven remaining useful life prediction of gearbox. *IEEE Access* **2025**, *13*, 111614–111627. [CrossRef]
23. Cao, Z.; Chen, L.; Li, C.; Zhu, L.; Deng, Z. Tool-health digital twin for CNC predictive maintenance via innovation-adaptive sensor fusion and uncertainty-aware prognostics. *Machines* **2026**, *14*, 335.
24. Feng, Y.; Fan, K. Digital twin–driven causal diagnosis mechanism for life health of high-speed spindle system. *Int. J. Adv. Manuf. Technol.* **2024**, *134*, 1077–1089.
25. Kobayashi, K.; Alam, S.B. Explainable, interpretable, and trustworthy AI for an intelligent digital twin: A case study on remaining useful life. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107620.
26. Zhang, M.; Amiri, A.; Xu, Y.; Bastin, L.; Clark, T. Self-adaptive digital twin of fuel cell for remaining useful lifetime prediction. *Int. J. Hydrogen Energy* **2024**, *89*, 634–647. [CrossRef]
27. Sumesh, C.S.; Ramesh, A. Digital twin-enabled surface quality prediction and optimization in dry turning of Ti6Al4V using ANFIS and genetic algorithm. *J. King Saud Univ. Eng. Sci.* **2025**, *37*, 17.
28. Wei, Y.; Hu, T.; Wei, S.; Ma, S.; Wang, Y. Digital twin technology applicability evaluation method for CNC machine tool. *Int. J. Adv. Manuf. Technol.* **2024**, *131*, 5607–5623.
29. Zhang, C.; Qin, F.; Zhao, W.; Li, J.; Liu, T. Research on rolling bearing fault diagnosis based on digital twin data and improved ConvNext. *Sensors* **2023**, *23*, 5334. [CrossRef] [PubMed]
30. Qin, Y.; Liu, H.; Mao, Y. Faulty rolling bearing digital twin model and its application in fault diagnosis with imbalanced samples. *Adv. Eng. Inform.* **2024**, *61*, 102513. [CrossRef]

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